INSTITUTO
UNIVERSITÁRIO
DE LISBOA

Unveiling the Far-Right's Political Ascent in Portugal: An NLP and AI Approach

Mauro Jorge Costa Cardoso

Master in Computer Engineering

Supervisor:
PhD Fernando Manuel Marques Batista, Associate Professor,
Iscte – Instituto Universitário de Lisboa

Supervisor:
PhD Eugénio Alves Ribeiro, Assistant Professor,
Iscte – Instituto Universitário de Lisboa

September, 2025

TECHNOLOGY
AND ARCHITECTURE

Department of Information Science and Technology

Unveiling the Far-Right's Political Ascent in Portugal: An NLP and AI Approach

Mauro Jorge Costa Cardoso

Master in Computer Engineering

Supervisor:
PhD Fernando Manuel Marques Batista, Associate Professor,
Iscte – Instituto Universitário de Lisboa

Supervisor:
PhD Eugénio Alves Ribeiro, Assistant Professor,
Iscte – Instituto Universitário de Lisboa

September, 2025

"Obsession beats talent every time."

Agradecimentos

Em primeiro lugar, dedico este marco da minha vida à minha querida Mãe Elsa, por ser a minha maior inspiração, a tua luta é a minha luta. Aos meus amigos/as, Ilzy, José por serem o meu maior suporte, por conseguirem transformar qualquer percalço ou obstáculo na minha vida em um trampolim. Agradecer-vos nunca será suficiente. Aos meus colegas de trabalho, minha segunda familia, os dias do vosso lado significam terapia. Muito obrigado.

Agradeço aos meus orientadores, Professores Fernando Batista e Eugénio Ribeiro, por acreditarem na ideia de um sonhador e pelos diferentes desafios, colaboração, orientação e disponibilidade na transmissão de conhecimentos, com opiniões e críticas, solidificando as minhas dúvidas e direcionando-me para o caminho certo. Por fim, não posso deixar de agradecer a esta enorme instituição que é o ISCTE.

Mauro Cardoso

Resumo

Este estudo analisa o discurso nas redes sociais de figuras de destaque do partido de extrema-direita português CHEGA, examinando 10.323 publicações no *X* (antigo Twitter) entre o final de 2019 e meados de 2024. Combinando *BERTopic* e classificação zero-shot com *Large Language Models (LLMs)*, foram identificados 59 tópicos latentes, organizados em duas grandes dinâmicas: (1) política, eleitoral e parlamentar; e (2) ideológica e pública.

Os resultados sugerem um discurso privilegiando a confrontação institucional, a retórica anti-elitista e estratégias de campanha, articuladas com narrativas sobre identidade, imigração, segurança e valores morais. Predominam emoções de raiva, ódio e repulsa, associadas a dispositivos retóricos como ironia, hipérbole e apelos à ação, que funcionam como mecanismos de mobilização, provocação e polarização. O recurso a formas indiretas e ofensivas de ódio, como ameaças simbólicas, estereótipos e negação, contribui para normalizar narrativas excludentes ao mesmo tempo que mitiga riscos reputacionais, refletindo padrões da comunicação da extrema-direita europeia.

Apesar de limitações como viés de seleção de dados, ausência de métricas de engajamento e 23% do conteúdo categorizado como *outliers*, o estudo evidencia o potencial das abordagens computacionais para compreender as estratégias discursivas. Mostra ainda como estas figuras recorrem às redes sociais para difundir e normalizar ideias radicais, reforçando o papel central da comunicação digital na ascensão da extrema-direita em Portugal.

Palavras Chave: *Discurso Político*; *Modelação de Tópicos*; *Classificação Zero Shot*; *Extrema Direita*; *CHEGA (Portugal)*; *Redes Sociais*.

Abstract

This study analyzes the social media discourse of prominent figures from the Portuguese far-right party CHEGA, examining 10,323 posts on *X* (formerly *Twitter*) between late 2019 and mid-2024. By combining *BERTopic* with zero-shot text classification using LLMs, 59 latent topics were identified and organized into two main dynamics: (1) political, electoral, and parliamentary; and (2) ideological and public.

The results suggest a discourse that privileges institutional confrontation, anti-elitist rhetoric, and campaign strategies, articulated through narratives on identity, immigration, security, and moral values. Emotions such as anger, hate, and disgust predominate, associated with rhetorical devices such as irony, hyperbole, and calls to action, which function as mechanisms of mobilization, provocation, and polarization. The use of indirect and offensive forms of hate, such as symbolic threats, stereotyping, and denial, contributes to normalizing exclusionary narratives while mitigating reputational risks, reflecting broader patterns in far-right communication across Europe.

Despite limitations such as data selection bias, the absence of engagement metrics, and 23% of the content categorized as outliers, the study highlights the potential of computational approaches to understand discursive strategies. It also shows how these figures use social media to spread and normalize radical ideas, reinforcing the central role of digital communication in the rise of the far right in Portugal.

KEYWORDS: *Political Discourse*; *Topic Modeling*; *Zero Shot Classification*; *Far-Right*; *CHEGA (Portugal)*; *Social Media*.

Contents

List of Figures

List of Tables

List of Acronyms

AI: Artificial Intelligence

API: Application Programming Interface

BERT: Bidirectional Encoder Representations from Transformers

CNE: Comissão Nacional de Eleicões

CRISP-DM: Cross-Industry Standard Process for Data Mining

FN: False Negatives

FP: False Positives

GPT: Generative Pre-training Transformer

HDBSCAN: Hierarchical Density-Based Spatial Clustering of Applications with Noise

HS: Hate Speech

LDA: Latent Dirichlet Allocation

LaBSE: Language-agnostic BERT Sentence Embedding

LLMs: Large Language Models

LSTM: Long short-term memory

MAI: Ministério da Administração Interna

ML: Machine Learning

NER: Named Entity Recognition

NLP: Natural Language Processing

NMF: Non-negative Matrix Factorization

POS: Part-of-Speech Tagging

PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses

SVM: Support Vector Machines

TF-IDF: Term Frequency-Inverse Document Frequency

TN: True Negatives

TP: True Positives

UMAP: Uniform Manifold Approximation and Projection

CHAPTER 1

Introduction

The growth of far-right politics in Portugal reflects a broader trend observed in several European countries in recent years. Factors such as economic crisis, social tensions, and discontent with political elites have created favorable conditions for the emergence of popular disgust and polarized discourse [1], [2]. CHEGA, one of Portugal's most prominent far-right parties, has capitalized on these tensions, consolidating its presence on the national political scene.

Using Natural Language Processing (NLP) and Artificial Intelligence (AI) techniques, this dissertation analyzes CHEGA's political growth and social media landscape, exploring emerging topics in its discourse, and how it unfolds, for example, in details such as the strategies used, the emotions evoked, and the impact of this discourse on Portuguese society. Data from *X* (formerly known as *Twitter*) and election results obtained from official sources will be used, allowing correlations to be established between the evolution of the discourse and the party's political performance between 2019 and 2024.

1.1. Introductory Concepts: Political Discourse and Social Media

The relationship between political discourse and social networks is emerging as an innovative area of study in the fields of political science and computational linguistics. These platforms in modern politics provide a space for political participation and open discourse. However, they facilitate the spread of polarized, often offensive, discourses aimed at minority groups. The presence of these types of speech directed at *Afro-descendant*, *Roma* and *LGBTQ+* communities in Portugal is notable on social platforms such as *X*, despite reformulations of its behavioral policies, or even efforts of the European Union [3], [4]. This discourse is often presented subtly, making it very difficult to identify, presenting itself masked in rhetorical tactics and questions, such as irony, sarcasm, humor, and euphemisms [5].

The role of social networks in the growth of far-right movements around the world is undeniable, allowing the rapid dissemination of ideologies that promote narratives that appeal to emotions and fears about issues such as immigration, national identity, and economic security [6], [7]. By communicating directly with their target audience to attract new members, bypassing traditional media, they often create echo chambers where their messages are amplified while divergent opinions are marginalized [8]. The rise of the far-right has been capitalized on within a context of global discontent with neoliberalism, intensified by financial crises in recent decades and, more recently, by the coronavirus

pandemic. This capitalization effort channels resentment toward global elites and social-democratic parties associated with neoliberalism. The failure of left-wing parties to offer viable alternatives has driven the search for radical answers, which resonate strongly with disillusioned working classes and marginalized communities [9].

Although it varies depending on the context, feelings of abandonment and resentment toward elites and immigrants bolster xenophobic and racist discourses. In Europe, immigration and economic decline in specific regions play a more significant role; however, in the United States, populist support is stronger in counties with a larger white population and declining economic conditions [10]. Conservative media and far-right websites have contributed significantly to the spread of conspiracy theories. Exposure to these media sources not only increases belief in conspiracies but also strengthens these beliefs when confronted with opposing views, further enforcing misinformation. Political ideology influences the acceptance of such theories, with alternative far-right media promoting antagonistic perspectives that exacerbate polarization and public distrust in governance [11]. This public distrust, which has been growing since the 2010s, has led to a normalization of nationalist and anti-democratic discourse infiltrating mainstream media, eroding liberal values. The media thus play a dual role by both carrying far-right messages and, simultaneously, becoming targets of these forces. Although the relationship between the media and far-right ideology is complex, on-line platforms facilitate the spread and normalization of these authoritarian discourses in public spheres, posing a threat to liberal democracy [12].

This growing complexity of narratives and their subtle rhetorical disguises pose significant challenges to researchers in analyzing social media data, detecting these forms of discourse goes beyond simple keyword identification, requiring sensitivity to cultural and linguistic nuances, irony, and implicit messaging [7], [13]. This analytical complexity has driven an increasing dependence on advances in NLP and AI, whose models - such as transformer architectures like Bidirectional Encoder Representations from Transformers (BERT) — can process large amounts of multilingual content and uncover latent patterns [13], [14]. Their application, including adaptations for political texts in languages such as Greek [15] and Portuguese [16], illustrates a growing convergence between computational linguistics and political science, allowing for deeper insight into the dynamics of online discourse.

1.2. Motivation

We live in an era where freedom of expression is widely protected, yet this freedom also fuels increasingly polarizing communication, especially online. Digital platforms have become key arenas for public debate, shaping opinions, and amplifying ideological divides. Studying these interactions helps us understand how people discuss sensitive issues, react to political and social events, and influence public opinion, while also revealing challenges such as harassment, misinformation, and radicalization.

The motivation for this research stems from the need to examine these digital dynamics and their connection to the political rise of the far-right in Portugal, a development that has raised concerns both nationally and internationally [8]. The growing influence of the far-right is often fueled by populist and divisive narratives, strategically deployed to shape public discourse, mobilize support, and reinforce ideological boundaries. Such narratives frequently target marginalized communities, spread misinformation, and foster a us versus them mentality that deepens social polarization. By analyzing these communication patterns and their underlying structures, this research aims to contribute to a more comprehensive understanding of the digital ecosystem and its impact on contemporary political landscapes. Ultimately, the findings will provide valuable information that can promote greater awareness, inform policy discussions, and support initiatives aimed at promoting a more informed and cohesive Portuguese society.

1.3. Research Questions

Political texts can often be polarizing, and identifying this phenomenon, its various forms, and the implications it has in different contexts is a challenging task. However, such challenges can be addressed through NLP and AI approaches. The research questions that follow will be addressed in this work:

RQ1: How do far-right political actors in Portugal use social media to shape, amplify, and sustain their discursive and mobilization strategies, and what thematic and temporal patterns structure this communication, especially in electoral cycles?

RQ2: To what extent are NLP and AI techniques (e.g., text classification, topic modeling) effective in identifying patterns of political discourse in the Portuguese language, and what specific challenges and potential solutions arise when applying these methods?

- Identification of suitable methods: Which NLP/ AI techniques are best suited to analyze Portuguese political text data?
- Measurement of effectiveness: How can we assess the performance of these techniques?
- Challenges and solutions: What linguistic and computational issues are common and how can they be addressed?

1.4. Objectives

The following phases identify the main success objectives of this work:

- Compile a database consisting of X posts (featuring political data on CHEGA key figures), along with corresponding political results for the CHEGA party, between 2019 and 2024.
- Determine the appropriate text analysis techniques, with an emphasis on NLP and AI techniques supporting the Portuguese language, to be able to identify and categorize the discursive patterns and contextual factors within the collected text data.

- Identify and categorize the discursive patterns and other contexts present in the collected texts to understand the context and possible implications of public discourse.
- Analyze the evolution of these discursive patterns and expressions over time, correlating them with political, public interest, and/or electoral data, to understand possible relationships between discourse content (political and social view).
- Develop an artifact that helps to raise awareness about the impact and implications of such discursive patterns in Portuguese society, contributing to academic and social debate.

1.5. Methodology

The Cross-Industry Standard Process for Data Mining (CRISP-DM) methodology was adopted, offering a structured framework that covers the full life cycle of data mining projects [17]. It consists of six phases, which are depicted in Figure 1.1: Business Understanding; Data Understanding; Data Preparation; Modeling; Evaluation; and Deployment.

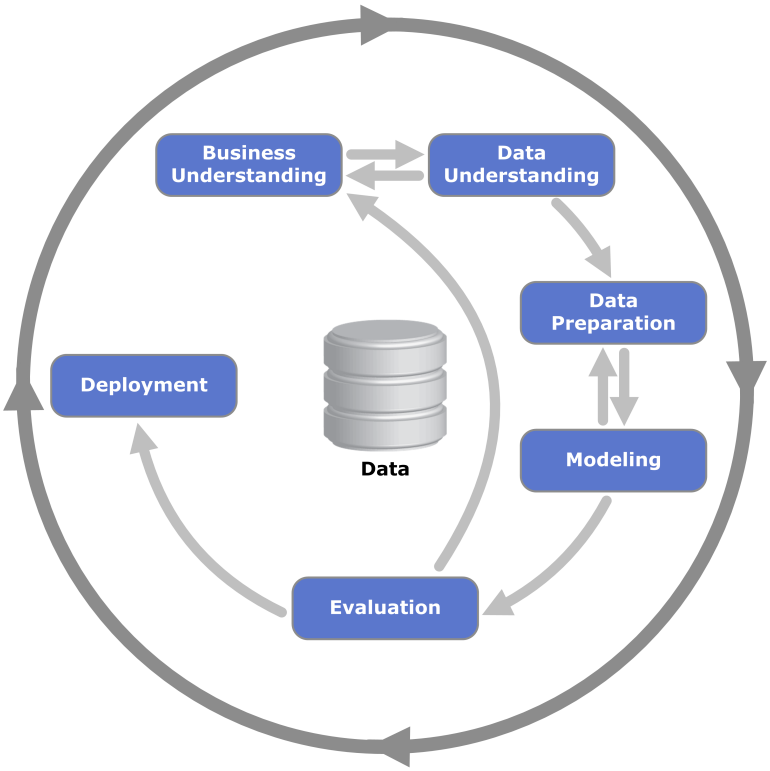


FIGURE 1.1. The data mining life cycle (CRISP-DM as a process model) [18].

Each phase has its own set of tasks and relationships and can be easily customized [17], [18]. The customized outline of each phase along with the work carried out is as follows.

(1) **Business Understanding:** The initial phase involved determining the project objectives, evaluating the current state, identifying the data mining goals, and creating a comprehensive project plan. This stage expanded significantly through

the analysis of sensitive social issues, enriched by academic literature and documentary media.

(2) **Data Understanding:** In this phase, relevant data sources, such as *X*, and elections data from Comissão Nacional de Eleicões (CNE) and Ministério da Administração Interna (MAI) were identified to facilitate analysis. The data from these platforms were subsequently extracted, collected, described, and explored, focusing on contextual relevance and ensuring that the data were appropriate for the project's goals.

(3) **Data Preparation:** The data retrieved from the aforementioned platforms were subjected to cleaning, standardization, and preprocessing procedures. This step was critical in preparing the data for NLP and AI tasks.

(4) **Modeling:** In this phase, appropriate modeling and evaluation techniques were selected. This involved generating a test design and building initial artifacts.

(5) **Evaluation:** The results of the modeling phase are thoroughly evaluated. This included reviewing the process, assessing the effectiveness of the techniques applied, and determining the subsequent steps for refinement and analysis.

(6) **Analysis/Business Understanding Checkpoint:** In this phase, the results of the evaluation were subjected to a deeper analytical process to extract meaningful insights. The analysis focused on interpreting the results, identifying patterns and correlations, and contextualizing them within the project objectives.

(7) **Deployment:** This final phase involves planning the deployment of the insights derived from the analysis and creating the final artifact, along with a comprehensive report of the results.

1.6. Outline of the Dissertation

Following the introduction provided in the current chapter, this dissertation is structured as follows:

- **Chapter 1** provides an overview of the research topic, detailing the motivation behind the study, the research questions, objectives, and the significance of the work. Introduces the methodological approach and outlines the key concepts that guide the dissertation.
- **Chapter 2** provides background on the topics under investigation and presents the state-of-the-art studies using the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) systematic review process. This chapter explores relevant studies on digital political discourse, the rise of the far-right in Portugal, and the application techniques in social and political analysis.
- **Chapter 3** explains the data used to apply the study methodology, including metadata analysis and preprocessing, preparation, and transformation methods. This includes an overview of data sources, data collection procedures, and the steps taken to clean and structure the datasets for analysis, ensuring their suitability for NLP and classification tasks.

- **Chapter 4** presents the complete proof of concept for the proposed solution, covering all stages of development, the application, and the evaluation methodologies. Provides a detailed account of the approaches implemented and the performance evaluation methods used, and the refinements introduced to optimize the results.
- **Chapter 5** focuses on the systematic evaluation of the approaches used and creates the artifact. The results are introduced with comparative analyzes that assess the effectiveness of the solution, identify strengths and weaknesses, and establish the general reliability of the results.
- **Chapter 6** provides a comprehensive interpretation of the evaluated results in the context of the research objectives. Highlights the most significant findings, identifies emerging patterns, and discusses their relevance to the underlying social and political issues studied.
- **Chapter 7** summarizes the key contributions of the research, reflects on the broader impact of the findings, and suggests avenues for future studies to expand on the work conducted. The research questions and objectives stated in the introduction are reviewed, and the work accomplished, and the outcomes are scrutinized. Describes any improvements that may be considered for future work and the conclusions are presented.

CHAPTER 2

Literature Review

This study explores the intersection of political discourse and digital communication, with particular emphasis on the Portuguese context. By examining how NLP and AI can be used to analyze political text data, it offers a new lens through which to understand the modern political discourse of the far-right in Portugal. This section outlines current methodological approaches and their application in various domains, providing the foundation for the design of an artifact that allows a nuanced examination of this type of speech in digital spaces, thus illuminating its broader implications for political debate. By focusing on the underlying dynamics of divisive language, this work aims to demonstrate how such speech can shape public opinion and ultimately redefine the boundaries of political discourse.

A systematic review was conducted following the PRISMA methodology, with a search for articles and conference papers in the Scopus databases, based on the following search queries.

(1) *(("far right" or "extreme right") and ("rise") and pubyear > 2019 and (limit-to(exactkeyword,"far right") or limit-to (exactkeyword,"populism") or limit-to (exactkeyword,"far-right") or limit-to (exactkeyword,"social media") or limit-to (exactkeyword,"hate speech")) and (limit-to (doctype,"ar") or limit-to (doctype,"cp")) and (limit-to (pubstage,"final")) and (limit-to (srctype,"j")) and (limit-to (language,"english") or limit-to (language,"portuguese")) and (limit to (oa,"all"))*

(2) *("text classification"*) and ("polarized speech" or "hate speech") and ("political") and pubyear > 2019*

As seen in Figure 2.1, the vast majority of the 346 results of both queries (First: 229, Second:117) were dated within the last three years, highlighting the relative currentness of this particular area of study. Mendeley Reference Manager was used to organize and store references following the steps illustrated in Figure 2.2. An initial screening was conducted to determine whether the title, abstracts, and keywords of the documents were relevant to the dissertation topic. Then a full-text analysis was performed to determine whether the chosen documents were appropriate for quantitative synthesis, finding references with techniques, strategies, and results that could be reviewed. 60 articles were included in this state-of-the-art review because, during this phase, documents were excluded if they used proprietary software, or redundant approaches.

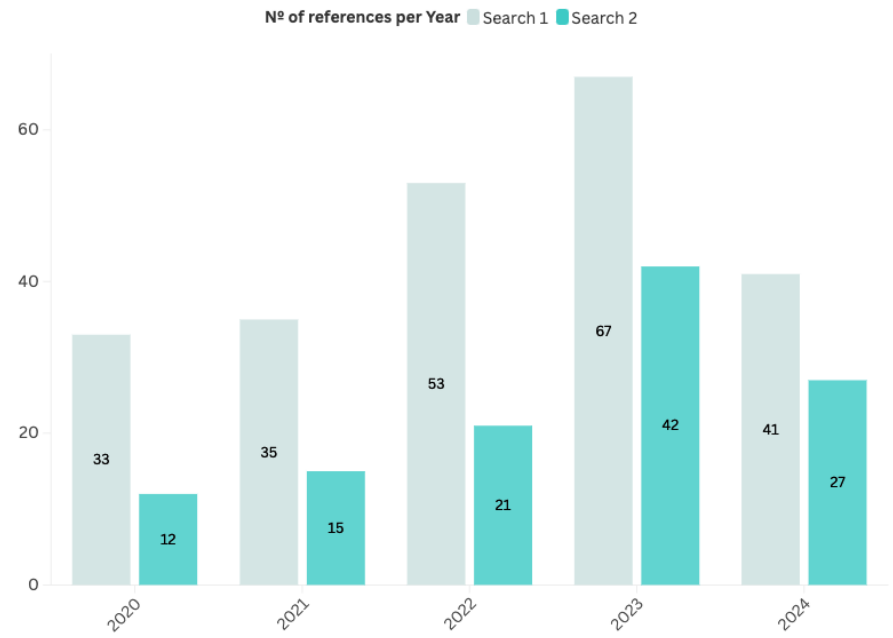


FIGURE 2.1. Number of documents per year retrieved on November 5, 2024.

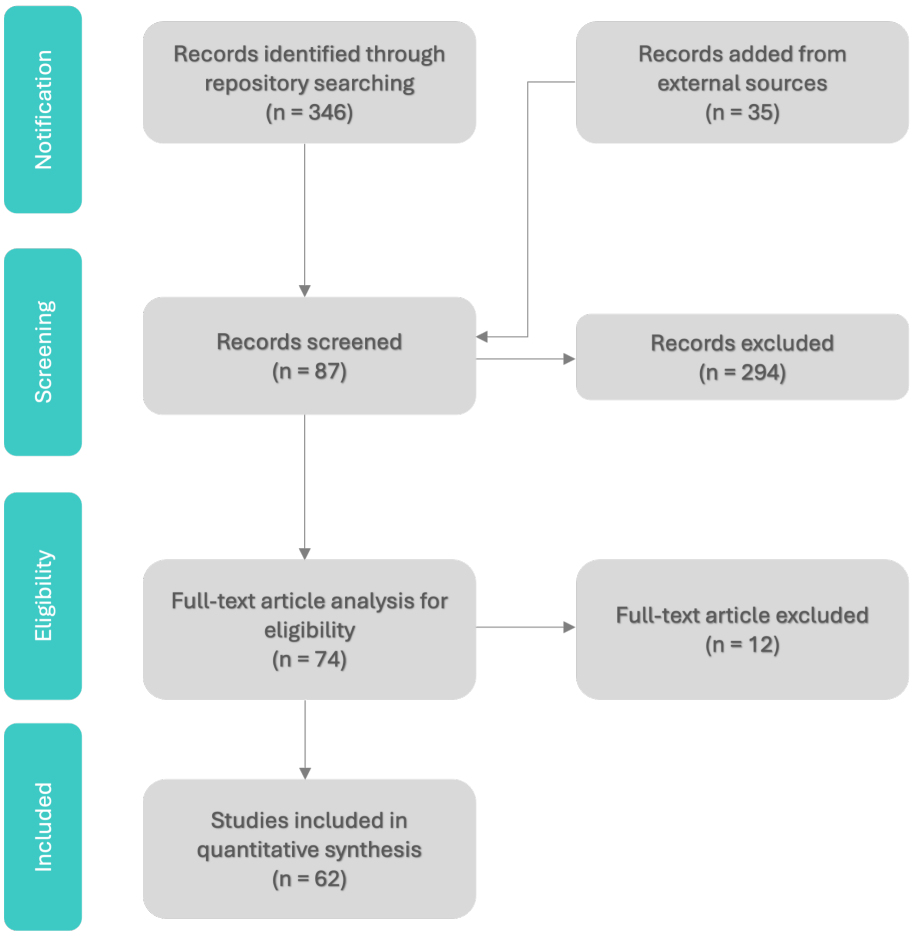


FIGURE 2.2. PRISMA workflow diagram.

2.1. Social Context

The rise of far-right movements in recent decades has generated a substantial interdisciplinary debate, where scholars identify structural and sociocultural drivers, economic pressures, perceived cultural insecurity, and shifting structures of political opportunities while also underscoring how the technical advantages of social media platforms shape mobilization dynamics. Across national settings, far-right figures rely on a remarkably stable discursive repertoire anchored in nationalism, anti-globalism, anti-elitism, anti-immigration, and cultural nativism. These themes coalesce in narratives that portray society as being under attack on multiple fronts: by external enemies such as immigrants or supranational bodies, by internal moral decay, and by betrayal from domestic elites - all presented as simultaneous threats. By positioning themselves as the sole authentic voice of the people, such figures advance xenophobic policy prescriptions while explicitly distancing their programs from fascism and biological racism. Notably, European far-right parties frequently frame the European Union's freedom-of-movement regime as an existential menace, conflating immigration with both economic decline and cultural degradation. Anti-immigration frames therefore remain indispensable for recruiting sectors worried about the growing societal heterogeneity [19]. The logic of populism feeds the view that only the far-right movement or party represents true national interest, generating a concept of discursive opportunities, drawing attention to how certain narratives and frames available in public space favor or make the legitimization of radical nationalist and xenophobic ideas difficult [20], [21]. Represented in a vector of symbolic legitimacy, with an increasing articulation between actors, themes, and channels, even when the main figures do not have direct access to traditional media, coverage of culturally sensitive topics, such as national identity and security, facilitates their penetration into political debate [22].

Portugal shares dynamics with other European countries regarding both political supply and electoral demand. The historical absence of an institutionalized radical right until 2019 appears to reflect not a lack of societal predisposition, but the absence of a credible political offer. The emergence of CHEGA marks the end of this phase of Portuguese exceptionalism, aligning the country with broader European populist trends [23]. The populist landscape in Portugal spans multiple dimensions, from historical figures to the intersection of populism and religion, highlighting ideological divergences between other parties, particularly in the instrumental use of religious references. Media discourse analysis reveals that populism in Portuguese journalism is frequently framed negatively, often associated with demagogy or simplistic rhetoric, and is attributed not only to radical actors like CHEGA, but also to centrist figures such as President Marcelo Rebelo de Sousa. At the attitudinal level, citizens with populist tendencies share traits with abstentionists, including institutional distrust, euroscepticism, and perceptions of identity threat, but differ in their higher political interest and engagement. These attitudes significantly predicted support for CHEGA's leader André Ventura in the 2021 presidential

election, though they were less predictive in legislative elections, where factors such as trust in government played a greater role [23]. Investigations of personality traits of people with populist attitudes reveal correlations with conscientiousness, extroversion, and neuroticism, as well as low openness to experience, suggesting a preference for rigid and immediate solutions in uncertain contexts. Although Portugal has often been overlooked in comparative populism research, its case offers critical insights: populist attitudes can persist without immediate electoral translation, and cultural, institutional, and contextual factors must be considered in tandem [23]. An online survey carried out between May and June 2021 (n = 3,183) offers the first systematic portrait of CHEGA's militants. Predominantly male, middle-class, educated, and politically active; many without any previous party affiliation, they express great dissatisfaction with democracy, without adopting antidemocratic principles. Instead, they support plebiscitary mechanisms that elevate the will of the people, defending a nativist logic, based on traditional customs, relegating religious identity. Although explicit biological racism is uncommon, essentialist views on culture and ethnicity, especially Roma, are widespread. Aligned with their leader André Ventura, their personalist and punitive populism, characterized by an anti-corruption zeal, calls for strong authority, rejection of multiculturalism, and opposition to progressive agendas, emerge as the main drivers of CHEGA's growth, placing the party between protest populism and cultural nationalism [24].

The electoral rise of CHEGA is linked to the confluence of political discourse and weak digital skills among large segments of the Portuguese population. The limited capacity to verify sources exposes citizens to misinformation, while social media algorithms favor sensationalism and polarization of content, reinforcing echo chambers [25]. CHEGA's communication repertoire: fearmongering, stigmatization, anti-elitism, narrative manipulation, achieves high visibility and engagement online. Comparative parallels with Ventura, Bolsonaro, and Trump underscore a shared playbook: manipulation of electoral narratives, strategic use of AI in communication, and selective appropriation of democratic symbols that ultimately erode institutional trust. Therefore, scholars recommend media literacy policies, platform regulation, source accountability, and the reinforcement of professional journalism as countermeasures [25].

2.1.1. Polarized Discourse and the Influence of Social Media

Digital platforms allow activists and supporters to produce content, share opinions, and interact directly with political figures, without the intervention of traditional news outlets [26], [27]. Thus, certain extremist groups can spread xenophobic discourse and conspiracy theories virally, creating an environment in which hate speech finds receptivity and normalization [21], [28]. Social networks not only reinforce exclusionary discursive patterns, but also serve to organize far-right communities, facilitating the formation of ideological bubbles. In these spaces, other participants validate xenophobic, racist, or conspiratorial arguments, generating a spiral that intensifies radicalization. This process of echo chambers can reinforce group identities and feelings of cohesion, while at the same

time antagonizing those who are seen as different. Such mechanisms can have offline consequences, whether in the form of street protests or the adoption of some sort of violence [21], [26]. Polarized discourse has become a defining feature of the political and social landscape in recent decades, especially on digital platforms [3], [4], [6], [7]. As social networks consolidate as primary communication platforms [29], the need to classify and interpret volumes of textual data has increased, especially in scenarios involving polarized discourse and extremist behaviors.

2.1.2. Polarized Discourse Detection in the Portuguese Context

In Portugal, the detection of Hate Speech (HS) in political and social discourse has gained relevance due to the growing prevalence of discriminatory narratives targeting historically marginalized groups. Research has focused particularly on *Afro-descendant*, *Roma*, and *LGBTQI+* communities, the most reported online HS targets in both European and national contexts [3], [5]. A significant contribution is the *FIGHT* corpus, which compiles 63,450 geolocated tweets in Portugal, posted between 1.5 years before and after the official declaration of *COVID-19* as pandemic by the *World Health Organization*, including public and deleted messages [3]. This dataset was designed to analyze HS dynamics before and after this pandemic, allowing temporal and spatial monitoring of offensive discourse. Addresses the limitations of earlier Portuguese HS corpora, which relied heavily on lexical approaches based on closed lists of offensive terms, often failing to capture covert forms of HS such as irony, sarcasm, humor, and metaphor. The initial manual annotation trial indicated that 40% of tweets automatically flagged as potentially offensive or hateful were confirmed as such, highlighting both the utility and the challenges of automated detection. Through focus groups with representatives of *Afro-descendant*, *Roma* and *LGBTQ+* groups, [5] revealed that covert HS, often disguised as compliments or humor, can be more harmful than overt attacks, as it normalizes prejudice and limits the ability of victims to respond. This aligns with the larger findings that detecting covert HS requires moving beyond isolated word analysis to incorporate contextual, pragmatic, and sociocultural dimensions. Methodologically, multiple works have investigated transfer learning and semi-supervised approaches for Portuguese HS detection [30]. *BERT*-based architectures (*BERT-LinearLayer, BERT-CNN, GAN-BERT*) were evaluated using the *CO-HATE* [31] *YouTube* comments corpus, explicitly distinguishing between overt and covert HS. The results showed that annotator subjectivity significantly impacts model performance and that preprocessing had limited effect on detection rates. In a complementary study, [14] proposed an ensemble of *GAN-BERT* and label propagation methods to semiautomatically annotate HS in *FIGHT*, achieving an *F1 score* of 0.66 for the HS class. This semi-supervised strategy reduced the dependency on large manually annotated datasets and proved effective in expanding Portuguese HS resources.

Recent advances in the application of state-of-the-art transformer architectures and generative Large Language Models (LLMs) for Portuguese HS detection [16] include a comprehensive comparative evaluation of models such as *BERTimbau*, *mDeBERTa*, and

generative systems from the Generative Pre-training Transformer (GPT), *Gemini*, and *Mistral* families, tested on two annotated corpora — one composed of *YouTube* comments and the other of tweets — carried out within the scope of the *kNOwHATE: kNOwing online HATE speech* project[1]. For the *YouTube* dataset, the best result was achieved by a *BERTimbau* variant retrained with European Portuguese tweets and fine-tuned for the HS task, reaching an *F1 score* of 0.87 for the positive class, more than 20% above baseline models and 1.8% higher than the base *BERTimbau*. For the *Twitter* corpus, GPT-3.5 obtained the best result, with an *F1 score* of 0.50 for the positive class. The study also assessed the impact of training in domain versus mixed domain, as well as the effect of adding contextual information to the generative model prompts, confirming that domain adaptation and prompt design substantially influence performance.

In general, Portuguese research on HS detection demonstrates a convergence of sociolinguistic insight and computational innovation. By combining qualitative community-based studies, targeted corpus construction, and advanced Machine Learning (ML)/ LLMs techniques, recent work has advanced the capacity to identify both overt and covert forms of polarized discourse in the national context. However, issues of class imbalance, annotation subjectivity, and contextual inference remain central obstacles for future research.

2.2. Computational Methods for Political Discourse Analysis

Text classification plays a key role in the analysis of political discourse by enabling the identification of underlying linguistic and thematic patterns through the use of ML techniques and deep learning models, which have proven effective in categorizing and interpreting political texts while overcoming the limitations of traditional rule-based methods. However, significant challenges remain, particularly in capturing semantic nuances, dealing with unbalanced datasets, and requiring substantial volumes of annotated data.

The increasing availability of short digital texts, such as news or psychosocial content, demands efficient and accurate methods for their automatic classification, recent research highlights the advantages of deep learning techniques over traditional ML approaches, particularly in the handling of short texts with greater accuracy and generalization capabilities [32]. Furthermore, hybrid approaches that combine both paradigms have shown promise in multilabel classification tasks, improving the organization and analysis of large-scale textual datasets [33]. Automatic classification of textual data, especially short texts or content with high linguistic complexity, faces several structural and methodological challenges. Among the most pressing issues are the limitations of traditional algorithms based on manual feature engineering (e.g., Term Frequency-Inverse Document Frequency (TF-IDF)), the imbalance between class distributions, and the high demand for annotated data. These factors can hinder generalization, reduce performance for underrepresented categories, and restrict scalability in domains such as political or news analysis [32], [33].

[1] `https://knowhate.eu/`

To mitigate these limitations, Transformer-based models have been increasingly adopted because of their ability to capture complex semantic and syntactic patterns without requiring extensive preprocessing. This type of approach, such as BERT, shows superior performance in short text classification tasks, outperforming classical models in terms of accuracy and generalization [32], [33]. However, their application is not without cost—Transformer architectures typically require substantial computational resources. In response, strategies such as *Random OverSampling* have been implemented to balance training data, dimensionality reduction, and inference optimization, ensuring that model performance is maintained even in scenarios with limited infrastructure. Furthermore, hybrid pipelines that combine contextual embeddings from deep learning models with traditional classifiers contribute to more robust and interpretable outcomes, particularly in multilabel scenarios [33].

In the analysis of LLMs in the automatic classification of the affiliation of a party from political texts [34] experiments were conducted on three different datasets containing parliamentary speeches from Norway, Germany and the UK. For each, traditional models (Naïve Bayes), transformer-based models (such as BERT), and adjusted versions of LLMs were tested. The analysis revealed that language models outperformed Naïve Bayes by up to 10.35% (Norwegian), 12.95% (German) and 6.39% (English), demonstrating their superiority in identifying the political affiliation of texts. However, challenges persisted, it was observed that some classes, especially parties with less representation in the data, underperformed, even with the most advanced models. This indicates that class imbalance affects the accuracy of the models, favoring the majority parties. To mitigate this problem, class weights and model refinement were explored by making adjustments to training data, resulting in improvements in overall accuracy. The results suggest that language-specific models offer superior performance compared to multilingual models, as evidenced by the greater accuracy of the models adjusted for Norwegian and German. In addition, they reinforce the need for balanced and specific datasets for political analysis, as the quality and representativeness of the data directly impacts the performance of the classifiers.

Detecting hateful or polarized speech and disinformation on social networks has become a priority for researchers and developers, given the social and political consequences associated with the spread of harmful and polarizing content. Social networks facilitate the rapid spread of information, including potentially harmful speech, which can amplify prejudices and stigmatize specific groups. To address these challenges, automated approaches based on deep learning and NLP have been developed, using robust models and transfer learning techniques to identify and classify different forms of problematic speech. Different methodologies are applied to detect hate speech on digital platforms such as *Twitter(X)* and *Facebook*, highlighting the complexity of this task in multilingual and multicultural contexts. Comparative experiments using models such as BERT, Long

short-term memory (LSTM)-based neural networks, and traditional Support Vector Machines (SVM) classifiers show distinct strengths: while SVM and LSTM approaches perform well in identifying surface-level patterns and sequential structures, transfer learning models such as BERT excel in capturing contextual and semantic nuances, particularly in highly diverse linguistic environments [35]. Through a systematic review of 64 studies [36] on the application of NLP in the analysis of extremism, the field has been mapped in terms of techniques, tools, datasets, and methodological approaches used to describe and detect extremist discourse. The reviewed literature is organized into five analytical dimensions: research topics, techniques employed, empirical applications, available software tools, and accessible datasets. The theoretical framework carefully distinguishes between extremism and radicalization, defining extremism as an antidemocratic ideological movement that may or may not involve violence, while radicalization is conceptualized as a psychological process of detachment from democratic norms. The operational definition of extremist discourse is articulated through five dominant narrative types: political, historical, sociopsychological, instrumental, and theological-moral. These narratives are often accompanied by discursive elements such as hate speech, war metaphors, and dehumanization strategies.

From a technical point of view, the studies reveal a predominance of classical NLP techniques such as TF-IDF, n-grams, and sentiment analysis with feature extraction techniques for syntactic and semantic analysis frequently include Part-of-Speech Tagging (POS), Named Entity Recognition (NER), and topic modeling with Latent Dirichlet Allocation (LDA). However, there is a growing adoption of deep learning models, showing particular promise in distinguishing extremist discourse, especially when combined with deep learning algorithms. Among the challenges identified are the limited explainability of deep learning models, the complexity of handling multilingual and code-mixed texts, and the lack of publicly available annotated datasets. However, NLP emerges as a robust and rapidly evolving set of tools to detect and characterize extremist narratives, with significant applications in both academic research and the development of public policies and preventive strategies against online radicalization.

2.2.1. Automatic Topic Discovery

Topic modeling has been instrumental in identifying discourse patterns in large volumes of text and has historically been dominated by methods such as LDA that marked a turning point in probabilistic topic modeling for textual corpora by proposing a generative approach grounded in hierarchical *Bayesian* principles [37]. Such approaches are based on bag-of-words representations, which limits their ability to capture semantic and contextual relationships between words. When applied to data from social media platforms, the technique must account for the challenges inherent in short, fragmented, and noisy content. LDA remains one of the most widely used probabilistic models for this task; however, its performance is limited in contexts involving microtexts such as tweets due to the sparsity and lack of contextual co-occurrence [38].

Recent studies comparing LDA, Non-negative Matrix Factorization (NMF), *Top2Vec*, and *BERTopic* in 31,800 tweets about travel during the *COVID-19* pandemic confirm the well-known obstacles to topic modeling in microtexts: brevity, lexical noise, and heterogeneity. The methodological pipeline included standard preprocessing steps (punctuation and stopword removal, lemmatization, and normalization). This type of common preprocessing techniques, including others such as textual noise reduction (e.g., eliminating hashtags, *URLs*, and slang), lower-casing, are critical to ensure data quality [39]. Furthermore, term weighting methods such as TF-IDF are often applied to optimize word relevance and topic representativeness [36], [38]. While LDA tends to generate generic and overlapping topics [38], [40], NMF improves thematic coherence but remains limited to TF-IDF representations. Embedding-based models, in particular *BERTopic*, overcome these barriers by producing semantically more cohesive, stable, and easily visualizable topics between domains [41], offering modern features that make it scalable and flexible for large, short, or polarized corpora, such as social networks and political discourse studies. However, they are limited to assigning one dominant topic per document and can generate many topics and outliers that require manual inspection.

In a large-scale application, *BERTopic* was applied to a corpus of more than 750,000 tweets about the German federal elections of 2021 [42]. The key topics identified included *COVID-19*, climate policy, tax policy, digitalization, anti-Semitism, gender equality, and the legalization of cannabis. Although some themes, such as the pandemic and digitalization, were common in all subgroups, others, such as humor and cannabis, appeared more prominently in public mentions. Specific global and national events, such as the Taliban's resurgence in August 2021 and debates surrounding Israel, were reflected in temporary spikes in topic-related activity. Sentiment analysis revealed that discussions around the pandemic and immigration were more frequently associated with negative sentiment, especially in tweets from official accounts. Temporal analysis of topic distributions reveals that certain themes gain prominence during specific time periods, reflecting changes in public attention and discourse engagement. These findings suggest that the integration of topic modeling with temporally-aware and linguistically-informed preprocessing pipelines provides a robust foundation for analyzing discursive trends in political social media data. In the same study, it is noted that although *BERTopic* was found to be effective in capturing latent themes, some limitations were observed, including challenges in interpreting topics with low coherence and the restriction of limiting the generation of topics, which may have excluded less frequent but significant content. The approach proved effective in extracting complex discursive patterns in electoral contexts, demonstrating the usefulness of *BERTopic* as a tool for computational investigation of political communication on digital platforms. The integration of multimodal data, temporal analysis, and sentiment analysis is the key to capture the dynamics and diversity of online political discourse [42].

2.2.2. Advances in Automatic Text Classification

Advances in NLP enable text classification without task-specific fine-tuning or large annotated corpora. Approaches like zero-shot or few-shot classification use large pre-trained language models to assign texts to predefined categories, often guided by natural language prompts. The effectiveness of such methods strongly depends on prompt design. These models were initially celebrated for their ability to translate task instructions expressed in natural language into predictive behavior. However, systematic evaluations suggest that model performance remains robust even when the prompts are irrelevant or misleading, asking if the improvements stem from real instruction understanding rather than statistical correlations between input and label space [43]. This highlights a critical methodological limitation: while LLMs achieve impressive results in zero-shot and few-shot setups, their sensitivity lies more in the selection of label words than in the semantic content of the instructions.

The potential of this type of approach has been especially relevant for abusive language and hate speech detection, domains where annotated corpora are scarce and context-dependent. Prompting LLMs can achieve results comparable to, and sometimes exceeding, supervised baselines in eight benchmark datasets in multiple languages [44]. The decisive role of the choice of verbalizers (for example, using *'hateful'* versus *'non-hateful'*) in maximizing performance, shows that instruction-tuned models such as *FLAN-T5* outperform traditional encoders such as *BERT* and *RoBERTa* in zero-shot settings [44].

Subsequent research further confirms the strengths and weaknesses of zero-shot classification in social media contexts. Benchmarking GPT-3.5 and GPT-4 on *Twitter* datasets for hate speech, offensive language and emotion detection shows that although these models may not consistently surpass fine-tuned systems, they offer competitive performance, particularly in scenarios where training and test data are misaligned in time or domain, a common issue in dynamic online discourse [45]. Similarly, zero-shot and few-shot learning approaches enables LLMs such as *LLaMA-3* and GPT-4, to detect implicit abuse, such as euphemistic, comparative or identity-based forms of hostility, achieving near-human performance on subtle linguistic phenomena often missed by supervised classifiers [46]. However, these advances do not eliminate fundamental challenges. Zero-shot prompting remains highly sensitive to prompt design, performance drops in adversarial or noisy contexts, and the detection of subtle abuse, particularly euphemisms and implicit comparisons, continues to be error-prone. Since even human annotators struggle with consistent labeling of such cases, the reliability and transparency of LLMs driven classifications remain limited, highlighting the need for a careful interpretation of results in applied research [46].

CHAPTER 3

Data

Understanding the data is crucial in any data mining process, as it forms the foundation for meaningful analysis and results. This chapter provides a comprehensive description of the data used in this dissertation, detailing its sources, characteristics, and methods used for its collection. In addition, it outlines the logic behind the choices made during the pre-processing of the datasets, ensuring their alignment with the research objectives. The workflow for processing these datasets encompasses multiple phases, using various platforms and mechanisms to extract, transform, and organize the information into analyzable formats.

3.1. Data Collection

The datasets cover multiple dimensions. A total of 11,003 posts were extracted from the *X* platform between late 2019 and mid-2024, focusing on posts from prominent figures of the CHEGA Party, including its leader, totaling 10,323 unique posts. The selection of these figures was based on their position within the official list of members of the party, as published on its website[1], as well as their number of followers, posting activity and the longevity of the account. Table 3.1 presents sample data prepared in both textual and tabular formats.

TABLE 3.1. Sample data extracted from *X API*.

Content
"RT @PartidoCHEGA: A lua hoje está quase do tamanho da dívida pública. #chega #desocialismo" r
"RT @PartidoCHEGA: Acabou o malabarismo, afinal o crescimento andava por aí, mas onde estão os frutos agora? Portugal está farto disto, a di..."

Furthermore, the party's political performance data by district was manually collected from officials at the Portuguese government site for the same period, with 2019 marking the party's foundation. An architecture was implemented to efficiently collect and process *X's* data based on a clear separation between the support functions and the main data extraction flow, ensuring modularity and ease of maintenance. Configuration and authentication were centralized, with credentials accessed via an external configuration file, ensuring security and preventing the exposure of sensitive information in the code. Communication with Application Programming Interface (API) focused on two main functionalities: querying by username (being the main feature used) or searching based on keywords and expressions. The data collected was then transformed, mapped,

[1] `https://partidochega.pt/index.php/orgaos-nacionais`

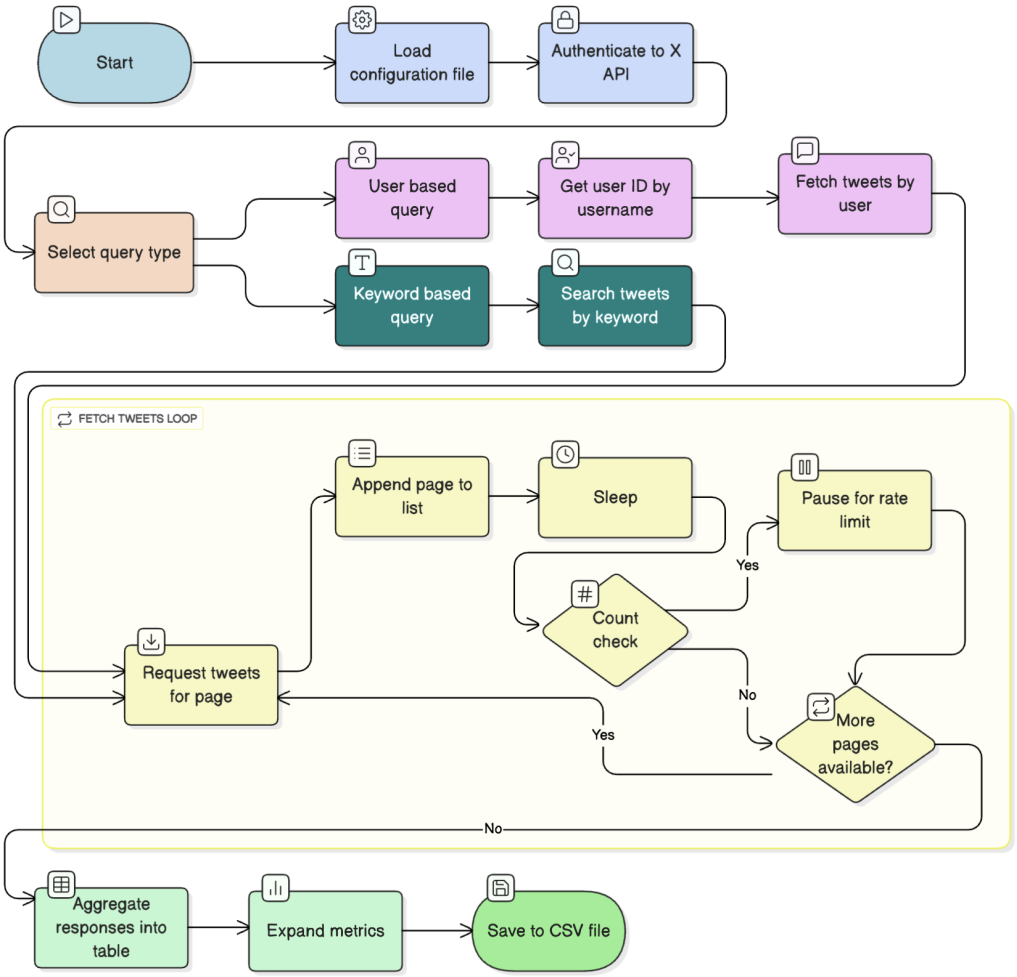


FIGURE 3.1. Data collection diagram.

and stored in tabular structures, facilitating subsequent analysis. Finally, the data are transformed, organized, and stored in tabular structures with two dimensions: posts and their respective comments. Figure 3.1 illustrates the implementation flow diagram.

3.2. Data Preprocessing

As discussed in Chapter 2, the preprocessing stage is essential when dealing with social network texts, as informality and network specific mention formats introduce additional problems for analysis and model refinement [36], [47]. The data were prepared in both textual and tabular formats. Although unique identifiers that could identify the author are included in the dataset, these were not utilized for the tasks carried out in this study. To better understand the content of the collected data, we performed a frequency analysis of the most common words within the datasets. Interestingly, functional words such as articles and prepositions *('de', 'a', 'que', 'e', 'o')* dominate both datasets. These are typical in any Portuguese text corpus, reflecting the language's inherent structure. However, such terms carry limited semantic value for content-based analysis. Figure 3.2 illustrates the temporal distribution of the data. Initially, from 2019 to early 2022, there was a relatively low level of activity, with sporadic spikes in the number of posts. From

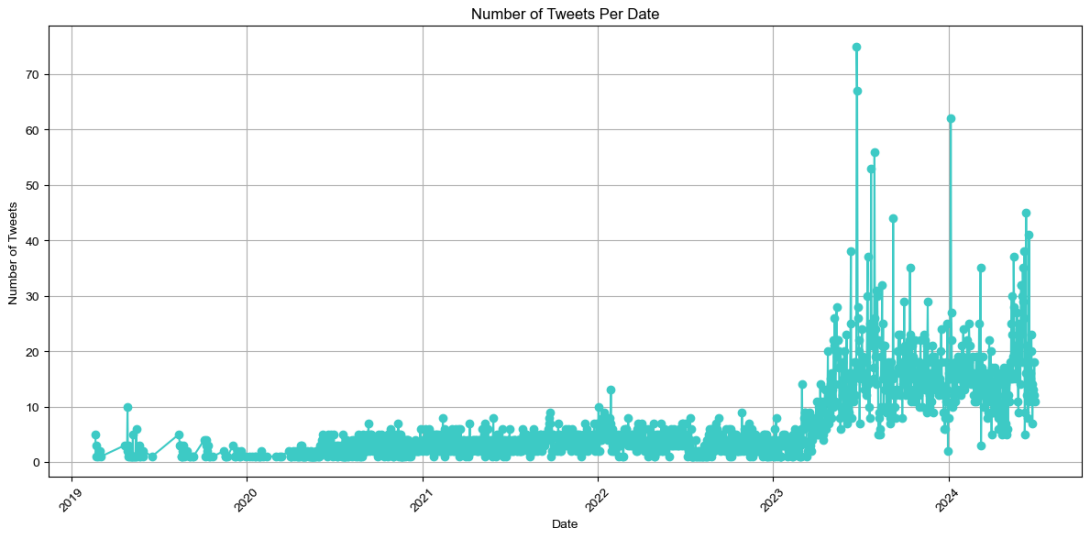


FIGURE 3.2. Number of X's posts from main party figures.

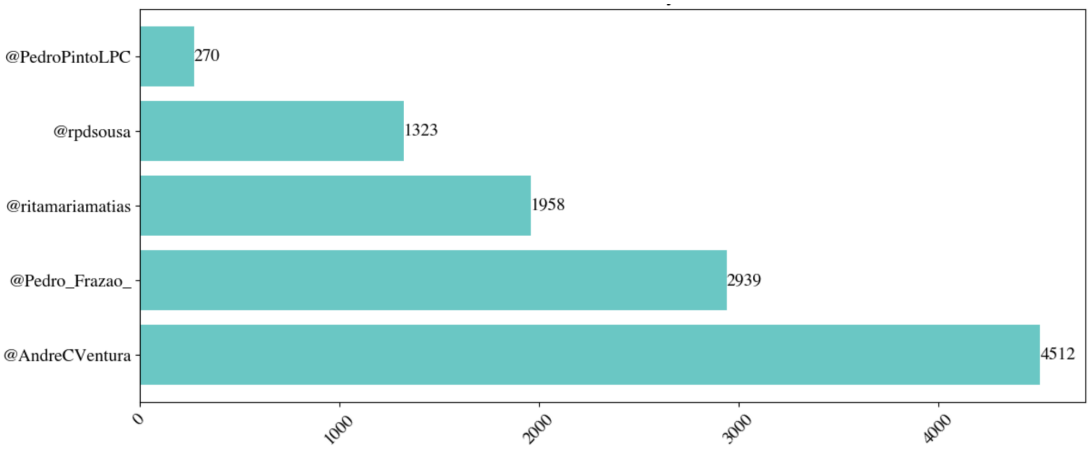


FIGURE 3.3. Number of tweets extracted from the account of each of CHEGA's main politicians.

mid-2022, there is a noticeable increase in the volume of tweets. This trend intensifies significantly throughout 2023 and into 2024, with consistent peaks reaching more than 20 tweets per day. Figure 3.3 illustrates the volume of tweets collected from the accounts of the main figures of the party. A significant disparity is evident, with party leader André V. accounting for the largest number of publications, followed by Pedro F. and Rita M.

For an initial exploration, prototypes were developed using *Orange Software* and *Python*, applying different preprocessing strategies for Portuguese text. These ranged from stemming and lemmatization to frequency-based filtering and enhanced approaches that combine refined libraries and models. These preprocessing pipelines were designed to identify the most effective configuration for corpus analysis, balancing linguistic normalization, noise reduction, and meaningful token retention, obtaining the following results visible in Table 3.2.

The stemming approach produces tokens by reducing words to their root forms using a heuristic process. For example, the word *'hoje'* becomes *'hoj'* and *'tamanho'* is reduced

TABLE 3.2. Comparison of tokenization approaches in orange software.

Approach	**Tokens and Tags**
Stemming	`rt, partidocheg, lua, hoj, quas, tamanh, div, public, cheg, desocial`
Lemmatization	`rt, partidochega, lua, hoje, quase, tamanho, dívida, pública, chega, desocialismo`
Frequent Tokens Filtering	`rt, partidochega, hoje, chega`
Enhanced Approach	`rt, partidochego, lua, hoje, quase, tamanho, dívida, público, chegar, desocialismo`

to *'tamanh'*. While stemming is computationally efficient and reduces the vocabulary size significantly, it often results in incomplete or non-standard tokens that may lose semantic clarity. As a result, while stemming can be advantageous for certain tasks such as search indexing or simple frequency analysis, it may not capture the nuances of language, which is critical for complex tasks like sentiment analysis or semantic understanding. In contrast, the lemmatization approach takes a more linguistically informed route, returning words to their base forms while maintaining grammatical accuracy. For example, *'hoje'* remains *'hoje'* and *'tamanho'* stays *'tamanho'* This approach retains more semantic meaning compared to stemming and is especially useful for analyses that require linguistic precision, such as named entity recognition or syntactic parsing. However, lemmatization is typically more computationally expensive and may not reduce vocabulary size as significantly as stemming. The frequent token filtering approach takes a different path by focusing only on the most frequent and relevant tokens. This method significantly reduces the number of tokens, as seen in the output where only terms like *'rt', 'partidochega', 'hoje' and 'chega'* remain. This reduction is ideal for tasks that prioritize interpretability and focus on the most impactful terms, such as topic modeling or keyword extraction. However, this approach sacrifices a substantial amount of contextual information, which may limit its applicability for tasks that require broader semantic understanding. Finally, the Enhanced approach refines the 'Lemmatization' approach, using more robust libraries.

In incorporating *Python* for more programmatic exploration, it was possible to process the data in a more mature way, understanding the nature of the information after exploring with the *Orange software*. From this point on, the use of lemma or stemming was abandoned, as it is not the intention to reduce the vocabulary, it is verified that the political text carries different connotations, and there is a need for this study to capture contextual nuances. In this understanding, it was possible to manually check the data for the representation of entities that appear in different variations, as can be seen in Table 3.3. To ensure the consistency of these entity representations, a normalization process based on NER techniques was implemented. Specifically, all variants of the same person or institution were converted into a common identifier (for example, 'andre_ventura' or 'partido _chega'), chosen because they were the most frequent in the data, thus facilitating subsequent tasks. This type of treatment is essential in social network texts, where

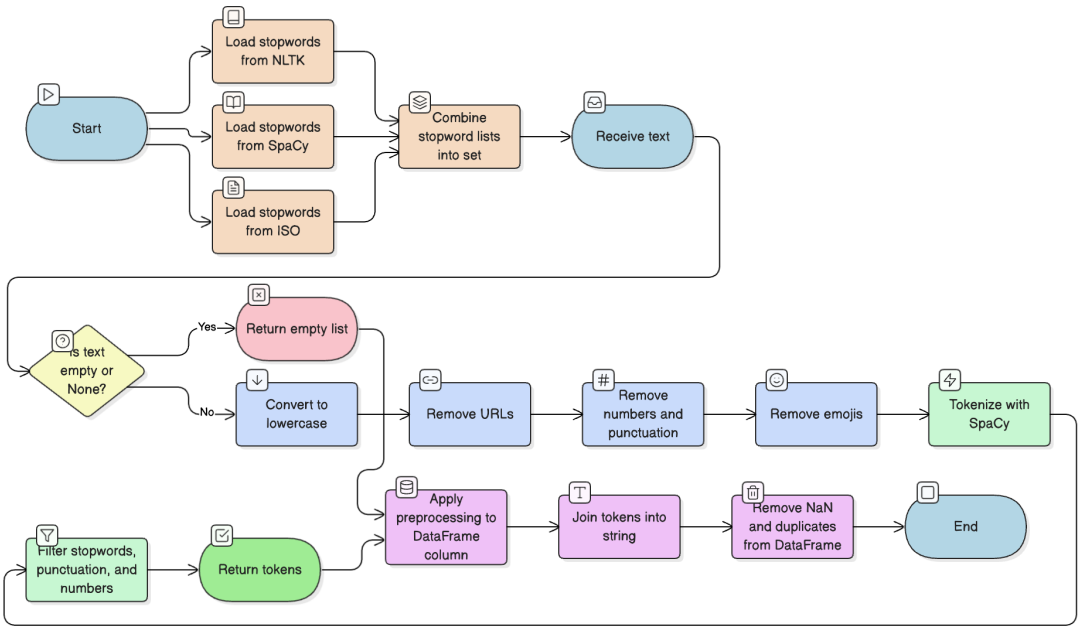


FIGURE 3.4. Preprocessing implementation steps.

informality and multiplicity of form of mention can make analysis difficult, also helping with model refinement [47].

TABLE 3.3. Substitutions of singular and pair terms.

Original Term	**Substitution**
'andréventura' OR 'andreventura' OR 'venturas' OR 'ventura' OR '@andrecventura' OR 'andré ventura' OR 'andre ventura'	andre_ventura
'@partidochega' OR 'partidochega' OR 'chega' OR 'grupoparlamentarchega' OR 'partido chega'	partido_chega
'costa' OR 'antonio costa' OR 'António costa'	antonio_costa
'montenegro'	luis_montenegro
'rocha' OR 'rui rocha'	rui_rocha
'cotrim' OR 'joão cotrim' OR 'joao cotrim'	joao_cotrim
'matias' OR 'rita matias'	rita_matias
'pns' OR 'pedro nuno santos'	pedro_nuno_santos
'ana gomes'	ana_gomes

A series of other steps, shown in Figure 3.4, were implemented to ensure the standardization and consistency of the collected texts. Generic mentions of user accounts were removed and/or replaced, the text is then cleaned of URLs, punctuation and special characters, as well as emojis that could interfere with token recognition. Next, the text is tokenized, simultaneously filtering out stopwords that have been aggregated from different sources, to eliminate terms without semantic relevance. After this process, any empty texts are removed, and finally duplicates are eliminated to avoid redundancies in subsequent tasks. It is important to note that while these preprocessing modifications were applied, both the original and the preprocessed texts were retained and used in the subsequent analytical steps.

CHAPTER 4

Modeling

Modeling is a fundamental phase of this study, as it seeks to extract useful information from the corpus, taking advantage of NLP and AI techniques. This section elaborates on the approaches used to classify the data into relevant categories, such as feelings, topical groups, or types of discourse, thus responding to the research objectives described above. We focus on the classification experiments performed on all key datasets, from the original to the preprocessed data. Figure 4.1 illustrates the workflow used in this study. This workflow served as the foundation for the project, which specifically focused on topic modeling and zero-shot classification. It integrates traditional NLP approaches alongside modern methodologies, leveraging transformer-based models.

4.1. Topic Modeling Methodology

To identify emerging topics, a topic modeling pipeline based on the *BERTopic* model was implemented, adapted to the linguistic and thematic context of the study. A TF-IDF vectorization technique was applied using 5,000 features, designed to reduce dimensionality and mitigate the impact of sparse terms while preserving a rich representation of textual data. The frequency thresholds were empirically defined, with a minimum document frequency of 1% and a maximum of 85%. These thresholds served two purposes: first, to exclude rare and potentially irrelevant terms that appear in only a small number of documents, thus reducing noise and improving topic stability; and second, to remove overly frequent terms (similar to contextual stopwords) that do not contribute meaningfully to

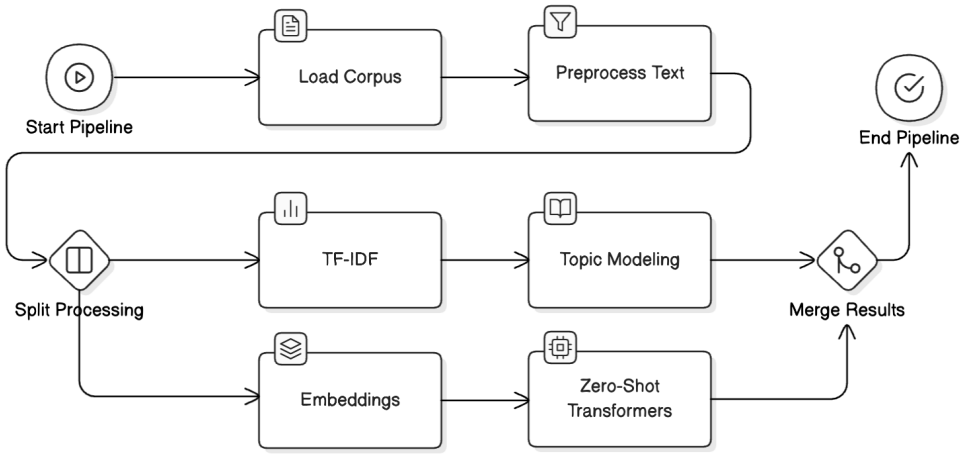


FIGURE 4.1. Pipeline for automatic topic discovery and text classification.

semantic differentiation across documents. This process ensured that the resulting feature space emphasized terms with greater topical relevance and discriminative value.

A customized list of stopwords was incorporated to remove common and politically irrelevant tokens, including elements specific to the social media platform, such as *'rt'* and *'id'*. These tokens were identified through multiple iterations of the model output. The pipeline was applied to both the preprocessed data and the raw data. Text embeddings were generated using the multilingual model Language-agnostic BERT Sentence Embedding (LaBSE) [48], which was trained to capture semantic similarity between languages. We opted for this model as, in preliminary experiments, it led to the generation of more consistent and interpretable topics than the multilingual Sentence *BERT* models [49] and the Portuguese-specific models of the *'Serafim'* family [50]. Dimensionality reduction was performed using Uniform Manifold Approximation and Projection (UMAP) with cosine distance and custom configurations for `random_state=30, n_neighbors=5`, and `min_dist=0.03`, ensuring replicability [51]. This step aimed to improve the separation of latent topics in the embedded space. Clustering was performed using the Hierarchical Density-Based Spatial Clustering of Applications with Noise (HDBSCAN) algorithm, configured for increased sensitivity to local point density (`min_cluster_size=50` and `min_samples=3`) [52]–[54].

In a context where data density varies, the adjustment and definition of these parameters were experimental to balance granularity and thematic sensitivity. This enabled the formation of more sparse clusters and significant semantic heterogeneity between documents. These parameter choices were not fixed a priori; instead, they emerged from multiple observable iterations of the full pipeline. Throughout this process, model output was repeatedly examined, compared, and refined, resulting in a manual assessment phase in which decisions were guided by direct inspection and sensitivity to the evolving structure of the topics. The final configuration therefore reflects an iterative, empirically grounded judgment of what produced the most coherent and analytically meaningful results.

To perform a solid analysis based on the topics identified by a topic model, it is important to assess the intrinsic quality of the model. For this purpose, we rely on two metrics. Semantic topic coherence is a fundamental step in validating topic models, particularly when the goal is to ensure that the generated topics are interpretable by humans. One of the most widely used metrics for this purpose is c_v coherence [55], which measures how well the words of a topic fit semantically by comparing their large-scale co-occurrence patterns with a vector metric, offering an efficient compromise between performance and interpretability. It has proven to be robust even when working with short texts or when topics are generated using modern approaches such as *BERTopic*. Topic models with higher coherence scores tend to be more semantically interpretable and meaningful. Beyond semantic coherence, topic diversity is another crucial dimension for evaluating topic model quality. Topic diversity quantifies the lexical overlap between

topics and is defined as the proportion of unique words across the top-n terms from all generated topics [56], [57]. Values near zero indicate significant topic redundancy, while values approaching one suggest a greater variety and distinction among topics.

4.2. Automatic Text Classification Methodology

To complement the exploratory analysis of the topics, a zero-shot prompt-based classification strategy was developed with the aim of obtaining a more extensive and interpretable characterization of the texts. Although topic modeling allowed for the identification of latent regularities, zero-shot classification provided an explicit semantic framework, associating each document with previously defined categories. The prompts were constructed based on a question–statement pair logic, in which each text was submitted to a LLMs and the answer was restricted to the options *'Yes'* or *'No'*. The classification schema covered multiple analytical dimensions central to this study, based on the guidelines of the project *kNOwHATE: kNOwing online HATE speech*[1] annotation scheme[2] [13], [16], [58], [59].

For each label, a short definition was provided to the model to reduce ambiguity and anchor the prompt in precise analytical terms. An example of such structure is presented below:

(1) *Question: 'Does this text mention or focus on the target group migrants?'*
(2) *Definition: 'Discourse of hate directed at individuals based on nationality or migratory origin, often associated with xenophobia.'*
(3) *Text for analysis: '...'*

The classification was executed across six state-of-the-art LLMs available through the OpenRouter API: *Mistral-7B-Instruct-v0.3*, *Mistral-Nemo*, *LLaMA-3.1-8B-Instruct*, *LLaMA-3.2-3B-Instruct*, *Gemini 2.0 Flash Lite*, and *Phi-4 Multimodal Instruct*, chosen based on the trade-off between capacity, cost and model author rank. Each text in the evaluation dataset was iteratively submitted to all configurations, with responses normalized to binary values (1 = *'Yes'*, 0 = *'No'*). Invalid or ambiguous outputs were coded as `-1`. A retry-and-backoff mechanism ensured robustness against API errors, while rate-limiting strategies (e.g., enforced pauses every 150 requests) avoided overload. The outputs for each label and model were stored systematically in structured files, allowing subsequent evaluation and comparison between models and categories.

The evaluation of the zero-shot classification experiments followed a supervised validation approach, in which the model predictions were systematically compared against the results of the gold annotated standard corpus provided by the *kNOwHATE* project. For each category label in the annotation scheme, a binary evaluation was performed with the following metrics: *Accuracy*, *Precision*, *Recall*, and *F1* Score. These measures allow for the assessment of the overall correctness of the predictions and their ability to

[1] `https://knowhate.eu/`
[2] `https://youtu.be/2S4yA1sF9Ao?t=1677`

capture minority classes, which is especially relevant in unbalanced datasets. In addition, confusion matrices were computed for each category, providing the distribution of true positives, false positives, true negatives, and false negatives. Since the evaluation corpus includes multiple annotation dimensions, the performance analysis was performed label by label, and to illustrate the prompt framework used, Table 4.1 presents a selection of labels drawn from different analytical dimensions, together with their corresponding guiding questions and operational definitions. These examples show how the classification schema integrates target groups, rhetorical devices, emotions, and discursive strategies into a unified structure. The complete list of labels, questions, and definitions can be found in Appendix B.

TABLE 4.1. Examples of zero-shot classification labels, questions and definitions.

Label	**Question**	**Definition**
Migrants	Does this text mention or focus on migrants?	Hate speech targeting individuals due to their nationality or migratory origin. Often related to xenophobia.
Irony	Does this text use the rhetorical device of Irony?	Indirect and ambiguous expression used to mock or humiliate, often in the form of sarcasm or humor.
Fear	Does this text express Fear?	Emotion associated with the perception of threat or danger (physical or symbolic) posed by the target group.
Dehumanization	Does this text use Dehumanization?	Representation of the target group as less human, facilitating hostility or violence against it.
Hope	Does this text express Hope?	Positive emotion associated with reconciliation, change, or improvement in intergroup relations.
Counter-speech	Does this text express Counter-speech?	Response aimed at contradicting hate or harmful speech, promoting respect and inclusion.

The evaluation procedure was automated by iterating over all models and categories. For each model–label pair, predictions were binarized using a decision threshold of 0.5 and compared against the ground truth. Ambiguous or invalid predictions were retained with a value of `-1` during the classification phase. However, the pipeline was rerun at the end for these specific texts to correct the prediction. Finally, a consolidated results table was generated, where each row corresponds to a model–label pair and includes the calculated metrics, the confusion matrix values (True Negatives (TN), False Positives (FP), False Negatives (FN), True Positives (TP)), and the distribution of class counts in the annotated data set. These results were exported to structured *.xlsx* files, ensuring reproducibility and facilitating cross-model comparison.

CHAPTER 5

Evaluation

This evaluation chapter provides both a validation of the methodological pipeline and an empirical basis for the subsequent interpretive analysis. First, the performance of the topic modeling approach is examined through quantitative indicators of coherence and diversity, assessing the extent to which the identified clusters capture semantically consistent and thematically distinct discursive patterns. Second, the results of automatic classification tasks are evaluated, measuring how far LLMs can detect rhetorical, emotional, and ideological characteristics in a custom prompt setting.

5.1. Topic Modelling Evaluation

The model identified 59 distinct topics as shown in Table 5.1, including the outlier class (topic -1). No application was used to avoid outliers, in order to maintain a more truthful nature of the data [57]. The topics achieved an average coherence score of 0.55, with values ranging from a minimum of 0.29 to a maximum of 0.73. These results indicate a relatively balanced distribution, with most topics displaying satisfactory levels of semantic coherence, particularly considering the nature of the collected data. Regarding topic diversity, the results indicate a high level of lexical diversity between topics, with a maximum value of 0.96 when considering the top 10 terms in each topic. Even with an increase in the number of words analyzed per topic, diversity remains high, 0.92 (top-20) and 0.88 (top-30), showing that the repetition of terms between different topics is low. The slight decrease in diversity as the number of terms evaluated increases is expected, given the increased likelihood of lexical overlap. However, the value of 0.81 for the top-50 terms continues to indicate a good thematic separation between topics, reinforcing the robustness of the model in capturing the distinct dimensions of the discourse.

In line with the experimental and iterative parameter adjustments described in Section 4.1, the interpretation of these results followed the same logic. Throughout these cycles, topic coherence and diversity metrics were systematically compared with direct inspection of topic terms and representative documents. This process ultimately became a manual assessment in which the stability, interpretability, and thematic separation of the topics were evaluated together. The decision to retain the final set of 59 topics therefore reflects an empirically guided judgment of what provided the most meaningful and analytically relevant representation of the discourse in this dataset.

A short description was attributed to each topic by providing its top 50 keywords and the domain context to the GPT-4o LLM [60], [61]. The generated descriptions were then manually curated. Considering the hierarchical nature of the topics generated by

TABLE 5.1. Topic-model labels assigned to the 59 *BERTopic* clusters.

ID	Definition
-1	Others
0	Communication, the Press and Political Mobilization
1	Political Vandalism and Ideological Conflict
2	Corruption and Abuse of Power
3	Political Coalitions, Controversies and Political Responsibility
4	Media Events and Political Events and Campaigning
5	Informal Expressions and Personal Interaction
6	Online Debates and Interpersonal Criticism
7	Social Tensions and Immigration
8	Indoctrination and Ideological Polarization
9	Parliamentary Activities and Constitutional Review
10	Elections and Political Strategy
11	Tax Burden and Social Costs
12	Crime and Public Safety
13	Political Events and Campaigning
14	Immigration and Border Control
15	European Right and Political Coordination
16	Economic and Social Crisis
17	National Identity and Patriotism, Youth and Education
18	Political and Institutional Conflicts
19	Agriculture, Fisheries and Rural Development
20	Local and Regional Elections
21	Family, Faith, Celebrations and Religion
22	Crisis and Collapse in Health
23	Corruption, Public Management and Clientelism
24	Criminal Justice, Violence and Crime
25	Alternative, Criticism and Political Confidence
26	European Right and National Identity
27	Gender Equality and Feminist Dynamics
28	Elections, Regional Campaigns and International Far-Right Movements
29	Insults and Ideological Conflicts, Disinformation and Public Exposure
30	Diplomacy, Budget and Government Relations
31	Governance, Controversies and Political Responsibility
32	Patriotism, Controversy and Mobilization
33	Opposition, Criticism and Political Management
34	Ideologies and Individual Freedoms
35	Social Interactions, Congratulations and Emotions
36	Controversies and Political Scandals
37	Expressions, Rhetoric and Emotions
38	Racism, Public Opinion and Prejudice
39	Public Safety, Crime and Minorities
40	Disinformation, Defamation and Social Networks
41	Colonialism, Historical Memory and Reparations
42	European Politics, Territorial Action and Institutional Participation
43	Patriotism, National Identity and Historical Memory
44	Indoctrination, Religion and Moral Education
45	Protest, Indignation and Political Outcry
46	Sexuality, Ideology and Child Protection
47	Corruption, Impunity and Criticism of the Political System
48	National Honor, State Careers and Patriotic Duty
49	Public Space, Mobility and Local Identity
50	Religious Conflicts and Terrorism
51	Justice, Cost of Living and Economy, Social Exclusion
52	Elections and Political Participation
53	Government Crisis and Public Administration
54	Political Participation and Institutional Action
55	Political Mobilization and Campaign Rhetoric
56	Public Health and the Hospital System
57	Media, Journalism and Television Representation
58	Participation and Political Figures

BERTopic, this process also allowed the preliminary grouping of topics into thematic categories, as shown in Figure 5.1. Together, the groups map ten broad domains of the party's discourse: institutional politics and electoral strategy; corruption and scandals;

FIGURE 5.1. *'Political Discourse'* thematic hierarchy map.

1.: Others: — **T:** -1 **D:** 2382
2.: Ideological and Public Dynamics
- **2.1:** Public Order, Clash of Values, Religious and Identitarian Issues
  - **2.1.1:** Conservative Approach, Public Order, Nationalism/Patriotism and Traditional Values
    - **2.1.1.1:** Social, Political Accountability and Dissatisfaction — **T:** 45, 47 **D:** 130
    - **2.1.1.2:** National Identity and Tradition, Ethnic Issues, Immigration and Security — **T:** 7, 12, 14, 21, 39, 41, 43, 48 **D:** 1017
  - **2.1.2:** Ideological Positions, Moral, Sexual and Religious Debates
    - **2.1.2.1:** Sexual Crimes, Feminism/Abortion, Religious Education and Gender/Sexuality Debates — **T:** 24, 27, 44, 46 **D:** 357
    - **2.1.2.2:** Religious Conflicts, Political Ideologies, Extreme Disputes, Relations between Belief, Morals and Politics — **T:** 8, 32, 34, 50 **D:** 440
- **2.2:** Discourse of Social and Digital Interaction, Internet Colloquial Tone — **T:** 5, 6, 29, 35, 37, 38, 40, 57 **D:** 1037

3.: Party, Electoral and Parliamentary Dynamics
- **3.1:** Governance, Economy and Social Policies, Public Health
  - **3.1.1:** Public Management, Health, Investigations/Scandals and Power Conflicts — **T:** 22, 30, 31, 33 **D:** 390
  - **3.1.2:** Socio-Economic Criticism and Debate — **T:** 11, 16, 23, 51 **D:** 539
  - **3.1.3:** Controversies and Disputes about Political Leadership and Government Management — **T:** 53, 56, 58 **D:** 165
- **3.2:** Political-Electoral Dynamics, Public Management, Ideological Debates and Media Coverage
  - **3.2.1:** Political Dispute, Ideological Positions, Communication and Media — **T:** 0, 1, 2, 3, 4, 10, 15, 18, 25, 26, 36, 42, 52, 54, 55 **D:** 2859
  - **3.2.2:** Political-Parliamentary and Electoral Cycle, Mobilization, Interaction and Event Organization — **T:** 9, 13, 17, 19, 20, 28, 49 **D:** 1007

ideological conflict and online polarization; national identity and historical memory; European and transnational far-right coordination; social tensions around migration, crime, and minority rights; religion and moral issues; economic hardship and public health crises; media and communication dynamics; and a residual class that gathers low-frequency or heterogeneous items.

Figure 5.1 shows the thematic hierarchy map, together with the corresponding topic identifiers (T) and the number of associated documents (D). The groups were supported and explored with hierarchical clustering (as provided by *BERTopic*).

This structure supported the assignment of broader interpretative meanings, resulting in a hierarchically organized thematic framework. Three overarching domains emerged: (0) Other or outliers (2,382; 23%), being a residual class; (1) Ideological and public dynamics (2 981; 29%), divided between security-identity discourses (1 017 documents) and moral-religious discourses (797), accompanied by 1 037 colloquial digital speeches that amplify these frames; and (2) Party, electoral, and parliamentary dynamics (4,960; 48%), covered by 1,094 texts on Social policies, health, leadership and economy, 1,007 on territorial mobilization, and above all, 2,859 disputative-style speeches combining accusations of corruption with campaign communication, highlighting the party's central media strategy. Taken together, the hierarchy exposes three interdependent rhetorical pillars: an identitarian securitarianism based on the 'us' versus 'them' dichotomy, a countercultural moralism oriented toward defending traditional values, and a dramatized antisystem denunciation,

disseminated through strong media and digital presence. The uneven distribution of the documents, with almost half dedicated to political confrontation and around a third to the identity-moral axis, suggests that the visibility gained in parliamentary debates and in the media serves as a vehicle for reinforcing emotive messages of security, nationalism, and moral order. This empirical framework thus offers a robust starting point for temporal analyses of the party's thematic evolution and for checking how peaks in public attention correlate with the strategic use of certain frames.

A closer examination of the 50 most relevant keywords per topic reveals distinct lexical clusters that structure the moral-religious and ideological domain. Items related to theological and geopolitical conflict — *('Jihad'), ('Israel'), ('Hamas'), ('Gaza')* — appear side by side with doctrinal and cultural markers such as *('Marxism'), ('Communism'), ('Expression')*. Beliefs, national belonging, and strategic positions are also evoked through terms such as *('Patriotism')* and *('Referendum')*. The colloquial-digital register highlights the interactive language of online platforms — *('posts'), ('tweet'), ('link')* — combined with affective and informal expressions like *('ahahah')* or *('hug')*. This register also includes the lexicon of conflict and disinformation — *('fake news'), ('defamation'), ('coward')* — alongside references to media outlets such as *CNN* and *SIC*. The result is a digital public sphere oscillating between sociability and polarization. Within the governance axis, the health crisis is evoked by references to *('emergency rooms')* and *('shutdowns')*, while political accountability emerges through terms such as *('audit')* and *('scandal')*. Socioeconomic criticism combines the tax burden and cost of living — *('taxes'), ('inflation'), ('pensions')* — with notions of *('clientelism')* and *('tolls')*, alongside markers of vulnerability such as *('disabled')* and *('income')*. This axis outlines a picture of scarcity, inequality, and suspicion toward public management. The political-electoral cycle and strategic communication amplify mediation and mobilization vocabularies — *('debates'), ('interview'), ('vote')* — together with campaign rhetoric *('victory'), ('alternative')*. Institutional conflict is marked by *('vandalism'), ('fascist'), ('censorship')*, linked to coalition dynamics as *('PSD')* and *('PS')*. European coordination appears through acronyms such as (*'VOX'*) and (*'Fidesz'*), while territorialized politics emerges in references to local rallies, fairs or Portuguese cities including *Lisbon* and *Porto*.

Taken together, these clusters illustrate a discursive ecosystem ranging from moral-identitarian to operative-electoral. The full set of top-50 keywords for each topic is available in Appendix A.

5.2. Automatic Text Classification Evaluation

The application of the zero-shot classification pipeline to the corpus enriches the *BERTopic* results and deepens our understanding of the discourse embedded in the data, while its evaluation shows how the layering of emotional and rhetorical signals on each topic sharpens thematic resolution and supports a broader interpretive reading, allowing the detection of latent strategies of persuasion, polarization, and resistance, as well as the intersections with *BERTopic* topics that reveal underlying patterns.

As mentioned in Section 4.2, a total of 34 distinct labels were used, capturing a wide range of rhetorical, emotional, and ideological categories — from hate speech types (e.g., *Direct Hate*, *Symbolic Threat*, *Dehumanization*) to rhetorical devices (e.g., *Hyperbole*, *Irony*), emotional markers (e.g., *Fear*, *Anger*, *Shame*), hate speech strategies (e.g., *Hope*, *Denial of Hate*, *Stereotyping*) and hate speech targets (e.g., *Migrants*, *LGBTI+*, *Roma*) based on the guidelines of the project *kNOwHATE: kNOwing online HATE speech*[1] annotation scheme[2]. Each classification was performed using natural language prompts designed to assess the presence or absence of a given label.

This approach was implemented in six LLMs - including *Mistral*, *Meta (LLaMA)*, *Microsoft (Phi)*, and *Google (Gemini)*. For each model-label pair, performance was evaluated using standard classification metrics such as precision, recall, F1 score, and accuracy, compared to the manually annotated corpus provided by the *kNOwHATE* project. The performance ranking was then applied, classifying the results as *Good*, *Mid* and *Bad*, as shown in Table 5.2. Although this provided a consistent framework for comparing models, the evaluation was interpreted in light of the inherent imbalance of the annotated dataset.

Several labels contained very few positive instances, for example, labels such as *Metaphor* and *Metonymy* totaling only 2 (FN and TP columns, as shown in Table 5.3), which limited the ability to fully observe how the models behaved when detecting these cases. Rather than introducing bias into the classification outcomes themselves, this imbalance mainly constrained how clearly the models' sensitivity to positive labels could be assessed. For almost all labels, the negative class represents 95–99% of the examples. As a result, part of this evaluation process required manual analysis of error distributions, ensuring that performance ratings reflected what could realistically be inferred from the available data.

To support this interpretation, a simple rule-based procedure was used to assign the *Good*, *Mid* and *Bad* labels across models. The decision relied on two elements: the F1 score of each model–label pair and the structure of its confusion matrix, particularly the number of true positives relative to the number of annotated positive cases. A pair was classified as *Good* when the model produced a detectable and interpretable signal for the positive class — typically with several true positives (rather than one or two isolated hits), a non-zero F1 score, and a negative class that remained mostly stable, i.e., not dominated by false positives. The *Mid* category captured cases where the model showed some ability to identify the label, but in a limited or unstable way. This included situations where recall was high but accompanied by a substantial number of false positives, or where the number of positive instances in the dataset was too small for the output to be considered robust despite a non-zero F1 score. Finally, the *Bad* category — which appears in the cross-model comparison but not in the Gemini-only table — was assigned to cases where the model did not produce a meaningful or reliable signal. This occurred when the F1 score was effectively zero or when the confusion matrix revealed degenerate patterns such

[1] `https://knowhate.eu/`

[2] `https://youtu.be/2S4yA1sF9Ao?t=1677`

as the absence of true positives or a prediction behavior dominated by systematic false alarms. In these scenarios, the model–label pair could not be considered analytically useful for downstream interpretation.

TABLE 5.2. Overall zero-shot classification performance by model-label pair.

Model	**Good**	**Mid**	**Bad**	**Total**
google/gemini-2.0-flash-lite-001	30	4	0	34
mistralai/mistral-nemo	21	9	4	34
microsoft/phi-4-multimodal	20	14	0	34
meta-llama/llama-3.1-8b	8	18	8	34
meta-llama/llama-3.2-3b	1	14	19	34
mistralai/Mistral-7B-Instruct	2	6	26	34
Total	83	64	57	204

Among all models tested, *Gemini 2.0 Flash Lite* consistently outperformed the others, with 30 of 34 classifications rated as *Good*, as shown in Table 5.3. Its performance was particularly strong in complex categories related to minority groups, such as *Roma* and *LGBTI+*, whereas it was comparatively weaker, still satisfactory, in detecting emotions or subtle rhetorical devices.

The model achieved high *Recall* in several categories, although at the expense of low *Precision*, which led to a considerable number of false positives. It performed poorly in rare emotional categories and showed uneven effectiveness across discursive axes: reasonably accurate in *Counter-speech* and *Rhetorical Questions*, but weak in *Metonymy* and *Hyperbole*, a limitation likely related to the scarcity of balanced data from the annotated corpus. Despite these limitations, the robustness observed across most categories reinforces *Gemini's* reliability for fine-grained classification tasks involving implicit content. In contrast, models such as *Mistral-7B Instruct* and *LLaMA 3.2 3B* produced poor *Recall* and *F1* scores, with a higher share of *Bad* classifications. Therefore, *Gemini 2.0 Flash Lite* was selected as the primary model for downstream analysis.

TABLE 5.3. Performance metrics per label for the Gemini 2.0 Flash Lite model.

Label	**Accuracy**	**Precision**	**Recall**	**F1 score**	**TN**	**FP**	**FN**	**TP**
Direct Hate	0.93	0.15	0.64	0.24	739	52	5	9
Indirect Hate	0.76	0.34	0.48	0.40	549	122	70	64
Counter-speech	0.81	0.60	0.41	0.49	575	49	106	75
Offensive Speech	0.46	0.04	1.00	0.07	353	435	0	17
Comparison	0.80	0.19	0.47	0.27	610	131	34	30
Metaphor	0.89	0.06	0.75	0.12	708	89	2	6
Metonymy	0.80	0.01	1.00	0.02	646	157	0	2
Lexical Creativity	0.92	0.03	1.00	0.06	742	61	0	2
Appeal to Fear	0.95	0.16	0.46	0.24	761	31	7	6
Call to Action	0.91	0.27	0.78	0.40	704	69	7	25
Intertextuality	0.80	0.11	0.56	0.18	630	143	14	18
Rhetorical Questions	0.82	0.27	0.78	0.40	616	129	13	47
Irony	0.61	0.32	0.77	0.45	364	274	39	128
Hyperbole	0.82	0.04	0.75	0.08	652	145	2	6
Disgust	0.85	0.02	1.00	0.03	685	118	0	2
Anger	0.67	0.11	0.94	0.20	506	264	2	33
Hate	0.81	0.02	1.00	0.04	649	153	0	3
Hope	0.96	0.03	1.00	0.06	775	29	0	1
Shame	0.94	0.04	0.67	0.07	753	49	1	2
Guilt	0.98	0.00	0.00	0.00	787	18	0	0
Fear	0.97	0.00	0.00	0.00	784	19	2	0
Stereotyping	0.65	0.17	0.88	0.29	465	274	8	58
Role Reversal	0.89	0.05	1.00	0.10	710	90	0	5
Denial of Hate	0.77	0.05	0.75	0.08	609	184	3	9
Symbolic Threat	0.75	0.03	0.88	0.06	596	201	1	7
Realistic Threat	0.96	0.23	0.37	0.29	763	23	12	7
Silencing	0.69	0.15	0.67	0.24	518	229	19	39
Dehumanization	0.86	0.07	0.89	0.13	686	110	1	8
Migrants	0.91	0.39	0.37	0.38	710	35	38	22
Roma	0.98	0.88	0.71	0.79	748	5	15	37
Racialized	0.86	0.46	0.62	0.52	628	75	39	63
Others	0.89	0.08	0.27	0.12	712	71	16	6
Intersectionality	0.98	0.07	1.00	0.13	790	14	0	1
LGBTI+	0.92	0.78	0.68	0.73	658	24	39	84

CHAPTER 6

Analysis

This chapter presents the analytical stage of the study, moving from the methodological and evaluative dimensions to the interpretation of the results. The analysis explores the evolution of discourse over time, the strategies used, and the ways in which it contributes to the processes of political communication and mobilization. The discussion begins with an overview of the temporal and thematic developments observed in the data, before turning to a closer examination of the discursive structures and their implications.

6.1. Political Discourse Evolution

From late 2019 to mid-2024, the aggregate monthly evolution of the three main thematic dynamics highlights key electoral moments (marked with red dots) and reveals patterns of discursive intensification, as illustrated in Figure 6.1.

The graph shows not only the volume of monthly mentions within each category, but also allows us to observe patterns of discursive intensification associated with key moments in the national political calendar, such as legislative and presidential elections. The overlap of curves and markers indicates a possible correlation between the increase

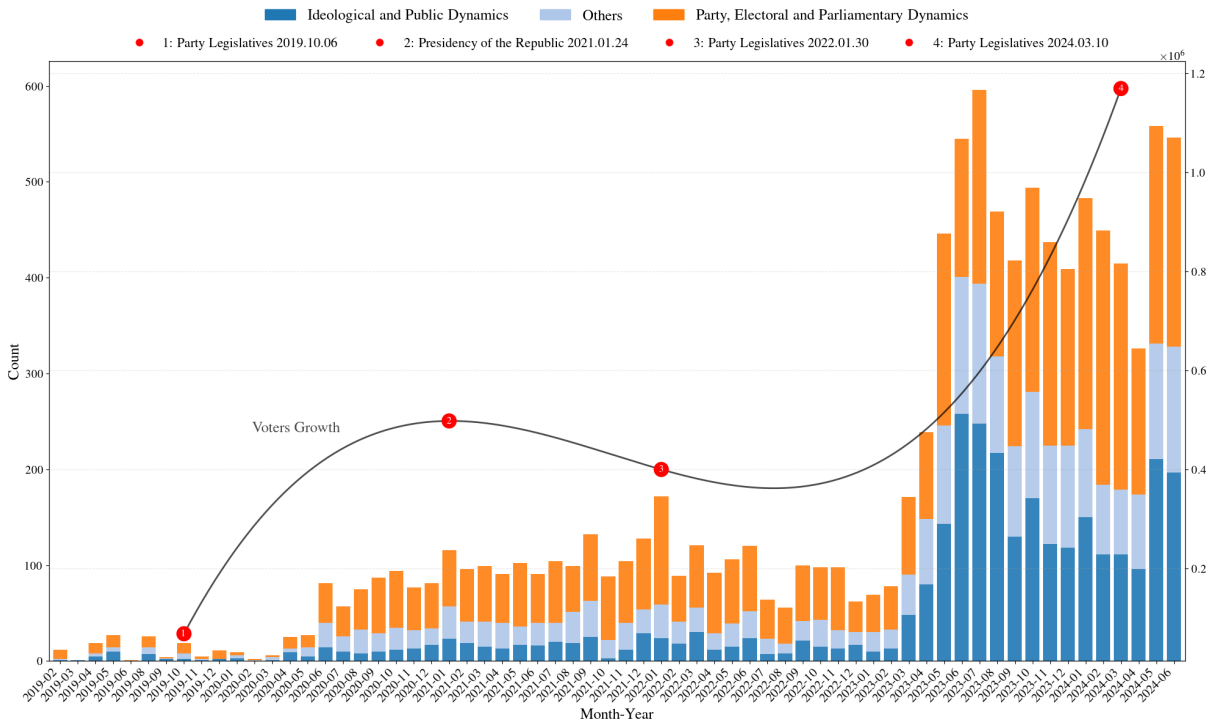


FIGURE 6.1. Monthly evolution of *'Political Discourse'* on *X (formerly Twitter)* associated with CHEGA (2019-2024).

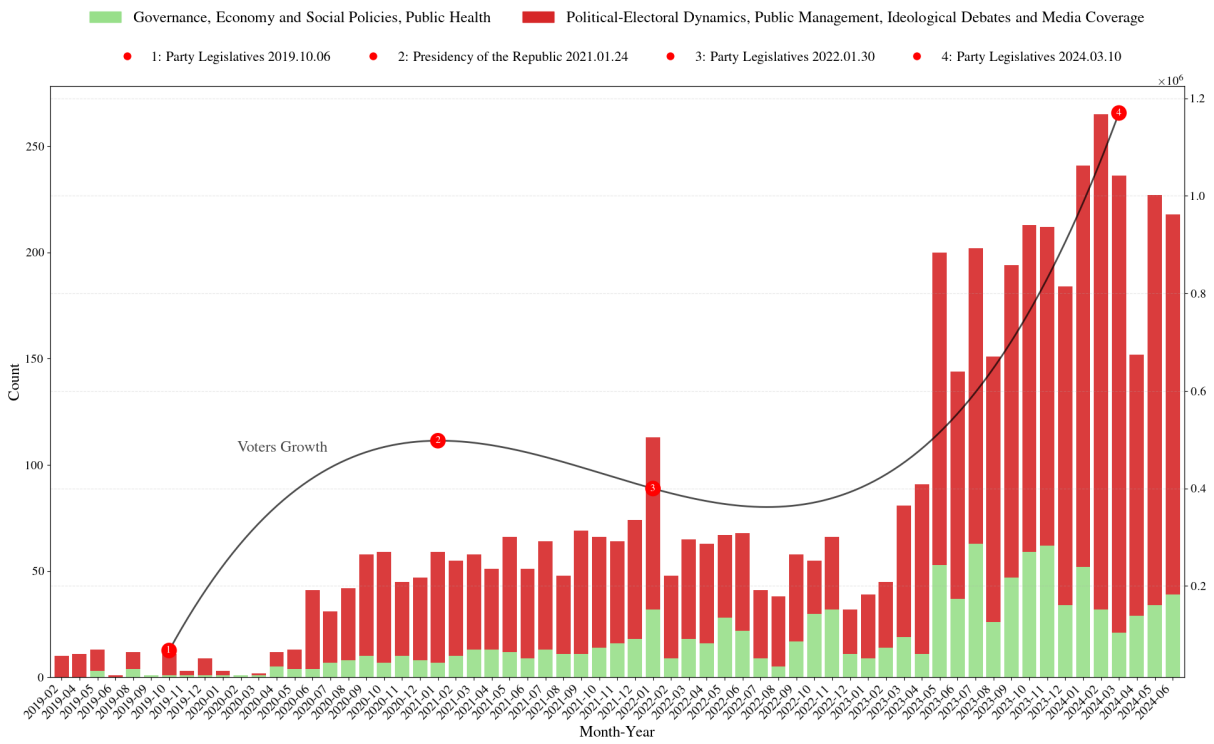


FIGURE 6.2. Monthly Evolution of *'Party, Electoral and Parliamentary Dynamics'* dimension.

in discursive activity on *X (formerly Twitter)* and electoral cycles, reflecting both the growth of the party's support base and its strategic positioning on social networks over time. Longitudinal analysis of the results shows that the discourse of the political figures involved is structured around two major complementary forces: political-electoral consolidation and ideological-moral polarization. Both are increasingly articulated throughout the different electoral cycles, reflecting a trajectory that accompanies the party's own institutional and communicational evolution. In the first few years after its founding on 9 April 2019, the volume of discursive interactions was still low and concentrated in moments of greater media visibility, especially during the 2019 legislative elections. In this initial period, party figures built their public presence around electoral dynamics, with some moments mixing with ideological dynamics, focusing on opposition to political elites, setting the tone for a rhetoric of protest. There has been a gradual expansion in discourse since April 2020, one month after the official declaration of a state of emergency due to *COVID-19* in Portugal, with a greater focus on political and electoral disputes and government criticism. This trend remained strong throughout the following years, as shown in Figure 6.2.

In this expansion, the political results in the 2021 presidential elections are promising, after a surprising candidacy by André Ventura, leader of the CHEGA Party, reaching third place[1], emphasizing that his political presence, linked to the founding of the party, would only be two years old. Although economic and public policy issues, such as healthcare

[1] `https://www.eleicoes.mai.gov.pt/presidenciais2021/resultados/globais`

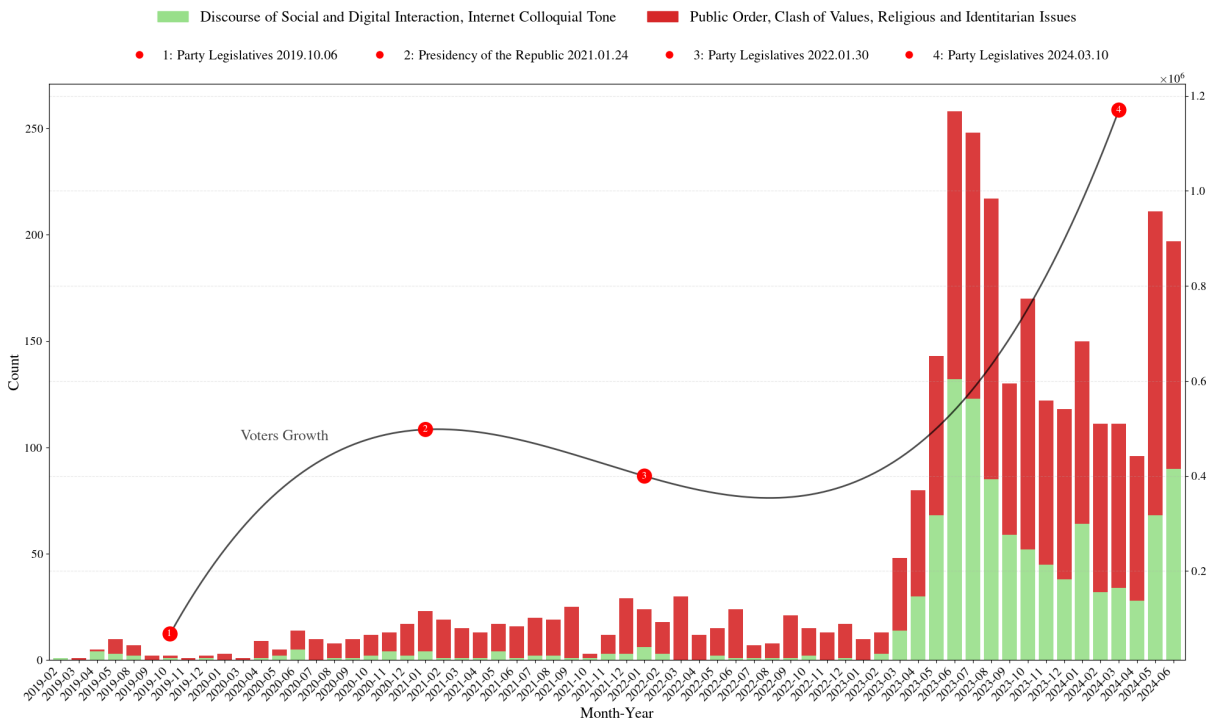


FIGURE 6.3. Monthly Evolution of *'Ideological and Public Dynamics'* dimension.

and management scandals, were relevant to some extent, the discourse remained strongly anchored in political dynamics, such as elections, the European right, scandals, and ideological themes. This discourse structure suggests that these figures of the CHEGA party sought both pragmatic legitimacy, through criticism of governance, and symbolic mobilization, through identity division. The political-electoral dimension highlights how media visibility became a central factor, with parliamentary debates and political controversies being thus transformed into discursive capital, simultaneously fueling digital engagement and strengthening the institutional projection of the party, with marked increases in messages related to local mobilizations, campaign events, constitutional revisions, and even references to international far-right movements. The discourse expansion pattern remained consistent until April 2023, with a strengthening of parliamentary presence during the 2022 legislative elections. In this period, the results suggest that the discursive institutionalization of these CHEGA figures becomes evident, with references to parliamentary disputes and electoral mobilizations taking on a central role. This evolution reflects the process by which the party is no longer seen solely as a protest movement but rather as a consolidated political actor in the party system. From 2023 onward, the ideological component of the party's discourse gains more prominence, displaying significant growth. Shown in Figure 6.1, and in more detail in Figure 6.3.

This dimension is reinforced through debates on public order, religion, sexual morality, gender, and nationalism alongside the increasing adoption of a digital and interactive register associated with social media. The topics related to digital interaction reveal a residual presence between 2019 and 2022, limited to informal expressions and occasional

online interactions of little significance. However, from 2023 onward this group undergoes an exponential increase, becoming one of the most prominent elements of party digital communication, with peaks exceeding 120 mentions. The most significant topics during the 2023–24 peaks include online debates and interpersonal criticism, informal expressions and personal interaction, and issues related to racism, public opinion, and prejudice. Other themes, such as disinformation, defamation, and social networks, as well as rhetorical and emotional expressions, appear on a smaller scale, but nonetheless reinforce the dynamics of digital polarization. The use of insults, ideological conflict, and public shaming accompanies this growth, consolidating the aggressive rhetoric typical of the digital environment. Overall, this suggests that from 2023 onward, CHEGA figures began to use *X* not only as a platform for political-institutional messaging, but also as a space for emotional mobilization and symbolic conflict. This increase coincides with the period of increased electoral visibility before the 2024 legislative elections, indicating that the party used digital rhetoric to strengthen engagement, stimulate controversy, and mobilize supporters.

The period between 2023 and the 2024 legislative elections corresponds to a phase of discursive intensification, in which two axes, political-electoral institutionalization and ideological-moral radicalization, emerge as mutually reinforcing. This discursive expansion directly mirrors the trajectory of electoral growth, with peaks linked to high-visibility political events, while also being sustained by a continuous presence in ideological debates. Within the strongest subdimensions of the ideological axis, *'Conservative Approach, Public Order, Nationalism/Patriotism and Traditional Values'* and *'Ideological Positions, Moral, Sexual, and Religious Debates'*, distinct patterns of thematic centrality emerge. Themes related to *'National Identity and Tradition, Ethnic Issues, Immigration and Security'* clearly dominate over those related to *'Social, Political Accountability and Dissatisfaction'*, highlighting the discursive emphasis on immigration, security, and the defense of national and traditional values, hallmarks of nativist rhetoric.

In contrast, for ideological positions, a more balanced distribution is displayed across topics such as *'Religious Conflicts, Political Ideologies, Extreme Disputes, Relations between Belief, Morals and Politics'*, alongside debates on *'Sexual Crimes, Feminism/Abortion, Religious Education and Gender/Sexuality Debates'*. The most significant growth comes from issues related to gender, feminism, and sexuality, which gained notable prominence and discursive strength during this intensification period, as shown in Figure 6.4. *'Social, Political Accountability and Dissatisfaction'* has been continuously present since 2019, although with residual intensity, mainly centered on criticism of governance and protest-driven or indignant rhetoric. Meanwhile, the axes addressing *'Sexual Crimes, Feminism/Abortion, Religious Education and Gender/Sexuality Debates'*, as well as *'Religious Conflicts, Political Ideologies, Extreme Disputes'*, remain relatively marginal until early 2023. From that point onward, however, a sharp surge emerges, with peaks exceeding 30 occurrences, both dominated by a common theme: indoctrination. These

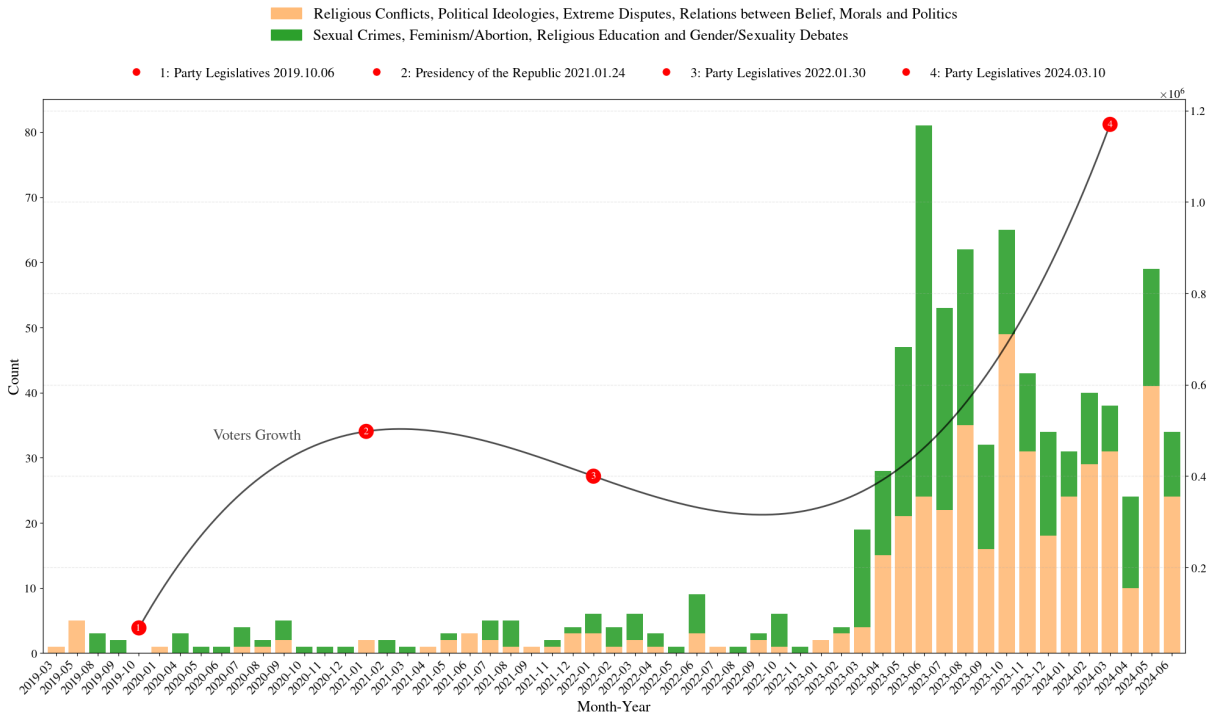


FIGURE 6.4. Monthly Evolution of *'Ideological Positions, Moral, Sexual and Religious Debates'*.

axes serve to reinforce cultural and moral cleavages, thereby consolidating a polarizing rhetorical framework.

When detailing, for example, the analysis for the dimension *'National Identity and Tradition, Ethnic Issues, Immigration, and Security'*, which is a predominant area of focus, it comprises 1,017 documents, representing approximately 10%. The three topics directly related to immigration and public order (topics 7, 12 and 14) share the words *'threat'*, *'violence'*, and *'lack of control'*, which reinforces the idea that internal security and border management merge into a single narrative about external risk. In contrast, topic 21 *('Family, Faith, Celebrations, and Religion')* shifts the focus from security anxiety to a register of values and affective belonging. Words like *'brotherhood'*, *'Fátima'*, *'fatherhood'*, and *'happyfamily'* refer to catholic festivities, family cohesion, and a positive imaginary community. This vocabulary acts as a discursive counterbalance: Where the migratory topics point to a threat, this one evokes the protection of traditions. Topics 39, 41, 43 and 48 fall between these two poles. Topic 39 takes up the *'crime'* but adds ethnic markers *('gypsies'*, *'negro')*, indicating that public security is also framed by specific minority categories. Topics 41 and 43 invoke the historical past *('carnation'*, *'Salazar'*, *'Camões')* to legitimize a selective memory of the nation, sometimes celebratory *('heroes'*, *'homeland'*, *'pride')* and sometimes resentful *('reparations*, *'traitors')*. Finally, topic 48 converges on the valorization of state careers *('military'*, *'firefighters'*, *'police officers')* and the *'homeland'* as a duty, connecting national honor with corporate recognition.

Figure 6.5 shows a progressive increase in the incidence of topics related to immigration, public safety, crime, nationalism, and social tensions. The graph shows not only

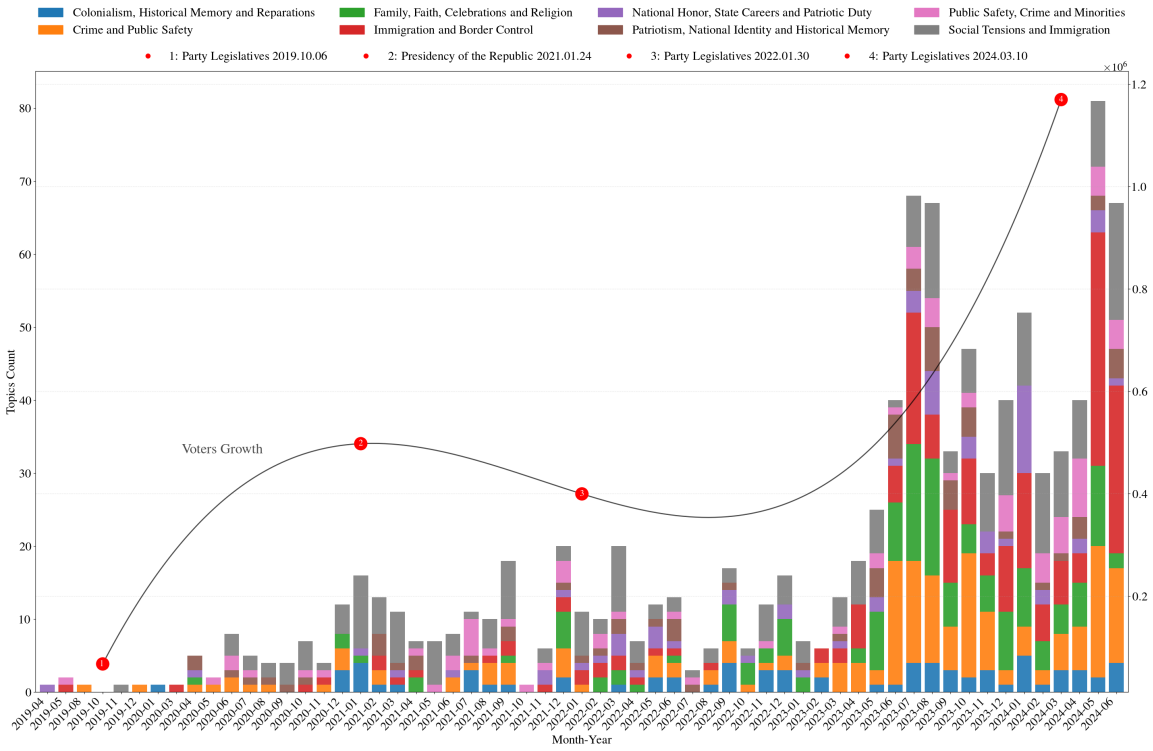


FIGURE 6.5. Monthly Evolution of Topics Related to *'National Identity, Tradition, Ethnic Issues, Immigration, and Security'*.

the volume of monthly mentions within each category, but also allows us to observe patterns of discursive intensification associated with key moments in the national political calendar, such as legislative and presidential elections.

The relationship between the curve and the electoral cycles is maintained, and, as already pointed out, this growth becomes particularly expressive throughout 2023 and into the early months of 2024, culminating in a sharp discursive peak just before and after the legislative elections in March 2024. This is reflected in the growth in voting numbers, with the range of figures in the 2022 parliamentary elections being around 400,000 rising to more than a million in 2024. Among the most recurrent subthemes are *'Social Tensions and Immigration'*, *'Crime and Public Safety'*, *'Immigration and Border Control'*, and *'Family, Faith, Celebrations, and Religion'*. In particular, *'Social Tensions and Immigration'* appear more consistently throughout the timeline, suggesting its centrality in the long-term discursive strategy of the party. Collectively, these themes account for a substantial portion of the discursive volume within this dimension, indicating a communication strategy grounded in the mobilization of negative affect, particularly fear, insecurity, and perceived threat, often linked to the presence of minorities or immigrants, and emotionally reinforced through references to family, faith, and religious values. Further strengthening this divisive narrative, topics such as *'Colonialism, Historical Memory, and Reparations'*, *'Patriotism, National Identity, and Historical Memory'*, and *'National Honor, State Careers, and Patriotic Duty'* also appear consistently.

These topics suggest a symbolic articulation between security, identity, and morality, constructing a vision of national cohesion under threat. The temporal overlap between the rise in discursive activity and the electoral cycles indicates an instrumental use of these narratives for political mobilization. The overlaid curve, representing the electoral growth of the party, appears to align with the intensification of the security and identity-oriented rhetoric, strengthening the hypothesis that these themes play a strategic role in consolidating and expanding the voter base of the party. Further visualizations are provided in Appendix C, for subsequent analysis.

6.2. In-Depth Analysis with Automatic Text Classification

From a broader perspective, Figure 6.6 shows that *'Migrants'* constitute the most frequently targeted group, particularly within the dimension *'Ideological and Public Dynamics'*. *'Racialized'* groups represent the second most prominent category, highlighting the ongoing salience of racialization and processes of exclusion in political and ideological debates. *'LGBTI+'* people also appear significantly, although to a lesser extent, highlighting the persistence of an approach that is sometimes quite hostile based on sexual orientation, gender identity or gender expression. In contrast, *'Intersectional'* groups have a very residual presence, revealing that texts related to this label were less frequently validated, suggesting that the speech surrounding these figures tends to be directed more toward singular identity labels rather than the overlapping social positions. The *'Roma'* label, with little presence, suggests a reduced stigmatization of communities when identified by the term 'gypsy'. The *'Party, Electoral, and Parliamentary Dynamics'* dimension registers comparatively lower levels of targeted groups overall, suggesting that formal political debates may involve less overt hostility compared to *'Ideological and Public'* discussions. The *'Other'* label captures diffuse forms of debates directed toward groups not explicitly represented within these major labels, also in a very significant way, and it is interesting to study these texts in order to explicitly understand the groups covered up within.

Anger emerges as the most prevalent emotion in all dimensions, followed by significant proportions of hate and disgust (Figure 6.7). A distinctive feature of the analysis is the behavior of the *'Others'* dimension, which, despite being classified as an outlier due to its size, shows a graphical distribution closely aligned with that of the *Ideological and Public Dynamics*. In particular, it exhibits an even stronger association with anger, highlighting the dominant presence of this emotion compared to all other dimensions. The prominence of anger underscores its centrality in mobilizing political discourse, often reinforcing hostile narratives. In contrast, emotions such as guilt and shame are largely absent, suggesting that these interactions in X within these political figures are driven more by confrontational expressions than by conciliatory or future-oriented affective frames.

Strategies like *Symbolic Threat*, *Stereotyping*, *Denial of Hate* and *Silencing* are among the most recurrent strategies (Figure 6.8), particularly in the *Others* and *Ideological and Public Dynamics* dimensions. These results highlight a recurrent rhetorical mechanism

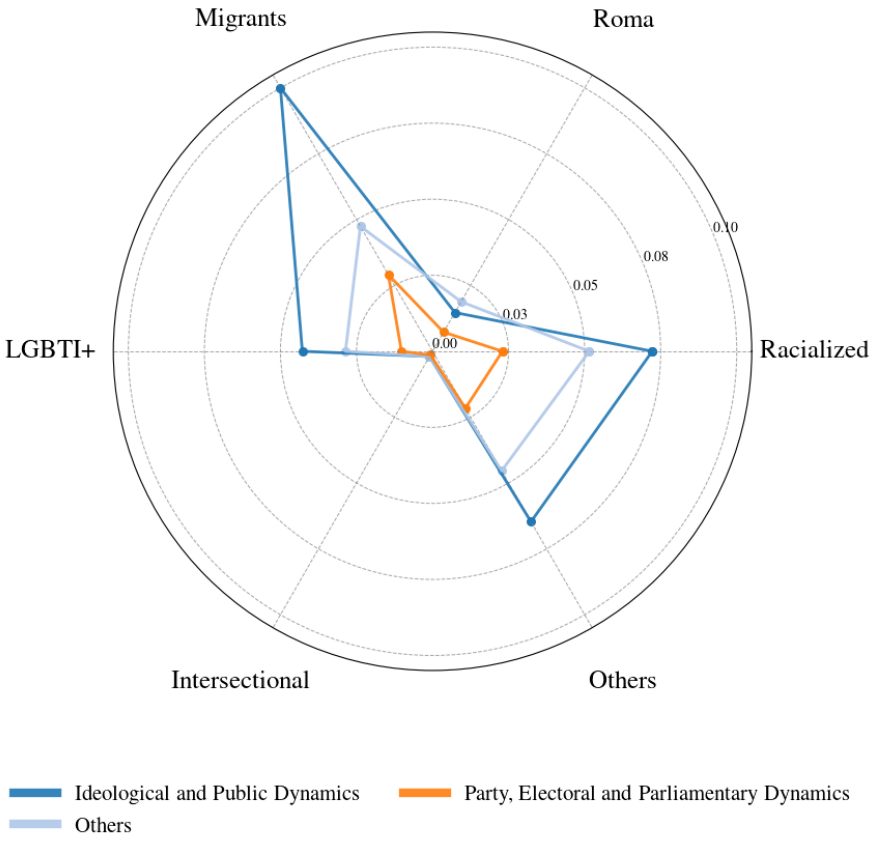


FIGURE 6.6. HS targets probability distribution across *'Political Discourse'*.

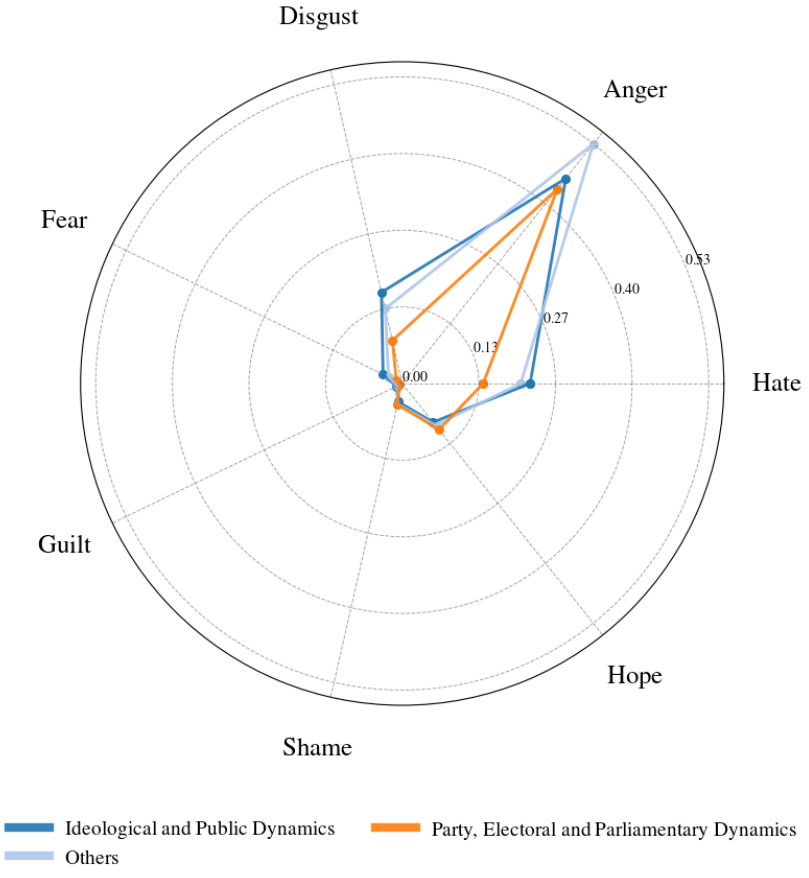


FIGURE 6.7. Emotions probability distribution across *'Political Discourse'*.

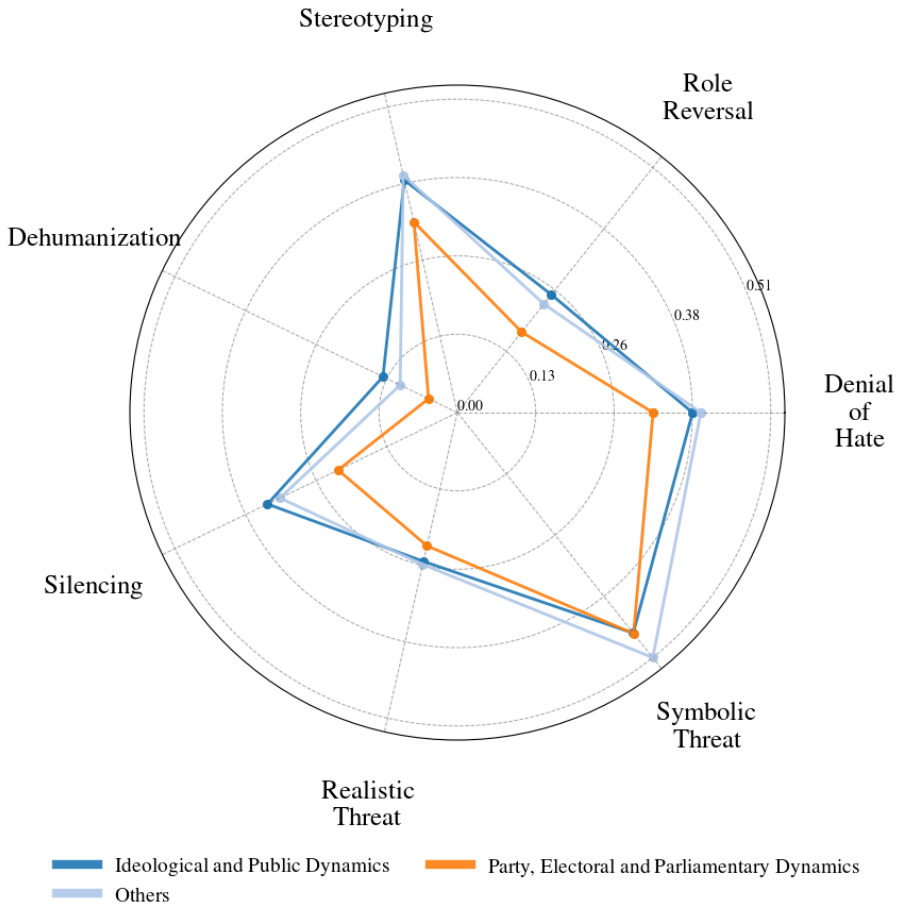


FIGURE 6.8. HS strategies probability distribution across *'Political Discourse'*.

in which party figures use symbolic threats as a dominant strategy, confirming that, in their political sphere, speech is expressed more often through the perception of threats to values, beliefs, and identities than through realistic threats or dehumanization. This may be related to the desire to avoid more explicit forms of incitement, which are less accepted in the political-institutional sphere. *Stereotyping*, *Denial of Hate* and *Silencing* also emerge as recurring elements, suggesting, respectively, the tendency to construct negative social images of minority groups, often claiming to be the victim (*Role Reversal*), and the strategic need to protect discursive legitimacy.

Irony is the predominant rhetorical device (Figure 6.9), which corroborates previous studies that identify humor as one of the most effective ways to cover up hate dissemination [62]. *'Hyperbole'* and *'Metonymy'* also play central roles, functioning respectively as emotional amplification and derogatory symbolic evocation. In contrast, the systematic use of *'Metaphors'*, *'Comparisons'*, *'Appeals to Fear'*, *'Intertextuality'*, and *'Rhetorical Questions'* appears at moderate levels, with *'Lexical Creativity'* basically undetected, suggesting that the discourses of these CHEGA party figures tend to favor shared interpretive resources, using analogies that, in a hateful context, border on the depreciative, also resorting to historical or cultural references, sometimes introducing questions without necessarily expecting answers, in order to legitimize arguments. Finally, the call to action

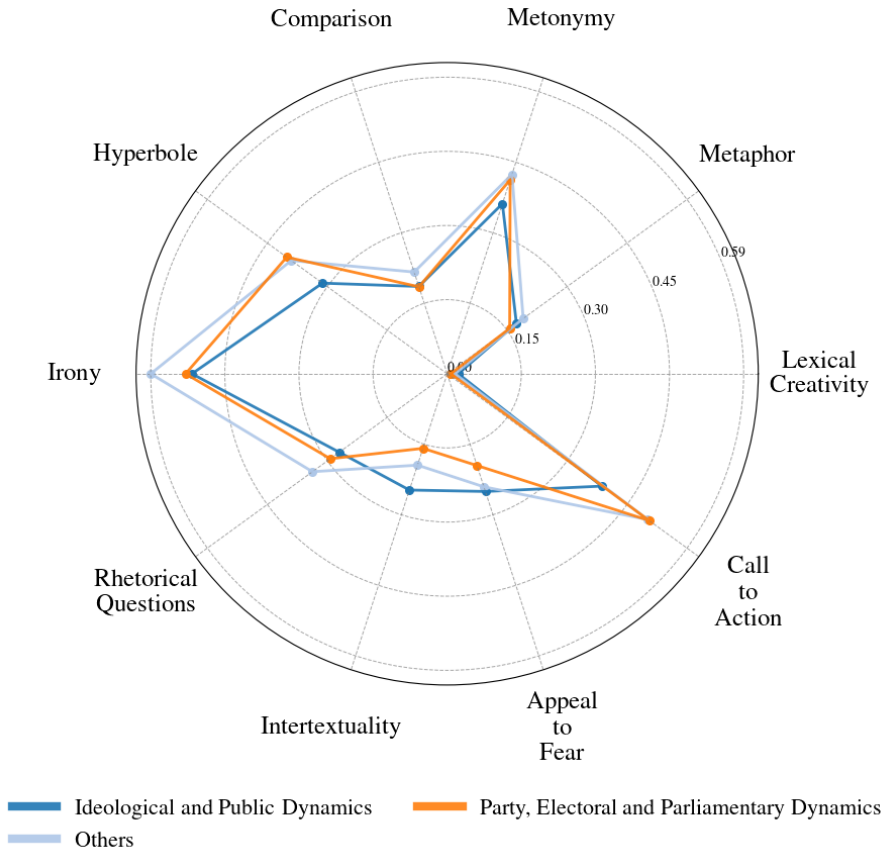


FIGURE 6.9. Rhetorical devices probability distribution across *'Political Discourse'*.

particularly distinguishes in the *'Party, Electoral and Parliamentary Dynamics'* and in the *'Others'* dimensions, highlighting the mobilizing function.

Offensive Speech is by far the most frequent in all high-level dimensions (Figure 6.10), along with *'Indirect Hate Speech'* showing moderate to high presence, particularly in the *'Ideological and Public Dynamics'* and *'Others'* dimension. *'Direct Hate Speech'* and *'Counter-Speech'* appear in much lower proportions, practically residual. This suggests that these political figures tend to avoid explicit expressions of hostility that could compromise their public legitimacy, while nevertheless resorting to more aggressive and personalized communication. There is a clear preference for implicit and indirect discourse, often supported by rhetorical devices such as irony, metonymy, or rhetorical questions, as seen in Figure 6.9. This type of discourse allows hostility to be spread and intergroup boundaries to be reinforced in a veiled manner, while protecting the speaker from direct accusations of inciting hatred. For the *'Party, Electoral and Parliamentary Dynamics'*, this type of expression occurs less frequently, which can be explained by the need for political figures to try to maintain an image of public credibility.

Reinforcing the detailed analysis for the subdimension *'National Identity and Tradition, Ethnic Issues, Immigration, and Security'* (T: 7, 12, 14, 21, 39, 41, 43, 48; D: 1017), which, as revealed with a discursive configuration strongly centered on emotions of confrontation, strategies of exclusion, and explicit targeting of minority groups, shown

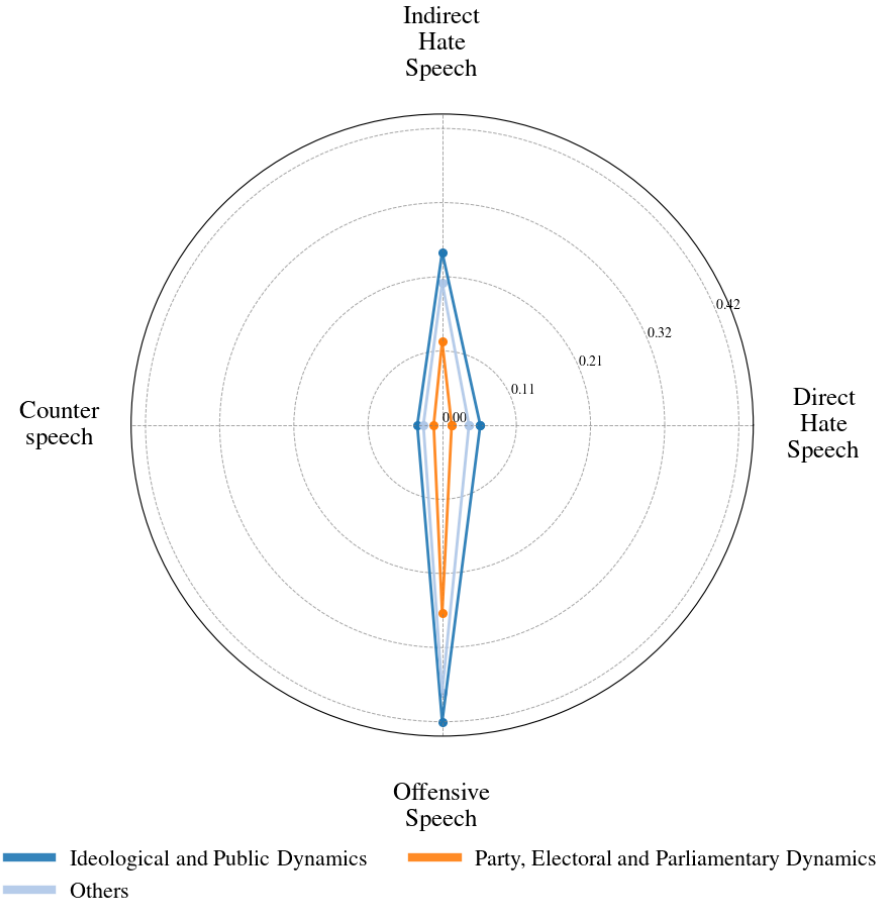


FIGURE 6.10. Speech types probability distribution across *'Political Discourse'*.

in Figure 6.11. *Anger* emerges as the most dominant emotion, reaching its highest levels, on topics that address *Public Safety, Crime (with or without minorities)*, *'Colonialism, Historical Memory and Reparations'* and *'Immigration and Border'*. This behavior was already expected due to the strong expression of the macrogroup. *'Hate'* and *'Disgust'* also appear consistently, further reinforcing the antagonistic tone. *'Fear'* is visibly present in *'Crime and Public Safety'* and *'Immigration and Border Control'*, suggesting how these political figures implement the idea that the external group represents a threat to the security of the internal group, with the need to protect themselves and other important people from these external groups. *'Hope'* assumes a central role, especially within *'Family, Faith, Celebrations and Religion'* and *'Social Tensions and Immigration'*. In the religious and family-related discourse (topic 21), hope is expressed through keywords such as *'jesus'*, *'fátima'*, *'fellows'*, and *'joy'*, framing politics in a moral and spiritual register that projects a collective future anchored in tradition and faith. In contrast, in *'Social Tensions and Immigration'* (topic 7), hope is mobilized through terms like *'trust'*, *'rely'*, and *'resist'*, which suggest a forward-looking orientation that combines resilience with a sense of national perseverance under perceived external threats. Thus, while anger sustains polarization and antagonism, hope provides a mobilizing affect that integrates spiritual identity and collective resistance. Guilt and shame remain marginal on all topics.

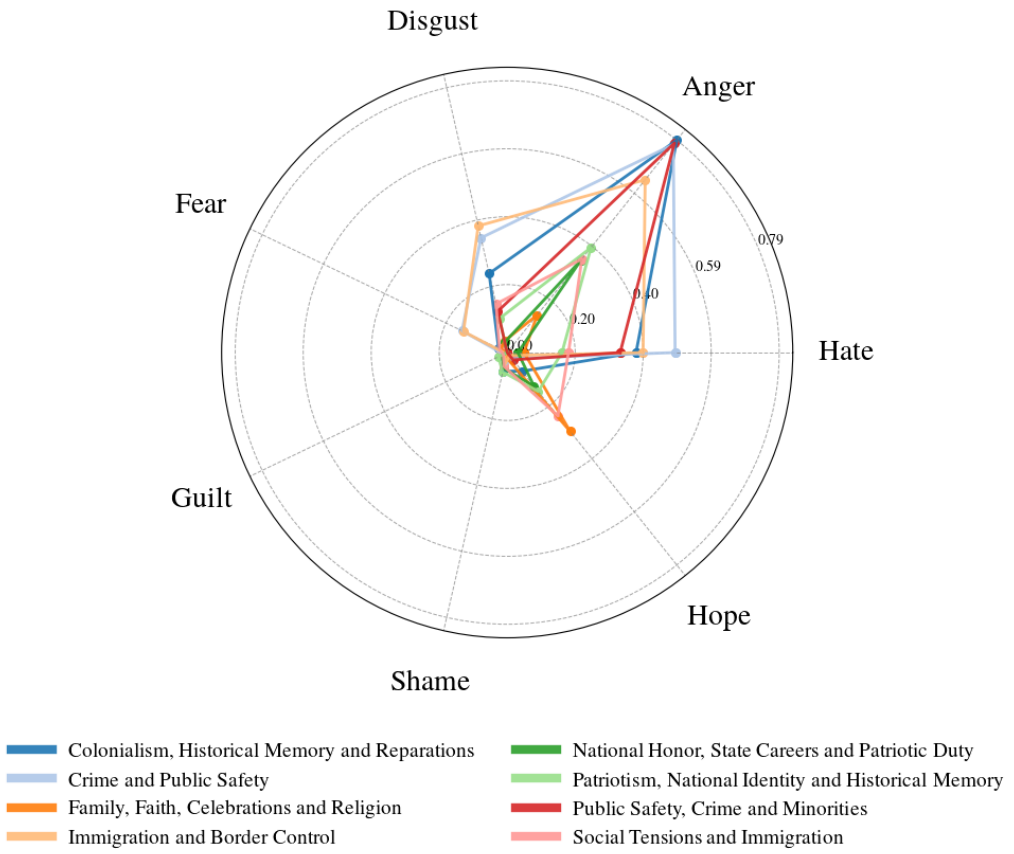


FIGURE 6.11. Emotions probability distribution across *'National Identity and Tradition, Ethnic Issues, Immigration, and Security'* topics.

Regarding speech types (Figure 6.12), the record of antagonistic predominance is consolidated through the significant frequency of the label *'Offensive Speech'*.*'Indirect Hate Speech'* reaches its highest levels in *'Immigration and Border Control'* and *'Crime and Public Safety'*, showing how hostility is often framed in implicit or coded forms. Importantly, *'Direct Hate Speech'* is not merely residual in this subdimension: it is particularly salient in *'Immigration and Border Control'* and in *'Crime and Public Safety'*, where minority groups are explicitly framed through criminality, illegality, or deviance. *'Counter-speech'* remains virtually absent, confirming the limited presence of defensive or corrective practices within these debates. Hate speech strategies (Figure 6.13), reveal a recurrent pattern built around *'Denial of Hate'*, *'Symbolic Threat'*, *'Realistic Threat'*, *'Silencing'*, and *'Stereotyping'*, particularly evident in four main topics — *'Immigration and Border Control'*, *'Crime and Public Safety'*, *'Public Safety, Crime and Minorities'*, and *'Colonialism, Historical Memory and Reparation'*.

This configuration suggests that these political figures explicitly reject accusations of intolerance while simultaneously reinforcing stigmatizing narratives, especially in debates on immigration and crime. Such a paradox functions to protect the legitimacy of exclusionary claims under the guise of neutrality or common sense. Operating as powerful tools of simplification, these strategies reduce minorities to figures of threat or deviance. They are most visible in the context of crime and public safety, where *'Migrants'*, *'Roma'*, and

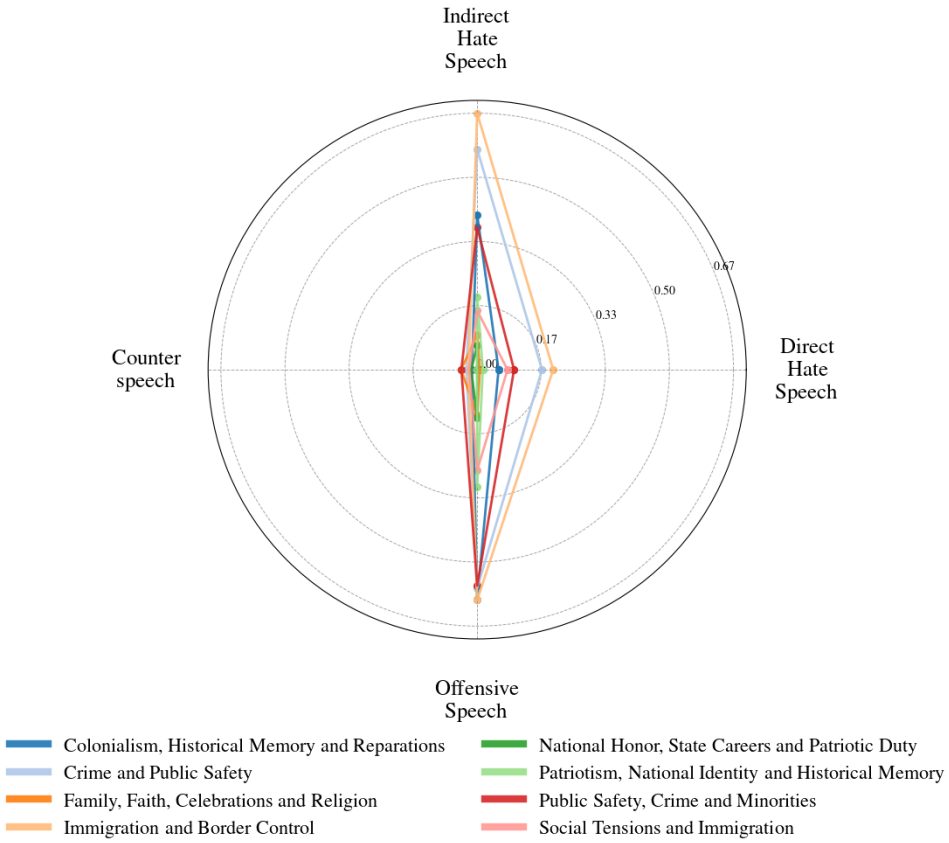


FIGURE 6.12. Speech types probability distribution across *'National Identity and Tradition, Ethnic Issues, Immigration, and Security'* topics.

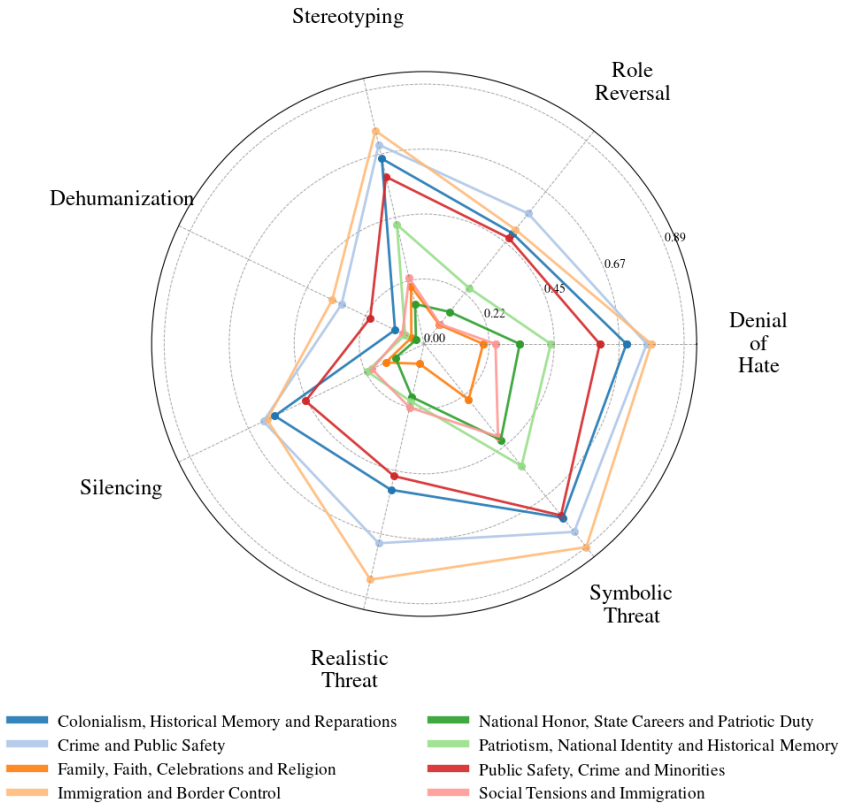


FIGURE 6.13. HS strategies probability distribution across *'National Identity and Tradition, Ethnic Issues, Immigration, and Security'* topics.

'Racialized' groups are not portrayed as individuals but as interchangeable embodiments of criminality. *'Silencing'* plays a complementary role in this dynamic: by discrediting or erasing alternative perspectives-such as those related to colonial memory or reparations-it prevents the emergence of counter-narratives that might destabilize dominant representations. The emphasis on *'Symbolic'* and *'Realistic Threats'* reflects the dual anchoring of these discourses. On one side, symbolic references evoke fears of cultural dilution, religious decline, or loss of national traditions; on the other hand, concrete references to security or territorial control sustain the material plausibility of hostility. Taken together, these strategies show that hate speech in this domain is less about overt insults and more about weaving a coherent framework of legitimacy for exclusion, combining denial, stereotyping, silencing, and threats into a structured grammar of antagonism. *'Migrants'* are by far the most frequent group mentioned (visualization provided in Appendix D), especially within *'Immigration and Border Control'*, strengthening the analysis in the previous paragraph. *'Racialized'* groups and the *'Roma'* community also appear significantly in *'Public Safety, Crime and Minorities'*, *'Crime and Public Safety'*, reflecting how ethnic minorities are framed through narratives of insecurity. *'LGBTI+* and *'Intersectional'* targeting remains marginal, but its appearance signals the overlap of vulnerabilities in identity categories. The label *'Others'* essentially reflects what was analyzed in the top cluster, capturing debates directed at groups that are not explicitly mentioned. It is particularly interesting to understand which groups these are.

The rhetorical devices shown in Figure 6.14 point to a communicative style that thrives on dramatization, exaggeration, and ridicule. *'Irony'* and *'Hyperbole'* stand out as central mechanisms, creating a discursive environment where opponents are not simply contested, but ridiculed, and social problems are inflated into existential threats. This dramatization is reinforced by the recurrent use of *'Metonymy'*, which allows complex phenomena such as immigration or crime to be condensed into symbolic figures or single events, giving abstract anxieties a concrete and emotionally charged form. Within this repertoire, *'Call to Action'* and *'Appeal to Fear'* perform a mobilizing function, particularly visible in debates on immigration and border control. These devices transform insecurity into a rallying cry, framing political participation as a defensive necessity against cultural or material decline. Rather than isolated rhetorical flourishes, such strategies operate as affective triggers designed to convert sentiment into mobilization. Although less prominent, *'Comparison'*, *'Metaphor'* and *'Intertextuality'*, this labels enrich the discourse with symbolic references, cultural analogies, and associative links, reinforcing familiarity and recognition without dominating the overall communicative style. Taken together, the rhetorical pattern shows that far-right communication in this domain is not merely about conveying information, but about staging a drama where fear, ridicule, and symbolic condensation work to legitimate exclusion and incite collective action. Further visualizations are provided in Appendix D, for subsequent analysis.

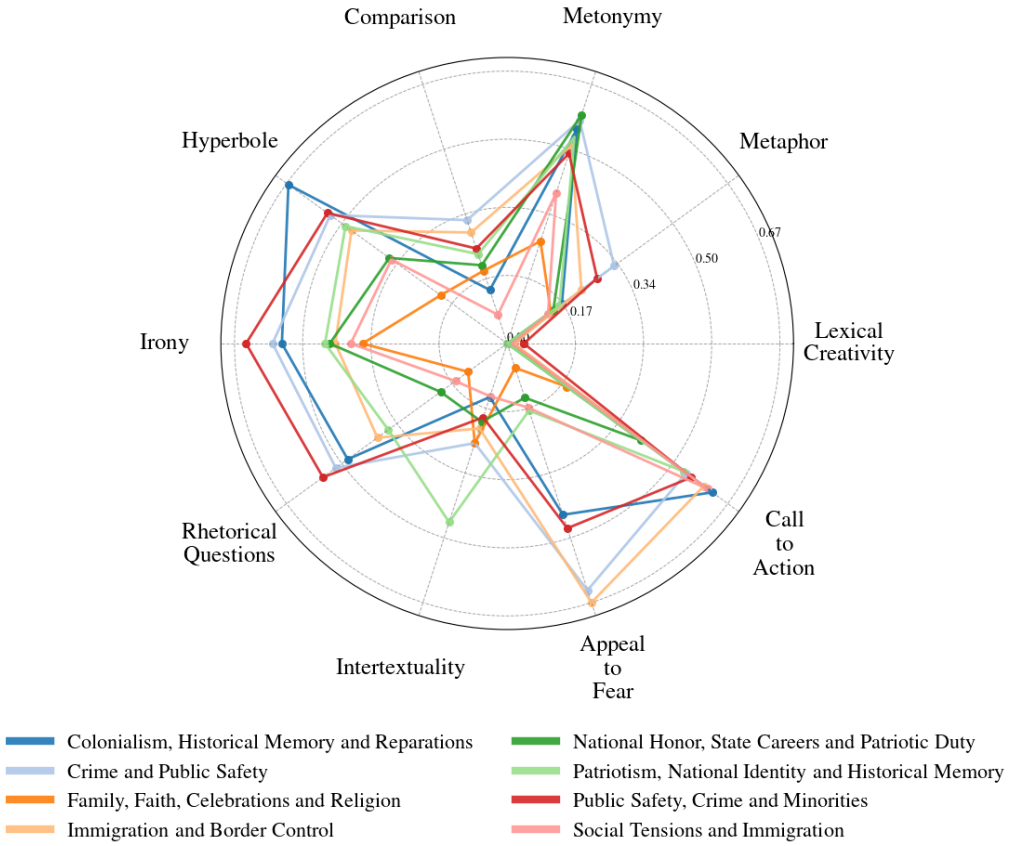


FIGURE 6.14. Rhetorical devices probability distribution across *'National Identity and Tradition, Ethnic Issues, Immigration, and Security'* topics.

CHAPTER 7

Conclusion

This study analyzes 10,323 unique posts on *X (formerly Twitter)* published by key figures of the CHEGA party between late 2019 and mid-2024, with the aim of uncovering latent dimensions of party discourse and examining how rhetorical and emotional strategies evolve over time in relation to broader processes of political mobilization and polarization.

To address RQ2, a combination of *BERTopic* and state-of-the-art LLMs for prompt-based (*zero-shot*) text classification was used to analyze Portuguese political discourse. This methodological framework proved highly effective in detecting latent topics, rhetorical strategies, and emotional patterns with significant interpretive value. The topic modeling process uncovered 59 latent topics, with an average topic coherence of 0.55 and a topic diversity above 0.80 among the top-50 terms. The pipeline integrated custom *LaBSE* embeddings, *UMAP*-based dimensionality reduction, and *HDBSCAN* clustering, combined with instruction-tuned LLMss (GPT-4o) and manual curation, which proved particularly effective for the analysis of multilingual and politically polarized content. In addition, six LLMss were evaluated in the manually annotated *kNOwHATE* corpus, with *Gemini 2.0 Flash Lite* achieving the best performance, outperforming 30 of 34 classification labels despite the imbalance present in the evaluation dataset.

However, the application of these techniques also revealed several methodological challenges, such as handling class imbalance in the evaluation phase, capturing the linguistic nuances of Portuguese language, and mitigating the interpretive subjectivity inherent to topic labeling. These limitations were partially addressed through the integration of multilingual embeddings, the use of instruction-tuned LLMss, and careful manual validation, which together enhanced the robustness and analytical depth of the study.

Regarding RQ1, hierarchical aggregation enabled the identification of two major discursive dynamics: (1) ideological and public, and (2) political, electoral, and parliamentary. These two forces, while distinct, are deeply interconnected, reflecting the dual logic of far-right political communication-the need to build institutional legitimacy and the strategic exploitation of identity, emotion, and polarization.

The political and electoral dimension accounted for almost half of the corpus and remained very consistent over time. The discourse progressively evolved into a sustained narrative of opposition, institutional confrontation, and campaign communication. The highest levels of activity consistently coincided with the electoral cycles, suggesting that social media were used not only as a channel of communication but also as a strategic amplifier of political presence. Themes such as corruption, government criticism, and

political scandals were prominent, often serving as vehicles to challenge established elites and reinforce an anti-system identity.

The ideological and public dimension, representing nearly a third of the corpus, showed marked intensification from 2023 onward, dominated by narratives related to national identity, immigration, security, and moral and religious values, mixed with typical digital interaction in this political sphere. These topics were frequently associated with high levels of emotional expression, especially anger, hate, and disgust, and were often structured around antagonistic framings of 'us' versus 'them'. This shift suggests the growing reliance of the party on cultural and identity-based issues as mobilization tools, aligning with the broader trends observed in the far-right discourse in Europe.

Automatic text classification (zero-shot) analysis added additional interpretive depth by mapping the emotional, rhetorical, and ideological layers embedded within the discourse. Emotions such as anger, disgust, and hate were the most frequently expressed in all thematic domains, underscoring their centrality as mobilizing forces. In contrast, conciliatory emotions such as guilt or shame were almost absent, indicating a preference for confrontational, rather than reconciliatory, modes of engagement. Rhetorical analysis suggested a consistent reliance on devices such as irony, call to action, hyperbole, and metonymy, which serve to ridicule opponents, exaggerate threats, and symbolically condense complex social phenomena into emotionally resonant images. The widespread appearance of other rhetorical devices that, such as intertextuality, rhetorical questions, comparison, and appeals to fear, further suggests how discourse is designed not just to inform, but to provoke, and mobilize. Offensive and indirect hate speech were the most frequent types of expression, suggesting that political figures deliberately avoid explicit forms of hostility that could damage their public legitimacy while still conveying exclusionary messages in more subtle and coded ways. Strategically, the results also suggested that the discourse rely heavily on mechanisms such as symbolic threat, stereotyping, denial of hate, and silencing. These strategies function to legitimize exclusionary narratives while deflecting accusations of intolerance. For example, minorities are often portrayed not as individuals but as generalized figures of criminality or cultural threat, while alternative perspectives, such as those related to colonial memory or reparations, are downplayed or discredited. These findings align with the existing literature on far-right communication, which suggests that implicit, emotionally charged, and morally framed narratives are more effective in mainstreaming radical ideas than overt hate speech.

7.1. Limitations

The combination of these techniques allowed the study to move beyond descriptive topic modeling toward a richer, multilayered interpretation of far-right political communication. However, it is important to acknowledge the potential biases in this study. Relying on a single data source introduces selection bias: the focus on some accounts of CHEGA key figures excludes interactions with the rest of the party figures, supporters, and counter-discourses. Additionally, the use of monthly snapshots limits the temporal

resolution of the analysis, preventing a more fine-grained estimation of causal relationships between offline events and topic fluctuations. Although it was attempted to minimize human-introduced bias in the topic generation process (e.g., by supporting decisions with quantitative metrics such as c_v coherence and lexical diversity, and by using an LLMs to generate short descriptions of topics), manual curation and interpretation of topics remains subjective and susceptible to bias.

In the *BERTopic* results derived from hierarchical clustering, a substantial proportion of cases (23%) were classified as outliers, grouped under the category *'Others'*. When included in the detailed analysis, this subdimension revealed patterns closely aligned with those of the other clusters. However, its considerable size justifies deeper investigation or the application of additional techniques to refine its structure and extract more meaningful insights from the texts it encompasses.

7.2. Future Work

Future work should focus on expanding the sample and triangulating across platforms by incorporating data from additional social networks to assess narrative consistency and detect cross-platform variation. Including replies, for instance, would help map the interactional ecology and organic reach of the messages. Estimating causal links between discursive peaks and real-world events remains a critical objective to understand the dynamics at play. An analysis was conducted from a broader perspective on the results, complemented by a more focused examination of the subdimension *'National Identity, Tradition, Immigration and Security* within the context of ideological dynamics dimension. However, further analysis will be required to achieve a more comprehensive and in-depth understanding of the remaining subdimensions, ultimately contributing to the development of an interactive and comprehensive artifact that integrates the complete topic and classification analysis.

7.3. Ethics Statement

It is acknowledged that analyzing this type of speech can be a sensitive issue; for this reason, we present the results of our experiments in a responsible and appropriate manner. Additionally, we emphasize that language models trained on large datasets can reproduce or amplify biases and discrimination. To mitigate these risks, we seek to carefully select and evaluate the models and verbalizers used, ensuring the promotion of impartiality and fairness.

7.4. Contributions

This study is important not only for its contribution to the academic area, but also for its political implications. The study's findings may be used to develop legislation and methods for effectively combating polarized speech, therefore building a healthier and more inclusive political atmosphere, being critical in combating its negative and manipulative impact on Portuguese society.

References

[1] C. Rodríguez-Aguilera, "The rise of the far right in europe," *IEMed Mediterranean Yearbook*, 2014. [Online]. Available: https://www.iemed.org/publication/the-rise-of-the-far-right-in-europe/.

[2] M. Golder, "Explaining Variation in the Success of Extreme Right Parties in Western Europe," *Comparative Political Studies*, vol. 36, no. 4, pp. 432–466, 2003. DOI: 10.1177/0010414003251176.

[3] P. Carvalho, B. Cunha, R. Santos, F. Batista, and R. Ribeiro, "Hate speech dynamics against african descent, roma and lgbtqi communities in portugal," in *Proceedings of the Thirteenth Language Resources and Evaluation Conference*, 2022, pp. 2362–2370.

[4] E. Commission, *Code of conduct on countering illegal hate speech online*, 2019. [Online]. Available: https://commission.europa.eu/strategy-and-policy/policies/justice-and-fundamental-rights/combatting-discrimination/racism-and-xenophobia/eu-code-conduct-countering-illegal-hate-speech-online_en.

[5] C. Silva and P. Carvalho, "When can compliments and humour be considered hate speech? a perspective from target groups in portugal," *Comunicacao e Sociedade*, vol. 43, 2023, ISSN: 21833575. DOI: 10.17231/COMSOC.43(2023).4135.

[6] T. Lapa and B. di Fátima, "Hate Speech Among Security Forces in Portugal," *Communication Library*, vol. 7, pp. 277–293, 2023. DOI: 10.25768/654-916-9.

[7] T. Davidson, D. Warmsley, M. Macy, and I. Weber, "Automated Hate Speech Detection and the Problem of Offensive Language," in *Proceedings of the International AAAI Conference on Web and Social Media (ICWSM)*, 2017, pp. 512–515. DOI: 10.1609/icwsm.v11i1.14955.

[8] M. Caiani and P. Kröll, "The Transnationalization of the Extreme Right and the Use of the Internet," *International Journal of Comparative and Applied Criminal Justice*, vol. 39, no. 4, pp. 331–351, 2014. DOI: 10.1080/01924036.2014.973050.

[9] W. Bello, "The far right: Formidable but not unbeatable," *Agrarian South*, vol. 9, pp. 388–398, 3 Dec. 2020, ISSN: 23210281. DOI: 10.1177/2277976020968318.

[10] A. Rodríguez-Pose, J. Terrero-Dávila, and N. Lee, "Left-behind versus unequal places: Interpersonal inequality, economic decline and the rise of populism in the usa and europe," *Journal of Economic Geography*, vol. 23, pp. 951–977, 5 Sep. 2023, ISSN: 14682710. DOI: 10.1093/jeg/lbad005.

[11] W. Yu, Z. Chen, X. Meng, and Q. Yan, "Propagating covid-19 conspiracy theories: The influence of right-wing sources," *SAGE Open*, vol. 14, 2 Apr. 2024, ISSN: 21582440. DOI: 10.1177/21582440241258026.

[12] M. Krzyżanowski and M. Ekström, "The normalization of far-right populism and nativist authoritarianism: Discursive practices in media, journalism and the wider public sphere/s," *Discourse and Society*, vol. 33, pp. 719–729, 6 Nov. 2022, ISSN: 14603624. DOI: 10.1177/09579265221095406.

[13] G. Ramos, F. Batista, R. Ribeiro, P. Fialho, S. Moro, A. Fonseca, R. Guerra, P. Carvalho, C. Marques, and C. Silva, "A comprehensive review on automatic hate speech detection in the age of the transformer," *Social Network Analysis and Mining*, vol. 14, p. 204, 1 2024, ISSN: 1869-5469. DOI: 10.1007/s13278-024-01361-3.

[14] R. B. Santos, B. C. Matos, P. Carvalho, F. Batista, and R. Ribeiro, "Semi-supervised annotation of portuguese hate speech across social media domains," in *OpenAccess Series in Informatics*, vol. 104, Schloss Dagstuhl- Leibniz-Zentrum fur Informatik GmbH, Dagstuhl Publishing, Aug. 2022, ISBN: 9783959772457. DOI: 10.4230/OASIcs.SLATE.2022.11.

[15] K. M. Kotsifakou and D. N. Sotiropoulos, "Greek political speech classification using bert," in *14th International Conference on Information, Intelligence, Systems and Applications, IISA 2023*, Institute of Electrical and Electronics Engineers Inc., 2023, ISBN: 9798350318067. DOI: 10.1109/IISA59645.2023.10345868.

[16] G. Ramos, F. Batista, R. Ribeiro, P. Fialho, S. Moro, A. Fonseca, R. Guerra, P. Carvalho, C. Marques, and C. Silva, "Leveraging transfer learning for hate speech detection in portuguese social media posts," *IEEE Access*, vol. 12, pp. 101 374–101 389, 2024, ISSN: 21693536. DOI: 10.1109/ACCESS.2024.3430848.

[17] P. Chapman, "Crisp-dm 1.0: Step-by-step data mining guide," 2000. [Online]. Available: https://api.semanticscholar.org/CorpusID:59777418.

[18] I. Corp, *Ibm spss modeler crisp-dm guide*, 2021. [Online]. Available: https://www.ibm.com/docs/it/SS3RA7_18.3.0/pdf/ModelerCRISPDM.pdf.

[19] S. N. Nourbakhsh, S. A. Ahmadi, Q. Y. Dero, and A. F. Rad, "Rise of the Far Right Parties in Europe: from Nationalism to Euroscepticism," *Geopolitics Quarterly*, vol. 18, no. 4, pp. 47–70, 2021. DOI: 20.1001.1.17354331.1401.18.68.3.5.

[20] R. Koopmans and S. Olzak, "Discursive opportunities and the evolution of right-wing violence in germany," *American Journal of Sociology*, vol. 110, no. 1, pp. 198–230, 2004, ISSN: 00029602, 15375390. DOI: 10.1086/386271. (visited on 02/16/2025).

[21] M. Wahlström and A. Törnberg, "Social Media Mechanisms for Right-Wing Political Violence in the 21st Century: Discursive Opportunities, Group Dynamics, and Co-Ordination," *Terrorism and Political Violence*, vol. 33, no. 4, pp. 766–787, 2021. DOI: 10.1080/09546553.2019.1586676.

[22] T. Völker and D. S. Gonzatti, "Discourse networks of the far right: How far-right actors become mainstream in public debates," *Political Communication*, vol. 41, pp. 353–372, 3 2024, ISSN: 10917675. DOI: 10.1080/10584609.2024.2308601.

[23] L. Manucci, "Portuguese Populism: People, Parties, and Politics," *Análise Social*, vol. 59, no. 251, 2024. DOI: 10.31447/202200.

[24] R. Marchi and J. P. Zúquete, "Far Right Populism in Portugal: The Political Culture of Chega's Members," *Análise Social*, vol. 59, no. 251, 2024. DOI: 10.31447/2022116.

[25] J. Jerónimo, "Comunicação na Era da Desinformação: o Crescimento da Extrema-Direita e a Iliteracia Digital," *The Trends Hub*, vol. 1, no. 4, 2024. DOI: 10.34630/tth.vi4.5683.

[26] R. Almodt, "The Right-Wing Perspective: Populist Frames and Agenda on Facebook in Central and Eastern Europe," *Central European Journal of Communication (CEJC)*, vol. 15, no. 3(32), pp. 434–463, 2023. DOI: 10.51480/1899-5101.15.3(32).6.

[27] G. Newth, "'Talking About' the Far Right and Common Sense. A Case Study of Matteo Salvini's Buon Senso Trope on Twitter (2018–2023)," *Acta Politica*, vol. 60, pp. 361–384, 2025. DOI: 10.1057/s41269-023-00327-1.

[28] O. Klein and J. Muis, "Online Discontent: Comparing Western European Far-Right Groups on Facebook," *European Societies*, vol. 21, no. 4, pp. 540–562, 2019. DOI: 10.1080/14616696.2018.1494293.

[29] A. Lüders, A. Dinkelberg, and M. Quayle, "Becoming "us" in Digital Spaces: How Online Users Creatively and Strategically Exploit Social Media Affordances to Build Up Social Identity," *Acta Psychologica*, vol. 228, 2022. DOI: 10.1016/j.actpsy.2022.103643.

[30] B. C. Matos, R. B. Santos, P. Carvalho, R. Ribeiro, and F. Batista, "Comparing different approaches for detecting hate speech in online portuguese comments," in *OpenAccess Series in Informatics*, vol. 104, Schloss Dagstuhl- Leibniz-Zentrum fur Informatik GmbH, Dagstuhl Publishing, Aug. 2022, ISBN: 9783959772457. DOI: 10.4230/OASIcs.SLATE.2022.10.

[31] P. Carvalho, D. Caled, C. Silva, F. Batista, and R. Ribeiro, "The expression of hate speech against afro-descendant, roma, and lgbtq+ communities in youtube comments," *Journal of Language Aggression and Conflict*, vol. 12, no. 2, pp. 171–206, 2024, ISSN: 2213-1272. DOI: https://doi.org/10.1075/jlac.00085.car.

[32] A. Fatima-Azzahrae, A. Rkia, A. Mehdi, and L. Lily, "Automatic classification of psychosocial concerns: From traditional approach to deep learning," in *Procedia Computer Science*, vol. 251, Elsevier B.V., 2024, pp. 390–397. DOI: 10.1016/j.procs.2024.11.125.

[33] N. Nagendra and J. Chandra, "Hybrid approach for multi-classification of news documents using artificial intelligence," in *Proceedings - 2024 5th International Conference on Intelligent Communication Technologies and Virtual Mobile Networks,*

ICICV 2024, Institute of Electrical and Electronics Engineers Inc., 2024, pp. 466–473, ISBN: 9798350385649. DOI: 10.1109/ICICV62344.2024.00079.

[34] T. M. Doan, B. Kille, and J. A. Gulla, "Using language models for classifying the party affiliation of political texts," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 13286 LNCS, pp. 382–393, 2022, Cited by: 2. DOI: 10.1007/978-3-031-08473-7_35.

[35] S. Modha, T. Mandl, P. Majumder, and D. Patel, "Tracking hate in social media: Evaluation, challenges and approaches," *SN Computer Science*, vol. 1, 2 Mar. 2020, ISSN: 26618907. DOI: 10.1007/s42979-020-0082-0.

[36] J. Torregrosa, G. Bello-Orgaz, E. Martinez-Camara, J. D. Ser, and D. Camacho, "A Survey on Extremism Analysis using Natural Language Processing," *Computing Research Repository*, vol. arXiv:2104.04069, 2021. DOI: 10.48550/arXiv.2104.04069.

[37] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent Dirichlet Allocation," *Journal of Machine Learning Research (JMLR)*, vol. 3, pp. 993–1022, 2003. DOI: 10.5555/944919.944937.

[38] D. Lopes-Teixeira, F. Batista, and R. Ribeiro, "Discovering Trends in Brand Interest through Topic Models," in *Proceedings of the International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management (IC3K)*, 2018, pp. 245–252. DOI: 10.5220/0006936202450252.

[39] M. Siino, I. Tinnirello, and M. L. Cascia, "Is text preprocessing still worth the time? a comparative survey on the influence of popular preprocessing methods on transformers and traditional classifiers," *Information Systems*, vol. 121, Mar. 2024, ISSN: 03064379. DOI: 10.1016/j.is.2023.102342.

[40] R. Egger and J. Yu, "A Topic Modeling Comparison Between LDA, NMF, Top2Vec, and BERTopic to Demystify Twitter Posts," *Frontiers in Sociology*, vol. 7, p. 886 498, 2022. DOI: 10.3389/fsoc.2022.886498.

[41] M. Grootendorst, "BERTopic: Neural Topic Modeling with a Class-based TF-IDF Procedure," *Computing Research Repository*, vol. arXiv:2203.05794, 2022. DOI: 10.48550/arXiv.2203.05794.

[42] N. C. Hellwig, J. Fehle, M. Bink, T. Schmidt, and C. Wolff, "Exploring Twitter Discourse with BERTopic: Topic Modeling of Tweets Related to the Major German Parties during the 2021 German Federal Election," *International Journal of Speech Technology*, vol. 27, pp. 901–921, 2024. DOI: 10.1007/s10772-024-10142-4.

[43] A. Webson and E. Pavlick, "Do prompt-based models really understand the meaning of their prompts?" In *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, M. Carpuat, M.-C. de Marneffe, and I. V. Meza Ruiz, Eds., Seattle, United

States: Association for Computational Linguistics, Jul. 2022, pp. 2300–2344. DOI: 10.18653/v1/2022.naacl-main.167.

[44] F. M. Plaza-del-arco, D. Nozza, and D. Hovy, "Respectful or toxic? using zero-shot learning with language models to detect hate speech," in *The 7th Workshop on Online Abuse and Harms (WOAH)*, Y.-l. Chung, P. R{\"ottger}, D. Nozza, Z. Talat, and A. Mostafazadeh Davani, Eds., Toronto, Canada: Association for Computational Linguistics, Jul. 2023, pp. 60–68. DOI: 10.18653/v1/2023.woah-1.6.

[45] N. Bauer, M. Preisig, and M. Volk, "Offensiveness, hate, emotion and GPT: Benchmarking GPT3.5 and GPT4 as classifiers on Twitter-specific datasets," in *Proceedings of the Fourth Workshop on Threat, Aggression & Cyberbullying @ LREC-COLING-2024*, R. Kumar, A. K. Ojha, S. Malmasi, B. R. Chakravarthi, B. Lahiri, S. Singh, and S. Ratan, Eds., Torino, Italia: ELRA and ICCL, May 2024, pp. 126–133. [Online]. Available: https://aclanthology.org/2024.trac-1.14/.

[46] J. Jaremko, D. Gromann, and M. Wiegand, "Revisiting implicitly abusive language detection: Evaluating LLMs in zero-shot and few-shot settings," in *Proceedings of the 31st International Conference on Computational Linguistics*, O. Rambow, L. Wanner, M. Apidianaki, H. Al-Khalifa, B. D. Eugenio, and S. Schockaert, Eds., Abu Dhabi, UAE: Association for Computational Linguistics, Jan. 2025, pp. 3879–3898. [Online]. Available: https://aclanthology.org/2025.coling-main.262/.

[47] X. Liu, M. Zhou, X. Zhou, Z. Fu, and F. Wei, "Joint Inference of Named Entity Recognition and Normalization for Tweets," in *Proceedings of the Annual Meeting of the Association for Computational Linguistics (ACL)*, vol. 1: Long Papers, 2012, pp. 526–535. [Online]. Available: https://aclanthology.org/P12-1055/.

[48] F. Feng, Y. Yang, D. Cer, N. Arivazhagan, and W. Wang, "Language-agnostic BERT Sentence Embedding," in *Proceedings of the Annual Meeting of the Association for Computational Linguistics (ACL)*, vol. 1: Long Papers, 2022, pp. 878–891. DOI: 10.18653/v1/2022.acl-long.62.

[49] N. Reimers and I. Gurevych, "Making Monolingual Sentence Embeddings Multilingual using Knowledge Distillation," in *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2020, pp. 4512–4525. DOI: 10.18653/v1/2020.emnlp-main.365.

[50] L. Gomes, A. Branco, J. Silva, J. Rodrigues, and R. Santos, "Open Sentence Embeddings for Portuguese with the Serafim PT* Encoders Family," in *Proceedings of the EPIA Conference on Artificial Intelligence*, 2024, pp. 267–279. DOI: 10.1007/978-3-031-73503-5_22.

[51] L. McInnes, J. Healy, and J. Melville, "UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction," *Computing Research Repository*, vol. arXiv:1802.03426, 2018. DOI: 10.48550/arXiv.1802.03426.

[52] R. J. G. B. Campello, D. Moulavi, and J. Sander, "Density-Based Clustering Based on Hierarchical Density Estimates," in *Proceedings of the Pacific-Asia Conference*

on Knowledge Discovery and Data Mining (PAKDD), 2013, pp. 160–172. DOI: 10.1007/978-3-642-37456-2_14.

[53] L. McInnes, J. Healy, and S. Astels, "HDBSCAN: Hierarchical Density Based Clustering," *The Journal of Open Source Software*, vol. 2, no. 11, p. 205, 2017. DOI: 10.21105/joss.00205.

[54] L. McInnes and J. Healy, "Accelerated Hierarchical Density Clustering," in *Proceedings of the IEEE International Conference on Data Mining Workshops (ICDMW)*, 2017, pp. 33–42. DOI: 10.1109/ICDMW.2017.12.

[55] M. Röder, A. Both, and A. Hinneburg, "Exploring the Space of Topic Coherence Measures," in *Proceedings of the ACM International Conference on Web Search and Data Mining (WSDM)*, 2015, pp. 399–408. DOI: 10.1145/2684822.2685324.

[56] A. B. Dieng, F. J. R. Ruiz, and D. M. Blei, "Topic Modeling in Embedding Spaces," *Transactions of the Association for Computational Linguistics*, vol. 8, pp. 439–453, 2020. DOI: 10.1162/tacl_a_00325.

[57] M. de Groot, M. Aliannejadi, and M. R. Haas, "Experiments on Generalizability of BERTopic on Multi-Domain Short Text," in *Widening Natural Language Processing (WiNLP)*, 2022. DOI: 10.48550/arXiv.2212.08459.

[58] C. Pontes, A. Fonseca, S. Moro, F. Batista, R. Ribeiro, C. Marques, P. Carvalho, C. Silva, and R. Guerra, "Unveiling patterns of hate speech in the portuguese sphere: A social network analysis approach," in Feb. 2025, pp. 70–81, ISBN: 978-3-031-73996-5. DOI: 10.1007/978-3-031-73997-2_7.

[59] P. Fialho, R. Ribeiro, F. Batista, G. Ramos, A. Fonseca, S. Moro, R. Guerra, P. Carvalho, C. Marques, and C. Silva, "Counter hate speech detection in youtube conversations," in Feb. 2025, pp. 94–105, ISBN: 978-3-031-73996-5. DOI: 10.1007/978-3-031-73997-2_9.

[60] OpenAI et al., "GPT-4o System Card," *Computing Research Repository*, vol. arXiv:2410.21276, 2024. DOI: 10.48550/arXiv.2410.21276.

[61] OpenAI et al., "GPT-4 Technical Report," *Computing Research Repository*, vol. arXiv:2303.08774, 2023. DOI: 10.48550/arXiv.2303.08774.

[62] F. Baider and M. Constantinou, "Covert hate speech: A contrastive study of greek and greek cypriot online discussions with an emphasis on irony," *Journal of Language Aggression and Conflict*, vol. 8, pp. 262–287, Oct. 2020. DOI: 10.1075/jlac.00040.bai.

APPENDIX A

Extended Topic Definitions Tables

FIGURE A.1. *'Political Discourse'* thematic hierarchy map.

1.: Others: — **T:** -1 **D:** 2382
2.: Ideological and Public Dynamics
- **2.1:** Public Order, Clash of Values, Religious and Identitarian Issues
  - **2.1.1:** Conservative Approach, Public Order, Nationalism/Patriotism and Traditional Values
    - **2.1.1.1:** Social, Political Accountability and Dissatisfaction — **T:** 45, 47 **D:** 130
    - **2.1.1.2:** National Identity and Tradition, Ethnic Issues, Immigration and Security — **T:** 7, 12, 14, 21, 39, 41, 43, 48 **D:** 1017
  - **2.1.2:** Ideological Positions, Moral, Sexual and Religious Debates
    - **2.1.2.1:** Sexual Crimes, Feminism/Abortion, Religious Education and Gender/Sexuality Debates — **T:** 24, 27, 44, 46 **D:** 357
    - **2.1.2.2:** Religious Conflicts, Political Ideologies, Extreme Disputes, Relations between Belief, Morals and Politics — **T:** 8, 32, 34, 50 **D:** 440
- **2.2:** Discourse of Social and Digital Interaction, Internet Colloquial Tone — **T:** 5, 6, 29, 35, 37, 38, 40, 57 **D:** 1037

3.: Party, Electoral and Parliamentary Dynamics
- **3.1:** Governance, Economy and Social Policies, Public Health
  - **3.1.1:** Public Management, Health, Investigations/Scandals and Power Conflicts — **T:** 22, 30, 31, 33 **D:** 390
  - **3.1.2:** Socio-Economic Criticism and Debate — **T:** 11, 16, 23, 51 **D:** 539
  - **3.1.3:** Controversies and Disputes about Political Leadership and Government Management — **T:** 53, 56, 58 **D:** 165
- **3.2:** Political-Electoral Dynamics, Public Management, Ideological Debates and Media Coverage
  - **3.2.1:** Political Dispute, Ideological Positions, Communication and Media — **T:** 0, 1, 2, 3, 4, 10, 15, 18, 25, 26, 36, 42, 52, 54, 55 **D:** 2859
  - **3.2.2:** Political-Parliamentary and Electoral Cycle, Mobilization, Interaction and Event Organization — **T:** 9, 13, 17, 19, 20, 28, 49 **D:** 1007

TABLE A.1. Topic-model labels assigned to the 59 *BERTopic* clusters.

ID	Definition
-1	Others
0	Communication, the Press and Political Mobilization
1	Political Vandalism and Ideological Conflict
2	Corruption and Abuse of Power
3	Political Coalitions, Controversies and Political Responsibility
4	Media Events and Political Events and Campaigning
5	Informal Expressions and Personal Interaction
6	Online Debates and Interpersonal Criticism
7	Social Tensions and Immigration
8	Indoctrination and Ideological Polarization
9	Parliamentary Activities and Constitutional Review
10	Elections and Political Strategy
11	Tax Burden and Social Costs
12	Crime and Public Safety
13	Political Events and Campaigning
14	Immigration and Border Control
15	European Right and Political Coordination
16	Economic and Social Crisis
17	National Identity and Patriotism, Youth and Education
18	Political and Institutional Conflicts
19	Agriculture, Fisheries and Rural Development
20	Local and Regional Elections
21	Family, Faith, Celebrations and Religion
22	Crisis and Collapse in Health
23	Corruption, Public Management and Clientelism
24	Criminal Justice, Violence and Crime
25	Alternative, Criticism and Political Confidence
26	European Right and National Identity
27	Gender Equality and Feminist Dynamics
28	Elections, Regional Campaigns and International Far-Right Movements
29	Insults and Ideological Conflicts, Disinformation and Public Exposure
30	Diplomacy, Budget and Government Relations
31	Governance, Controversies and Political Responsibility
32	Patriotism, Controversy and Mobilization
33	Opposition, Criticism and Political Management
34	Ideologies and Individual Freedoms
35	Social Interactions, Congratulations and Emotions
36	Controversies and Political Scandals
37	Expressions, Rhetoric and Emotions
38	Racism, Public Opinion and Prejudice
39	Public Safety, Crime and Minorities
40	Disinformation, Defamation and Social Networks
41	Colonialism, Historical Memory and Reparations
42	European Politics, Territorial Action and Institutional Participation
43	Patriotism, National Identity and Historical Memory
44	Indoctrination, Religion and Moral Education
45	Protest, Indignation and Political Outcry
46	Sexuality, Ideology and Child Protection
47	Corruption, Impunity and Criticism of the Political System
48	National Honor, State Careers and Patriotic Duty
49	Public Space, Mobility and Local Identity
50	Religious Conflicts and Terrorism
51	Justice, Cost of Living and Economy, Social Exclusion
52	Elections and Political Participation
53	Government Crisis and Public Administration
54	Political Participation and Institutional Action
55	Political Mobilization and Campaign Rhetoric
56	Public Health and the Hospital System
57	Media, Journalism and Television Representation
58	Participation and Political Figures

TABLE A.2. Topics under 2.1.1.1 - Social, Political Accountability and Dissatisfaction.

ID	Topic Definition	Top 50 Keywords
45	Protest, Indignation and Political Outcry	decente, arder, avisámos, sentir, façam, metade, entrar, acontece, pobreza, deixem, declarar, gritante, controlo, deixarem, miséria, bandalheira, caminha, asilo, mudamos, tsf, iniciou, estariam, continuem, estatística, nojo, sol, tejo, viagens, demitido, faltar, indignado, descanso, anular, estimação, vaise, ceder, firmes, pedem, proprietários, virem, governante, impunidade, atacado, autêntica, incomodar, critério, farão, chegamos, venham, deixamos, estupidez, ilhas, tentam, instituição, brincadeira, inimigo, pátria, ouviram, massa, deixámos
47	Corruption, Impunity and Criticism of the Political System	ladrão, vendemos, coligações, bandido, mexe, trabalharam, machico, demonstrado, bandalheira, mudanças, metadados, mira, terminar, darão, relatórios, território, imoral, arruada, vende, cúmulo, caminhada, calhar, perdão, contam, habituemse, ouve, governa, acompanhado, trás, absurdo, residência, prisão, pôr, fugir, contrário, falou, seguro, governar, restrições, acreditar, tolerância, dificuldades, sexo, demasiado, subsidiodependência, vir, tremer, sermos, outdoor, acabar, impunidade, passos, noção, privilégios, dúvidas, aconteceu, crescer, absurdas, leiria, vergonhoso

TABLE A.3. Topics under 2.1.1.2 - National Identity and Tradition, Ethnic Issues, Immigration and Security.

ID	Topic Definition	Top 50 Keywords
7	Social Tensions and Immigration	veremos, confiar, martim, moniz, oferecer, envergonham, atlântico, ex, seleção, limpo, limpeza, nepalês, louvor, paris, lembrome, policial, juntar, acordar, suíça, convosco, terminamos, amadora, ficaria, contamos, prontos, resistir, podridão, traficantes, mostram, hino, comunidade, destino, revolta, segredo, falava, ferro, sai, bocado, indivíduos, merecem, fraudes, frança, anedota, ocasião, abortar, marine, básicos, pais, islamização, pedras, rainha, símbolo
12	Crime and Public Safety	terror, faca, assaltos, adulto, tiroteio, esfaqueou, multiculturalismo, imigrante, moradores, indivíduo, islâmico, permitimos, incêndio, chamas, rixa, afegão, provocam, agressões, feridos, policial, insegurança, destruídos, armado, incendiários, agredir, orações, ourém, pornografia, plena, drogas, consumo, atirados, martim, moniz, terrorista, fumar, acontecem, ferida, islamização, metro, homicídio, brutalmente, violentos, cenário, agressor, bárbara, agente, infantil, ignoram, droga, psp
14	Immigration and Border Control	lampedusa, ilegais, deportação, aima, imigrantes, integração, imigrante, deportar, migração, massiva, tendas, quotas, refugiados, agência, descontrolados, espadas, asilo, facas, desembarque, refugiado, gratuito, migrações, migrantes, habitantes, apresentarem, dupla, entram, seguras, pacto, indiano, legalização, pedidos, controlada, utilizam, descontrolada, descontrolo, afegãos, idade, regras, chegada, fronteiras, fiscalização, zonas, alterar, chegam, ilegal, nacionalidade, chegaram, recorde, encontram
21	Family, Faith, Celebrations and Religion	brotherhood, tabuleiros, familyparty, happyfamily, fatherhood, familianumerosa, brothers, amam, catedral, parentalidade, family, santo, califórnia, festas, ventre, toiro, espírito, áfrica, irmãos, bispo, salomé, espada, mães, alegria, jesus, magnífica, colegas, natal, jerusalém, estátua, humildes, fátima, companheiros, international, nascimento, santidade, desempenha, paixão, reta, partidária, santuário, activista, dedicação, henriques, palestra, muçulmanos, emocional, católico, empresarial, cartaxo
39	Public Safety, Crime and Minorities	ciganos, magistrados, policiais, criminalidade, elogiar, gnr, seguro, monsaraz, negro, cigano, estatísticas, agir, probabilidade, crimes, violentos, reguengos, polícia, polícias, ligações, criminosos, bandidos, autoridade, publicações, luxo, agressores, agressões, mortos, diariamente, resistiremos, delinquentes, referente, trata, nacionalidade, acontecem, seixal, serem, culpas, coitadinho, homicídio, cigana, colonial, escravatura, prisionais, morto, erro, bandido, nulo, forças, arriscar, perderem, meios
41	Colonialism, Historical Memory and Reparations	cartoon, excolónias, orgulha, rejeitar, traidores, lgbti, desculpa, domínio, mansão, territórios, obscuros, humilhar, vergonhosa, antigas, degradação, antepassados, indemnizações, armadas, impostos, moçambique, desrespeito, forças, governa, pagar, promove, contentes, manter, colapso, corruptos, segurança, tratados, ações, pedimos, protege, pediu, ana_gomes, agem, insultar, compensações, online, cultural, colonial, paga, ilegal, imóveis, sintra, garantir, humanos, prec, polícia, fatura
43	Patriotism, National Identity and Historical Memory	milenar, históricos, lutam, tradições, agente, cravos, camões, dou, colectiva, antepassados, carneiro, perdida, amamos, psp, necessárias, desculpa, brutalmente, sá, portugalidade, símbolos, comunidades, palhaçada, lutaram, identidade, medalha, agredido, espírito, feitos, pátria, orgulho, portuguesas, heróis, hino, fantasma, espadas, protege, serviu, londres, veste, continuarmos, endémica, metro, racismo, independência, escumalha, fábio, deixarmos, bonitos, pergunta, salazar
48	National Honor, State Careers and Patriotic Duty	feriado, reivindicações, queridos, divulgação, policiais, militares, justa, inspiração, novembro, homenagem, entidades, exército, militar, agradecimento, prometido, heróis, sacrifício, protesto, lutaram, carreiras, combatentes, forças, lealdade, data, altas, bombeiros, antigos, empenhados, segurança, corpo, isabel, celebrar, esquecidos, básicos, ideologias, totalitarismo, membro, rainha, prisionais, castelo, seguido, delito, particularmente, operacional, proteção, humana, pátria, armadas, testemunhar, patriotismo

TABLE A.4. Topics under 2.1.2.1 - Sexual Crimes, Feminism/Abortion, Religious Education and Gender/Sexuality Debates.

ID	Topic Definition	Top 50 Keywords
24	Criminal Justice, Violence and Crime	suspensas, antimulher, alargamento, homicidas, química, castrados, penas, pedófilo, perpétua, prescrição, castração, graciosa, violadores, penal, abusadores, viola, pedófilos, prazos, monstros, balneários, expulsão, julgamento, defendo, agravamento, agressor, fábio, prisão, punir, agredida, inocentes, detenção, crimes, incendiários, vandalismo, suicídio, suspensa, duras, condenado, redução, prevenção, castigo, código, violador, tipos, artigo, impunes, sexual, pedofilia, violação, cabo, suspeito, agressores, cuidados, presos, reconhecer, selo, sexuais, supremo, essencial, vítima
27	Gender Equality and Feminist Dynamics	abortos, biológicos, feministas, abortar, sozinhas, feto, tóxica, grávidas, feminista, sexos, aborto, femininas, competir, biologia, testemunhos, desportiva, igualdade, princípio, masculinidade, autodeterminação, irresponsável, globalista, feminina, desporto, perversidade, patriarcado, mães, pintar, inocência, probabilidade, movimentos, relevante, feminismo, presunção, casar, género, dep, ciência, conhecidas, regimes, terceiros, filhas, mera, afirma, tirania, feminino, inclusão, perseguem, joana, medicina, síndrome, defensores, mulher, homossexuais, humano, coxi, hipócritas, meninas, precisamente, arte
44	Indoctrination, Religion and Moral Education	culto, piadas, educar, religioso, religiosa, educativo, precoce, noticiar, retrocessos, ensina, abusadores, indigna, ombro, colectiva, religião, lavar, educação, intelectualmente, eutanásia, testa, professores, amar, conteúdos, inclui, supostamente, conheço, padre, igreja, revolução, criança, tradição, ideologia, igrejas, doutrinação, islâmica, peito, abusos, lgbt, quadro, escolher, química, argumento, fé, inocente, crianças, procura, sexual, género, digno, vítima, cartazes, escola, projecto, santo, assédio, testemunho, filhos, disciplina, castração, super
46	Sexuality, Ideology and Child Protection	sexualmente, lucrar, meninas, abusadores, colo, assalto, ética, violadores, agendas, tomou, renunciar, casar, digitais, cirurgias, ideologia, partilha, ensino, castração, género, sexuais, moral, química, menina, menores, conteúdos, loucura, penas, plataformas, conhecimento, chumbou, escolas, bloqueio, avança, cores, racial, pedófilos, resistir, crianças, doméstica, superior, assédio, agressão, reunidos, retiradas, ficarão, humanas, universidades, menino, tentativa, seres, sexos, etnia, reduz, estrutural, passava, penais, ideológicas, aliança, instrumentalizar, cancelado

TABLE A.5. Topics under 2.1.2.2 - Religious Conflicts, Political Ideologies, Extreme Disputes, Relations between Belief, Morals and Politics.

ID	Topic Definition	Top 50 Keywords
8	Indoctrination and Ideological Polarization	encetado, dont, verá, pesquisar, jihad, ovos, catecismo, rapazes, artigos, expressar, admiração, esquerdistas, of, culto, rita_matias, in, doutrina, machista, marxismo, uso, leaders, cristãos, feio, this, israel, indignado, from, the, fixação, almoços, rapariga, desvalorizam, tabaco, young, usei, gostei, audiências, is, olhem, interferir, marcante, humano, to, imt, adoram, pagamento, derrotar, selo, imposto, cardeal, privilegiados, mistas, cristo, marxista, contraditório, lagos, directo, homossexual, perseguidos, bilhetes
32	Patriotism, Controversy and Mobilization	resgatar, leaders, or, to, young, brasileiro, patriótica, referendo, the, palestina, curioso, pandémico, solução, pesar, esplendor, aprovado, of, estados, demonstra, pescoço, esfaquear, endireitar, instalado, exceção, ve, símbolo, explicar, cubano, financiado, ótima, this, bloqueada, is, patriotismo, festejar, divulgação, lula, sirva, veem, governado, pp, prof, ong, percebeu, preparados, tratado, ministérios, voto, lembremse, rita, computador, literalmente, mostra, prometer, envolvido, posto, defendemos, sis, lusa, israelita
34	Ideologies and Individual Freedoms	alheia, comunismo, vacinado, docentes, limitações, evangelho, expressa, animais, matou, nazismo, ensino, nazi, bonitas, foice, casasdebanho, expressão, centros, balneários, martelo, terra, duarte, feminista, destruição, vergonhas, ativistas, climáticos, felizes, tradições, espaços, comunistas, publicação, deveriam, terminou, cordeiro, liberdades, loucura, internet, marxista, maduro, cães, fiel, confesso, jesus, family, limitar, luz, mobilidade, cales, cuba, terrenos, inocência, estudei, judeus, respeitem, comunista, alegria, propriedade, apóstolo, presunção, limpo
50	Religious Conflicts and Terrorism	isis, massacre, reféns, cristã, gaza, idf, cristãos, israel, hamas, islâmico, terroristas, islâmicos, islâmica, terrorista, bárbaro, sequestrado, lançada, frio, palestinianos, hesitações, missa, ocidental, terrorismo, jesus, jerusalém, israelita, fundamentalismo, civis, armas, bebés, cardeal, árabes, cristo, sangue, reuni, fachada, perseguidos, animais, igrejas, religião, irmãos, cores, madrugada, palestino, coordenadora, matriz, palestinos, falámos, ataque, condenação, nasceu, refugiados, festa, palestina, idosos, jihad, domésticos, bebé, provou, dor

TABLE A.6. Topics under 2.2 - Discourse of Social and Digital Interaction, Internet Colloquial Tone.

ID	Topic Definition	Top 50 Keywords
5	Informal Expressions and Personal Interaction	maluco, apaguei, ahahah, apaga, carmo, not, baby, referências, abraço, iam, toca, convidar, ai, leia, consegues, procura, completa, link, chora, esforço, fontes, julgar, pés, amiga, detalhe, escrever, carlos, limite, excelentes, passamos, mando, nenhumas, cales, paixão, identifica, exato, financiado, falamos, grândola, pesquisar, mete, exatamente, repente, falo, lelo, olha, cautela, camarada, devidas, filomena, salvação, grato, farei, dou, ligar, concordo, actuais, fila, ah, passámos
6	Online Debates and Interpersonal Criticism	post, evoluído, argumentação, miúda, deves, print, mortal, putin, olhe, percebi, intelectual, adoro, esperava, personagem, perdida, percebido, chora, apaga, entende, correção, teste, ouve, falso, bloqueado, honestidade, site, cor, situações, produção, factos, bloqueio, consegues, destacado, hugo, visto, disponíveis, inocente, devias, olha, camilo, nomes, abraço, partilhar, protestos, senhora, carácter, palco, indispensável, ido, deportar, fugiu, chamem, incoerência, argumentos, falei, bio, episódios, extremamente, indecisos, isenção
29	Insults and Ideological Conflicts, Disinformation and Public Exposure	banalizam, arremessandoos, nom, petit, caluniosamente, vulgarizarem, horrível, fascista, facho, mentem, desvalorizam, partilhei, lusófona, perdem, crédito, vêse, louçã, insulto, produtos, petição, ponte, amantes, publicou, estudei, fake, feio, apóstolo, esquerdistas, desportiva, odeia, inês, bots, religiosa, propositadamente, fascismo, filha, rasgam, vestes, biden, queixa, conheço, minuto, privada, comportamento, fotografias, noticiado, intelectual, partilho, opinião, bilhetes, termo, publicamente, francisco, mecanismos, compreendem, dirigente, escreveu, aulas, odeiam, site
35	Social Interactions, Congratulations and Emotions	felicidades, publicações, sororidade, todes, dados, natal, verá, comentar, olá, ignorante, vivo, exposição, parabéns, apreço, jerusalém, entendem, beijinho, específico, fama, sabias, casamento, portal, motivos, estatística, bonitas, idiotas, segundos, desinformação, pombal, nojo, plenária, definição, bonita, beijinhos, sara, conhecimento, petição, ouvido, bemvindo, denúncias, termo, felicidade, ventos, proposta, privados, consultar, conto, filipe, alerta, úteis, senti, repetir, estudar, procurar, mensagem, obsessão, feliz, felizes, espaços, completa
37	Expressions, Rhetoric and Emotions	camilo, morrem, digo, chora, som, comentar, capa, dizes, novidade, nazis, ahahah, esqueceramse, nepalês, andre_ventura, louvor, gostem, dissemos, sorriso, créditos, perde, atribuir, junteime, entrevistado, fase, exclusividade, daniel, engolir, encéfalo, joaquim, teste, doer, doa, descer, imunidade, diploma, olhão, cego, nojo, vês, presença, discutir, agradar, prometido, ofende, chiado, voltaram, convosco, albuquerque, anúncio, bemvindo, pan, isolado, correcto, dizendo, sofrem, medalha, leiam, usei, acordos, irmãos
38	Racism, Public Opinion and Prejudice	racista, sabes, contraditório, xenófobos, salmo, lelo, cúmulo, infantil, intelectual, brancas, tweet, inteligência, afirmei, mental, dum, ofensa, razões, quaresma, bancadas, racistas, gonçalo, bater, necessidade, queimado, caricatura, acto, aborto, twitter, comentários, admite, perfeito, disparate, enviado, atacou, preconceitos, contexto, terrenos, amigo, falamos, precoce, ótima, lamenta, preconceito, tentarem, cumprimentos, monsaraz, toque, lembrete, carro, armado, entendimento, bloqueiam, lembreime, queimar, jogar, armadilha, preocupante, escrevo, assinaturas, ignorância
40	Disinformation, Defamation and Social Networks	montagem, difamação, fake, irónico, idiota, news, devias, torna, fakenews, entender, telejornais, cobarde, senso, sabiam, violação, tweet, falso, perseguição, twitter, vendido, comentário, mostraram, mentiras, chamado, herdeiro, plataforma, delito, salazar, decoro, atacam, humorista, dita, daquilo, sexual, continuou, link, floyd, suspenso, falsidades, miúda, ignorância, imagina, anónimo, rede, passámos, dando, crime, pembrolizumab, ditadores, confusão, esfaqueado, existisse, achas, lavar, acusados, ridículo, autodeterminação, tapete, desaparece, fundador
57	Media, Journalism and Television Representation	reportagens, missa, postura, entrevistado, acompanha, actualidade, federal, acompanhamento, isento, aguardar, jornalismo, televisão, sic, independentes, rádio, jornalista, cnn, observador, jornalistas, expresso, morre, condenação, estarei, jornais, entrevista, jesus, jornal, mantenho, televisivo, acompanhar, morreu, amanhã, evento, podes, activista, televisivos, conversão, daqui, próvida, tvi, disfarçar, debater, comentadores, perguntei, perde, diana, invés, estúdio, conclusão, nele, típico, direto, palácio, coxi, corresponde, tsf, cmtv, conservadores, brasileiros, sinto

TABLE A.7. Topics under 3.1.1 - Public Management, Health, Investigations/Scandals and Power Conflicts.

ID	Topic Definition	Top 50 Keywords
22	Crisis and Collapse in Health	urgências, morrerem, hospitais, ambulância, grávida, consultas, colapso, inem, médicos, médico, vmer, hospital, encerramentos, ppp, cirurgias, grávidas, utentes, caos, enfermeiros, demissões, reestruturação, terrível, assistência, encerrados, mortes, onehealth, bebé, recursos, diretor, temido, vagas, inferno, sns, marta, medicina, cannabis, ceo, mandada, estatutos, faltará, morreu, doentes, deficientes, veterinária, ambulâncias, etc, pasta, bastonária, médica, operacional, serpa, profissionais, serviços, morta, maternidade, especialista, competência, enfermeiro, cuidados, cuidar
30	Diplomacy, Budget and Government Relations	audições, méxico, embaixada, finanças, especialidade, audição, iniciamos, âmbito, relativas, votações, chefes, orçamento, embaixador, ministra, concentração, infraestruturas, ucrânia, soares, bruno, oe, secretária, amorim, moçambique, ciência, fundação, comissões, israel, medina, gabinete, solidariedade, galamba, assembleia, amizade, restante, diplomática, magnífico, comissão, visitamos, intervenção, pacheco, votação, audiência, chefe, convite, reerguer, ministro, fernando, diárias, teatro, privada, ensino, conferência, peça, preocupações, opções, leão, república, negócios, link, operações
31	Governance, Controversies and Political Responsibility	escária, repente, adjunto, mendonça, auditoria, advogado, interferir, relevo, projetolei, detidos, encontrado, marise, secretário, galamba, velho, desiludir, adn, motoristas, portosantenses, nomeou, mendes, sis, marido, diretos, mano, cumpre, empresa, demitido, gabinete, cúmplice, notas, pia, cargo, ajustes, preparar, alexandre, relatório, irmão, banco, antonio_costa, gastos, madeirenses, escândalo, ana, marcelo, interessa, chefe, condições, acreditam, renda, descontrolados, cidadão, familiares, cheios, desperdício, funcionários, demissão, imediata, recomendação, isabel
33	Opposition, Criticism and Political Management	luis_montenegro, supranacionais, insultado, luís, pandémico, alterne, colo, maravilhas, legado, custou, sanitário, pedrogão, regulamento, apoiam, derrota, albuquerque, tacho, limitar, chumba, tratado, subsidiodependentes, antonio_costa, reconhece, enganado, impressionante, soberania, melo, escândalos, imitar, pé, tentando, destruiu, máscara, fofinha, fizermos, escária, desastre, esconder, plano, confrontei, gostava, achamos, tavares, apoiar, malta, cama, portugalprecisadeumalimpeza, diplomático, vergamos, empenhado, conselhos, usam, democrata, gasolina, transferência, liberais, sentar, darlhe, indigna, dissolução

TABLE A.8. Topics under 3.1.2 - Socio-Economic Criticism and Debate.

ID	Topic Definition	Top 50 Keywords
11	Tax Burden and Social Costs	combustíveis, lucros, inflação, roubo, cobrar, pensões, isenções, veículos, gasóleo, multa, pagam, preço, bancos, consumidores, suportar, pagamos, aumentos, património, fiscais, preços, imi, custam, beneficiam, erário, baixar, contribuintes, impostos, empresas, isentos, taxar, recusa, caça, aumentar, benefícios, carga, inundar, gamanço, aumenta, banca, percebemos, apertar, efetivamente, salários, planeamento, subiu, milionárias, nazis, artista, dignos, fiscal, níveis, pagar, transportes, migalhas, paga, rendimentos, miseráveis, admissível, enormes, taxas
16	Economic and Social Crisis	rendas, pensões, coreia, salários, gastamos, renda, miseráveis, pensionistas, asfixia, empregos, banca, proprietários, empresas, baixos, subsídios, lucros, estarmos, recebe, crise, gasta, afegãos, fortunas, pagamos, empresários, fiscalizar, sobreviver, prestações, fundações, salário, expolíticos, polvo, refugiados, escravatura, imóveis, cubano, rendimento, gastar, achamos, vitalícias, famílias, electricidade, trabalhou, gnr, idosos, médica, subvenções, comida, vivem, ucraniano, acolher, miséria, apoios, reais, desastrosa, queixamse, medicamentos, dívidas, aliados, transportes, migalhas
23	Corruption, Public Management and Clientelism	paliativos, denunciámos, desviar, darnos, clientelismo, exigimos, quietos, confusão, património, clientes, relógio, reina, bandidagem, estradas, generalidade, gasta, preços, fartos, dinheiros, bomba, desmascarada, entregar, sucessivos, sofrem, dinheiro, pequenos, portagens, lesados, uso, pagar, gastos, alimentar, carros, contribuintes, prometi, imensa, fico, impostos, euros, pedofilia, sofre, paciência, associações, banca, cuidados, empresários, precisam, presos, encerramentos, repudiar, ficado, passando, sectores, enganar, doentes, amiguismo, manter, prioridade, faltam, integridade
51	Justice, Cost of Living and Economy, Social Exclusion	deficientes, péssimo, donos, achar, conspiração, imi, aprovar, salário, gostaram, psd, pj, pia, chumbar, aprovação, pedofilia, benefícios, acabam, subsídios, manter, extraordinária, moralista, corruptos, desastrosos, estarem, contribuinte, poderiam, mínimo, perceba, prefere, apertar, combatentes, culpa, etc, ruptura, funciona, rendimento, pedófilo, prescrição, rescaldo, disposto, pântano, leis, discutimos, antiracistas, finais, batemos, narrativas, pagar, indevido, transforma, colocando, italiano, regimento, preocupassem, il, secretários, condenados, gasolina, entendimento, tvs

TABLE A.9. Topics under 3.1.3 - Controversies and Disputes about Political Leadership and Government Management.

ID	Topic Definition	Top 50 Keywords
53	Government Crisis and Public Administration	mandou, governantes, demissionário, beijar, juízo, interna, assessor, esclarecimentos, louco, ministros, administração, institucional, arrogância, funções, mentiu, brutal, prestar, restauração, provocou, explicação, comércio, anterior, credibilidade, parede, primeiroministro, sef, regular, electricidade, beneficiar, milionárias, alterne, silêncio, exsecretário, ridículos, comerciantes, seco, presta, consecutivo, antigo, ligou, reuniões, aposto, boca, conhecidas, carro, negócios, suficiente, discursa, caluniar, mp, acidente, preocupante, ministra, tornarse, email, envolvidos, usurpar, computador, inverno, absoluta
56	Public Health and the Hospital System	brasileiras, pizarro, temido, preocupações, delegação, foge, médicos, galveias, marta, hospital, diálise, hospitalar, particularmente, sindicato, bastonária, preocupantes, utentes, levanta, ignorados, down, relativamente, mentiu, síndrome, doente, cunha, hospitais, queixas, saúde, vacinas, médica, favores, vieram, jorge, instituto, centros, maternidade, profissionais, sns, gémeas, audiência, sector, dr, transporte, rise, comparticipação, crohn, instituição, distrital, representantes, distribuição, cuidados, manuel, partidários, nomeação, centro, algarve, neto, metropolitana, ordens, administração
58	Participation and Political Figures	acaso, nuno, orbán, espanhol, santos, cravinho, medos, pedro, dono, derrota, conversar, leva, luis_montenegro, coelho, ido, protagonista, minutos, comprovar, planeamento, irmão, cnn, exato, meter, implacável, hipócrita, exclusivamente, pinho, entrevista, serpa, tocar, luís, antecipadas, indiano, peço, capazes, hesitações, contou, esmagadora, monte, armadilha, tentou, espetáculos, rasca, dissolução, currículo, anunciada, passos, firmeza, km, paternalismo, seguem, estreia, pesada, antonio_costa, ando, tap, anunciar, pinto, bilhete, lusa

TABLE A.10. Topics under 3.2.1 - Political Dispute, Ideological Positions, Communication and Media (Part I).

ID	Topic Definition	Top 50 Keywords
0	Communication, the Press and Political Mobilization	pdf, bancas, leu, folha, edição, nº, versão, online, inúteis, cópia, participem, fimdesemana, rapazes, leia, folhanacional, assumem, revista, ratos, capa, disponível, publicar, esquecido, avisámos, sobreviver, pensem, deixado, comprar, piada, acabei, depressa, notícias, puro, original, mensagens, sextafeira, esquecemos, sábado, rendimentos, censura, servem, percebemos, leram, fortes, feliz, cerco, saudades, acompanhem, senso, jornal, votem, hipocrisia, funchal, roma, caminhar, ia, arruada, comando, digo, carinho, conselheiro
1	Political Vandalism and Ideological Conflict	democratas, provocação, sedes, vandalizada, danificar, insultos, instrumentalizar, corrigir, vandalismo, outdoors, vandalizar, totalitarismo, facebook, cartazes, minar, mostraram, golpe, fascistas, ditadura, podre, cravos, opositores, antifas, queimaram, obrigados, usam, destruídos, concorda, fascista, afirmar, protestar, impedidos, insultam, intimidam, lápis, únicos, antidemocráticas, plena, receio, inacreditável, insultado, poderes, continuidade, valer, antidemocratas, pensamento, mandasse, incendiários, devo, separação, dominante, arena, intervir, agem, respeitar, televisivo, instalado, atitudes, ameaças, cerco
2	Corruption and Abuse of Power	jair, fantasma, globalismo, endémica, derrotar, sanitária, favorecimento, bolsonaro, promiscuidade, criminal, rebelo, bissau, empobrece, ditadores, bazófia, máxima, únicos, proibido, gaveta, adesão, dorme, legislativo, estrutural, poderes, diplomática, guiné, berardo, terrível, corrupto, novidade, degradante, influências, velha, patriotismo, esquerdistas, ditador, península, ladrão, cordão, utiliza, referência, trump, socialismo, resistência, empregos, acidente, covid, duras, diretamente, degradação, colectiva, compadrio, influência, privado, expresidente, amnistia, junteime, aceitar, joe, empobrecimento
3	Political Coalitions, Controversies and Political Responsibility	cópia, psdps, cds, psdcds, maia, saudação, abstenção, salgueiro, rap, psd, escolheu, pacto, muleta, nomeações, queimaram, essenciais, projectos, vereadores, discordam, cumplicidade, princípio, centrão, adversário, sis, votada, rural, manipulação, deriva, colar, sabias, pdf, fingir, chumbada, cavaco, rangel, admite, refugiado, lm, opção, nenhumas, seca, cpi, dividir, dadas, ruptura, pan, iva, roubo, páginas, apresenta, elogiar, aprovada, pm, chumbaram, cartão, central, vender, exministra, trazem, antidemocrático
4	Media Events and Political Events and Campaigning	voarem, super, live, acompanha, acompanhe, rui_rocha, incendiário, rn, erc, terçafeira, pedras, tino, nulo, vinha, percam, co, utilizam, votachega, famalicão, aquando, ventura, interrompida, encenação, tavares, tsf, qu, tiros, perca, recebida, atacou, deixarmos, imigrante, sic, acompanhem, dinâmica, debates, arrasar, auditório, pen, anjos, requerimento, caravana, direto, espetáculos, moção, perguntas, calados, apresentadas, ratos, agredido, lágrimas, cnn, entrevista, barata, le, estreia, pesada, muleta, coordenadora, centrão
10	Elections and Political Strategy	vencermos, académico, abascal, les, diálise, ordena, santiago, intercampus, intenções, implacável, praticamente, mobilizar, chumbada, subir, confiam, sobe, vierem, farto, podermos, irs, votem, sonhos, emigração, distribuir, vox, distritos, autárquicas, vencer, crescer, março, mote, resultados, fique, elevado, aconteça, legislativas, estarão, atinge, elites, independentemente, miss, governos, jogos, negativo, podendo, média, eleições, extraordinário, resgatar, perceberam, eleitorado, percebam, alternativa, profunda, procura, governação, tornouse, outdoor, vitória, diferente
15	European Right and Political Coordination	fidesz, wilders, geert, rompeu, implementada, aprovada, reeleito, diziam, coordenação, redação, aprovadas, mandatária, baixos, previsto, apoiam, bonitos, retrete, orbán, marisa, húngaro, caldas, demonstrar, prometeu, superiores, voluntários, apresentadas, desígnio, gastou, apoiantes, perceberam, relações, iniciativas, monumental, xvi, apresentou, limpar, assessores, documento, existiam, objectivos, simpatizantes, segundos, futurália, esperança, especialidade, direcção, bipartidarismo, casasdebanho, aprovado, mistas, quartel, programas, presidência, tiago, legislativa, candidato, dra, contigo, prazer, estaríamos
18	Political and Institutional Conflicts	augusto, anular, explique, chão, ilegalização, assumiu, proximidade, acusa, ilegalizar, agressões, santos, silva, lobo, agredidos, david, solene, bancada, ferro, deputada, manifestações, condenar, respeita, regulamento, eleitos, auditoria, distritais, alexandra, ceder, dra, denúncias, abertura, aprovação, cede, expresidente, inquérito, rita_matias, eleito, lula, outdoors, dever, quantas, lusobrasileiras, acordaportugal, to, suspeitas, exige, rodrigues, constitucional, agressão, explicação, profissão, experiências, campos, considerado, gémeas, comemoração, recebidos, freguesia, estarem, prémio

TABLE A.11. Topics under 3.2.1 - Political Dispute, Ideological Positions, Communication and Media (Part II).

ID	Topic Definition	Top 50 Keywords
25	Alternative, Criticism and Political Confidence	consolida, dispostos, exclusivamente, diárias, intenções, agradecemos, ombro, centrão, alternativa, abraçar, afastar, sondagem, central, atacado, diferenças, prontos, confiam, oposição, panorama, seguidores, principal, podes, causas, conseguimos, confiança, acreditam, ppp, capaz, construção, socialismo, rompeu, incapacidade, única, prioridade, insistir, histórico, importância, aliança, intelectual, regional, discordo, apesar, preparar, universidade, mando, participem, assumiu, plataforma, seleção, edifício, cresce, assassinato, democrático, denunciar, winston, destrói, partidárias, voltam, estabilidade, naturalmente
26	European Right and National Identity	identity, preparação, party, pensadores, comissões, conservadora, publicar, gaia, amp, nações, tristes, destino, povos, intensa, grau, representar, líderes, health, homossexuais, pedroso, acabei, identidade, instalada, encontro, assembleia, iii[a], netflix, europeus, diminuição, consulta, formalmente, europeia, international, regionalização, irmão, plenário, pasta, tino, enfrentamos, wilders, presidentes, república, confinamento, gerir, velho, europeu, transformação, jordan, geert, resolvido, ascensão, infame, participação, podíamos, eficaz, orador, preparase, preferia, boasvindas, externos
36	Controversies and Political Scandals	exclusividade, imunidade, avença, câncio, recebia, mariana, fernanda, mortágua, indevido, mensal, imobiliária, robles, advogada, apanhada, judeus, quadros, acusa, especulação, difamar, roubou, espanhola, cadeira, mentiras, despesas, acreditar, ilegalização, milagre, bes, salários, violou, ex, merecia, zolgensma, sérgio, carmo, reféns, deputada, farão, ofensa, panorama, revelações, secretário, joana, mentir, mansão, twitter, conceito, afonso, explicar, jornalistas, be, catarina, segundafeira, falso, condenado, graciosa, ciganos, saiu, recibos, martins
42	European Politics, Territorial Action and Institutional Participation	europe, we, are, envia, manifesto, mithá, in, estradas, tv, beira, funcionários, ambição, percorrer, organizada, passos, madrid, retirar, distribuir, to, ribeiro, modelo, sul, interior, justo, estrada, ue, valer, antidemocratas, artv, crise, desrespeitar, campagnolo, loucos, limpar, promover, insultar, cigana, norte, círculo, plena, and, relevante, live, from, encontram, agricultores, custos, realizou, estabilidade, financeira, celebra, março, propor, politécnico, iii, participação, registo, acompanhe, amanhã, palestra
52	Elections and Political Participation	populacional, elegemos, votem, congressos, diretas, xvi, autárquicas, escolhidos, círculos, regiões, intenções, ultrapassa, legislativas, distritos, populares, estudo, fortes, eleições, academia, dados, recentemente, estando, realizado, independentemente, aximage, queríamos, espetacular, pensado, lm, regionalização, proximidade, semanas, órgãos, ouvimos, quo, dependência, escolhem, sondagens, excepto, eleitorais, concorrer, frutos, conhecidas, juntam, unidade, curiosamente, trans, algarve, implementação, sessões, fila, estruturas, municipais, ministérios, anual, câmaras, apresentadas, duma, secreto, assento
54	Political Participation and Institutional Action	iniciativa, fazêlo, liberal, televisões, lata, condicionar, bombeiro, comunista, hipócrita, nous, trapalhada, enfermeiro, despedida, emergência, adiado, dgs, recebo, tentativas, respeitam, candidatura, identificar, divulgado, floresta, morreram, assento, choque, fraco, pescador, national, escolheram, prevenção, vandalizada, atacar, pensam, estúdio, cumprimentar, votou, informação, ceo, incondicionalmente, direto, joao_cotrim, pude, pátria, desobediência, avante, entendeu, discursos, fiquei, autarquias, quis, presidencial, darão, contar, percurso, insustentável, republicana, figueiredo, tristes, proibição
55	Political Mobilization and Campaign Rhetoric	cedemos, mexer, milímetro, irs, melhorar, natalidade, nosacreditamos, viemos, fizemos, baixar, transformar, imediatamente, elegeram, foco, revelam, mentem, estamosprontos, convicções, pré, beneficiar, europeia, parceria, comícios, aceitam, incansável, politica, confiem, currículo, arranjar, dispara, necessário, national, liderar, cantos, voltaremos, graças, corruptos, formar, passo, aceitaremos, mudança, mudar, alternativa, precisamos, portal, recordamos, figueiredo, simpatizantes, nascimento, consolidar, direcção, mordaça, governativa, vendemos, desistirei, sérgio, continuarem, vêse, meias, provocar

TABLE A.12. Topics under 3.2.2 - Political-Parliamentary and Electoral Cycle, Mobilization, Interaction and Event Organization.

ID	Topic Definition	Top 50 Keywords
9	Parliamentary Activities and Constitutional Review	legislativa, revisaoconstitucional, reunidos, apareçam, estatuto, revisão, taxar, afd, posse, corvo, paços, seminário, represento, jornadas, liderada, tc, conheci, esclarecimento, admito, delegação, césar, tomada, decoro, sessão, permanente, regional, javier, seco, constitucional, vicepresidente, iniciamos, excombatentes, alemão, portugalisrael, propôs, bancada, assessor, participei, federal, transparência, xvi, parlamentares, abandonou, marcar, reunião, terminou, participou, legislatura, pacheco, cerimónia, eleitos, plenário, bélgica, viseu, convenção, trabalhos, objetivos, chiado, conferência, autarcas
13	Political Events and Campaigning	almoçocomício, jantarcomício, ribeira, golegã, recandidatura, mercado, viana, abascal, comício, machico, jantar, arruada, festival, visitamos, santiago, concelho, palmela, lourinhã, feira, olhão, seguido, exportação, portalegre, castelo, distrito, festas, aflito, natal, benavente, matosinhos, convívio, brava, recepção, zona, extraordinário, tradicional, almoço, cascais, concelhias, gaia, viçosa, cheia, eventos, tradições, começamos, beja, inauguração, guimarães, sentese, aveiro, partidochega, onda, chiado, terminamos, organizada, regiões, rentrée, ourique, pediram, multidão
17	National Identity and Patriotism, Youth and Education	inhocsignovinces, secundária, deuspátriafamíliatrabalho, poramoraosportugueses, poramoraportugal, pelobempelaverdadepelavida, umlíderumpartidoumdestino, núcleo, academia, ribatejo, bissau, sandra, ourém, estudantes, estrutura, jornada, universitários, secundário, verão, bonita, corrêa, cm, distrito, santarém, vereadores, pelouro, popular, municipais, cartaxo, distrital, conhecer, cobardes, alunos, juventude, cumprimentar, gomes, lista, viçosa, falámos, reforço, promovido, caparica, conquistar, acolhimento, círculo, fantástica, infarmed, vereador, isento, presença, orgulhoso, estudo, comissões, escola, municipal, derrotar, universidade, plenária, passada, tejo
19	Agriculture, Fisheries and Rural Development	produtores, pescas, rural, agroglobal, pecuária, comissãoparlamentardeagriculturaepescas, pescadores, agriculturafamiliar, agricultura, golegã, pesca, mundorural, ovibeja, agricultores, cap, cavalo, agrícola, primário, florestas, agricultor, caça, correio, cabeças, produtos, corvo, sector, afetados, feira, pescador, produção, prometeu, setores, martinho, publicada, respeitar, compensações, empreendedores, seca, respeitem, exportação, mar, valorizar, indústria, seco, desenvolvimento, painel, orador, concretas, requerimento, vivam, ilha, setor, bolsa, flores, abandono, turismo, emissão, associação, ribatejo, partidário
20	Local and Regional Elections	autárquicos, regionais, autónoma, região, nous, candidatos, afd, autárquica, juntaram, recandidatura, autárquicas, regional, have, delgada, delegação, açores, pré, mandatos, apresentação, amadora, funchal, açorianos, atingiu, ponta, encerramento, sonho, donos, apresentei, chegada, solene, vitórias, distrito, europeias, abandonou, roma, candidaturas, acácia, berlim, tornaram, eleição, candidato, oficial, sampaio, legislativas, arruada, termina[1], ilha, termina, eleitoral, bela, sensação, esperamos, municipal, madeira, eleições, diretamente, posse, pára, aproveitar, setembro
28	Elections, Regional Campaigns and International Far-Right Movements	soberanas, rumar, valença, tângercorrêa, italianas, juntate, milfontes, corrêa, javier, europeias, nações, fiques, presidenciais, elvas, aguiar, livres, primárias, correio, açorianos, desiludir, contigo, mem, varzim, tânger, conto, junho, arruada, minoria, póvoa, joaquim, decisivo, aproximase, vencer, urnas, olhão, vila, matteo, antónio, juntamente, eleições, massiva, regionais, seixal, termina, milei, argentina, aberto, rumo, distrito, fantástico, salvini, emigrantes, lábios, domingo, outdoors, tremer, concelho, évora, terra, largo
49	Public Space, Mobility and Local Identity	formato, lida, isabel, semabrigo, moreira, edição, online, leiria, braga, viana, anjos, augusto, lojistas, paris, aumentou, almirante, moradores, trânsito, chegaram, aveiro, prometido, faro, setúbal, londres, obras, reis, visitamos, castelo, zona, aeroporto, gabinetes, porto, ruas, elitista, aima, dezembro, coimbra, inveja, évora, deputada, regresso, elvas, provocam, santuário, centrodireita, saudar, aprenda, lisboa, senegal, joelhos, selvagens, ofendida, dignas, santos, semana, envergonhar, pescadores, centena, prof, canais

APPENDIX B

Zero Shot Classification Tables

TABLE B.1. Target Groups: Labels, Questions, and Definitions.

Label	**Question**	**Definition**
Racialized	Does this text mention or focus on racialized communities?	Hate speech targeting groups in processes of racialization, attributing a constructed racial identity that often leads to social exclusion.
Roma	Does this text mention or focus on the Roma (gypsy) community?	Hate speech targeting racialized groups identified through the stigmatizing term "gypsy" or similar.
Migrants	Does this text mention or focus on migrants?	Hate speech targeting individuals due to their nationality or migratory origin. Often related to xenophobia.
LGBTI+	Does this text mention or focus on LGBTI+ people?	Hate speech targeting sexual orientation, gender identity/expression, or sex characteristics. Encompasses homophobia, biphobia, transphobia, and intersexphobia.
Intersectional	Does this text mention or focus on intersectional groups (e.g., Black lesbian woman)?	Hate speech targeting individuals simultaneously belonging to multiple marginalized groups. Intersectionality considers how different social identities interact and combine.
Others	Does this text mention or focus on other minority groups outside the main categories (Racialized, Roma, Migrants, LGBTI+, Intersectional)?	Hate speech directed at groups or communities not explicitly included in the other categories of the project.

TABLE B.2. Rhetorical Devices: Labels, Questions, and Definitions.

Label	**Question**	**Definition**
Lexical Creativity	Does this text use the rhetorical device of Lexical Creativity?	Creation or adaptation of insulting or offensive terms, often with coded language or neologisms, to attack individuals or groups.
Metaphor	Does this text use the rhetorical device of Metaphor?	Transfer of attributes between distinct concepts with the aim of devaluing or dehumanizing the target group.
Metonymy	Does this text use the rhetorical device of Metonymy?	Substitution of a group by a characteristic associated with it, often reinforcing stereotypes or negative associations.
Comparison	Does this text use the rhetorical device of Comparison?	Association of a group with another through analogies or comparisons, aiming to depreciate or devalue.
Hyperbole	Does this text use the rhetorical device of Hyperbole?	Use of intentional exaggeration to provoke emotional reactions or reinforce negative perceptions of the target group.
Irony	Does this text use the rhetorical device of Irony?	Indirect and ambiguous expression used to mock or humiliate, often in the form of sarcasm or humor.
Rhetorical Questions	Does this text use Rhetorical Questions?	Questions that do not require an answer, but aim to reinforce a negative opinion or stereotype.
Intertextuality	Does this text use Intertextuality?	Reference to other texts, symbols or discourses to reinforce discriminatory or exclusionary messages.
Fear Appeal	Does this text use the rhetorical device of Fear Appeal?	Strategy that uses fear to persuade, suggesting negative outcomes if no action is taken.
Call to Action	Does this text use a Call to Action?	Explicit or implicit summons to act against the target group or a situation portrayed as negative.

TABLE B.3. Hate Speech Types: Labels, Questions, and Definitions.

Label	**Question**	**Definition**
Direct Hate	Does this text express Direct Hate Speech?	Explicit and aggressive expression of hate against a group based on its social identity.
Indirect Hate	Does this text express Indirect Hate Speech?	Implicit or covert expression of hate, using subtleties or irony to avoid direct accusation.
Counter-speech	Does this text express Counter-speech?	Response aimed at contradicting hate or harmful speech, promoting respect and inclusion.
Offensive Speech	Does this text express Offensive Speech (not necessarily hate)?	Aggressive or insulting language with an individual focus, not necessarily based on social group.

TABLE B.4. Emotions: Labels, Questions, and Definitions.

Label	**Question**	**Definition**
Hate	Does this text express the emotion of Hate?	Intense aversion and desire to eliminate or harm a group perceived as evil or immutable.
Anger	Does this text express Anger?	Negative emotion motivated by perception of injustice, often with a desire for confrontation or change.
Disgust	Does this text express Disgust?	Moral or physical rejection of a group, often associated with ideas of contamination or devaluation.
Fear	Does this text express Fear?	Emotion associated with the perception of threat or danger (physical or symbolic) posed by the target group.
Guilt	Does this text express Guilt?	Recognition of moral responsibility for transgressions committed by the in-group.
Shame	Does this text express Shame?	Desire to hide or distance oneself due to shameful conduct attributed to one's own group.
Hope	Does this text express Hope?	Positive emotion associated with reconciliation, change, or improvement in intergroup relations.

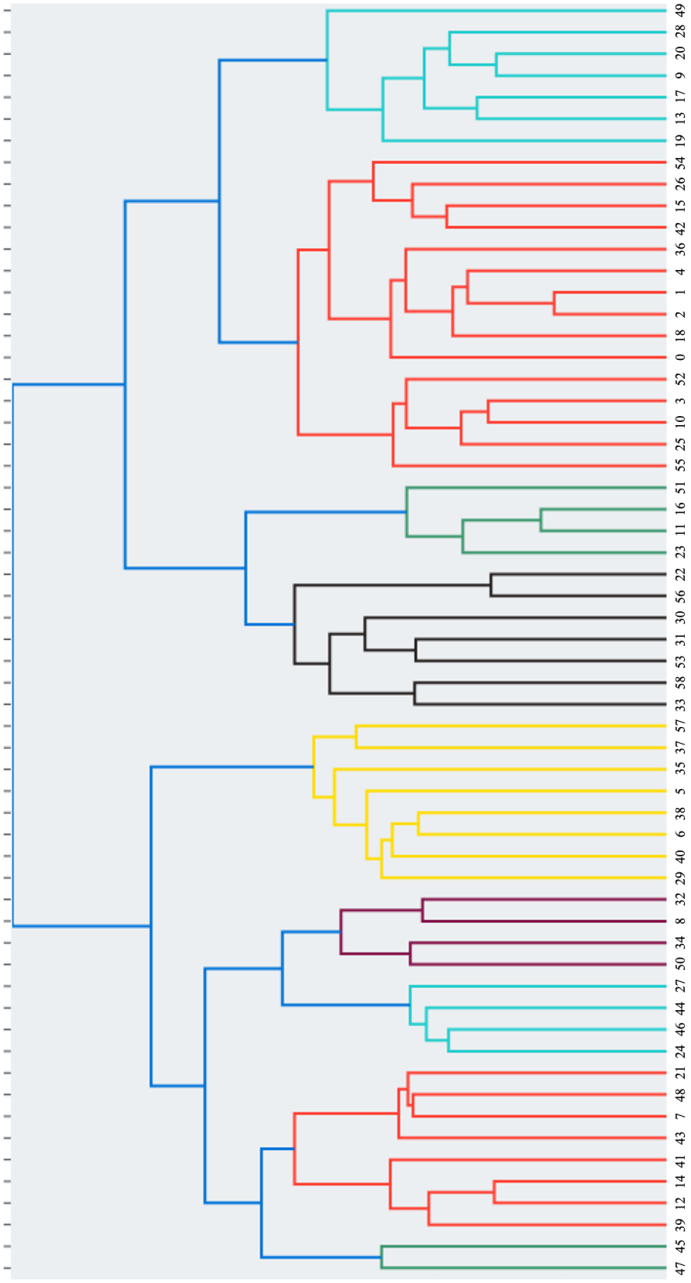
TABLE B.5. Discursive Strategies: Labels, Questions, and Definitions.

Label	**Question**	**Definition**
Negation of Hate	Does this text use the strategy of Negation of Hate?	Strategy of presenting one's own discourse as legitimate and non-hateful, even when transgressing social norms.
Role Reversal	Does this text use Role Reversal?	Presenting the dominant group as the victim and the marginalized group as the aggressor.
Stereotyping	Does this text use Stereotyping?	Attribution of fixed and negative characteristics to a group based on incorrect generalizations.
Dehumanization	Does this text use Dehumanization?	Representation of the target group as less human, facilitating hostility or violence against it.
Silencing	Does this text use Silencing?	Disqualification or humiliation of the target group to prevent its voice or participation in public discourse.
Realistic Threat	Does this text use the strategy of Realistic Threat?	Perception of concrete threat to the resources, security, or well-being of the dominant group.
Symbolic Threat	Does this text use the strategy of Symbolic Threat?	Perception of threat to the values, beliefs, or identity of the dominant group.

APPENDIX C

Topics Evolution Figures

FIGURE C.1. Hierarchical clustering - topic representation of the hierarchy Level.



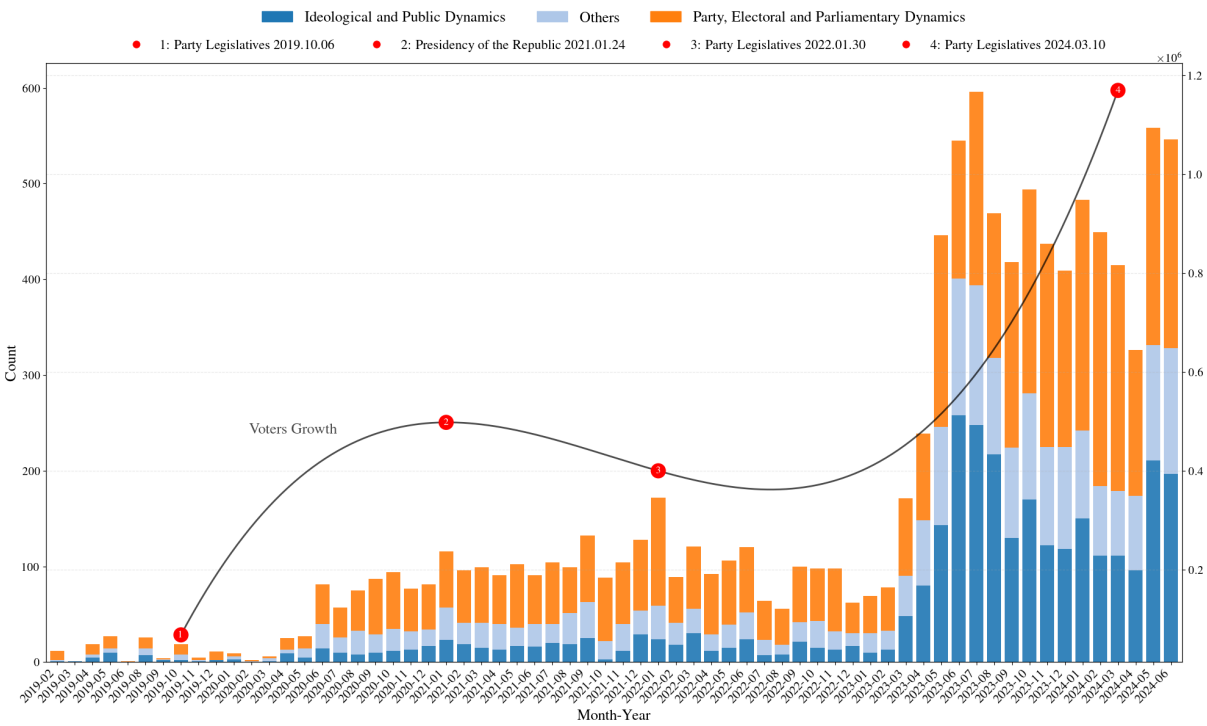


FIGURE C.2. *'Political discourse'* on *X (formerly Twitter)* associated with CHEGA (2019-2024).

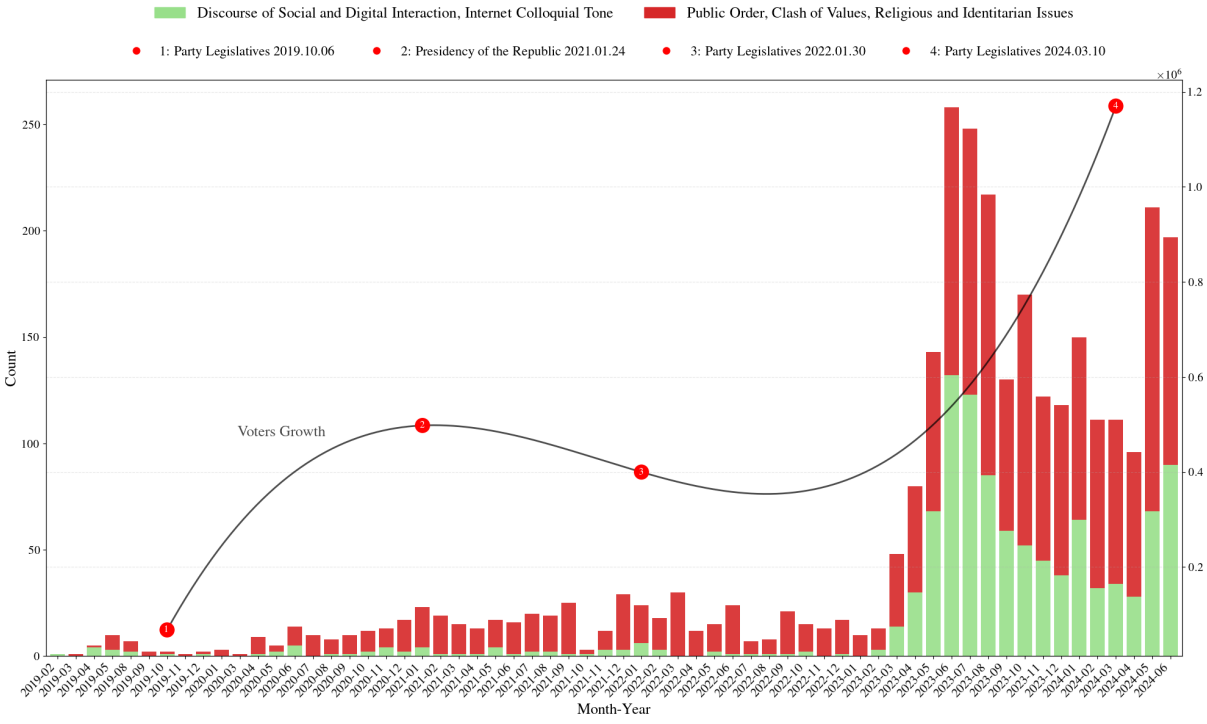


FIGURE C.3. Monthly evolution of 2: *'Ideological and Public Dynamics'* Dimension.

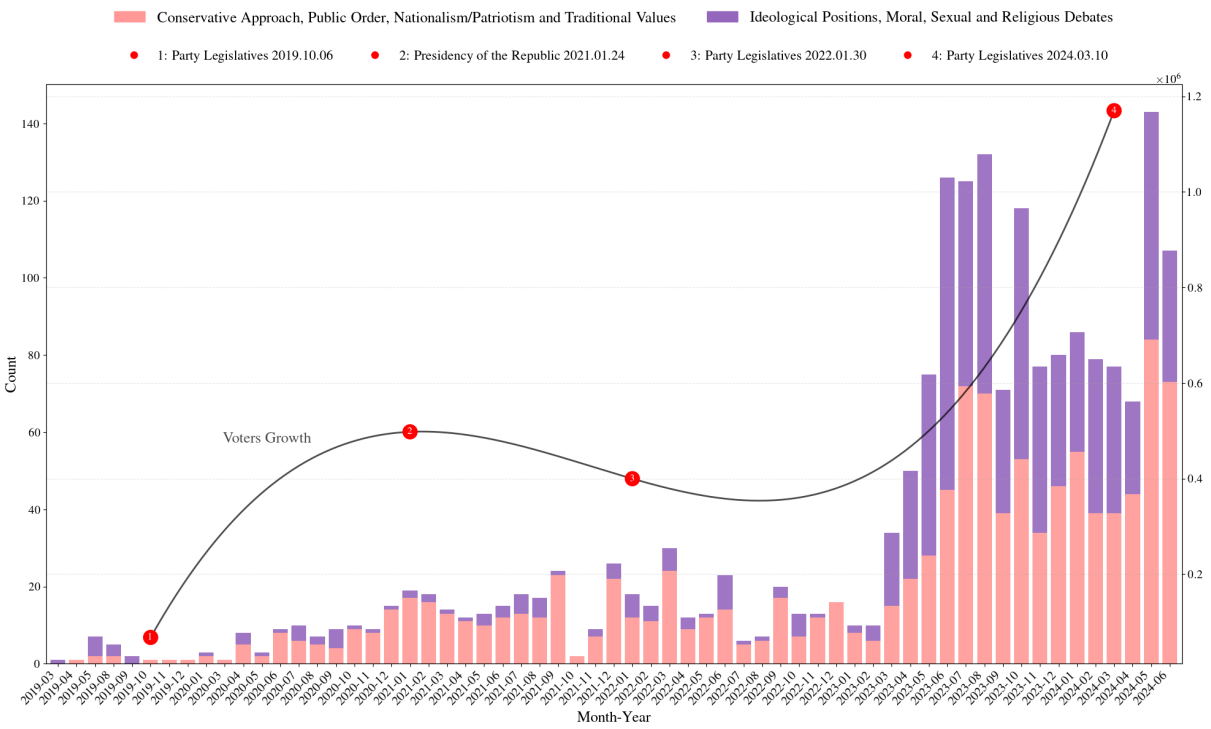


FIGURE C.4. Monthly evolution of 2.1: *'Public Order, Clash of Values, Religious and Identitarian Issues'.*

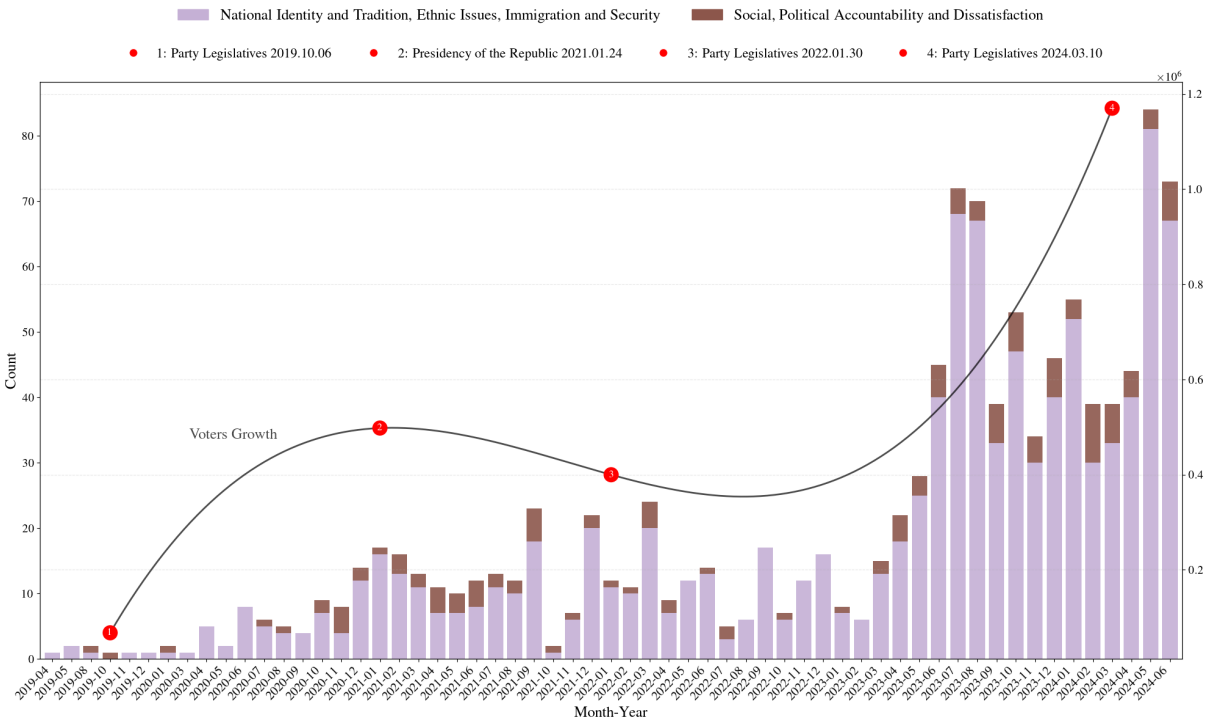


FIGURE C.5. Monthly evolution of 2.1.1: *'Conservative Approach, Public Order, Nationalism/Patriotism and Traditional Values'.*

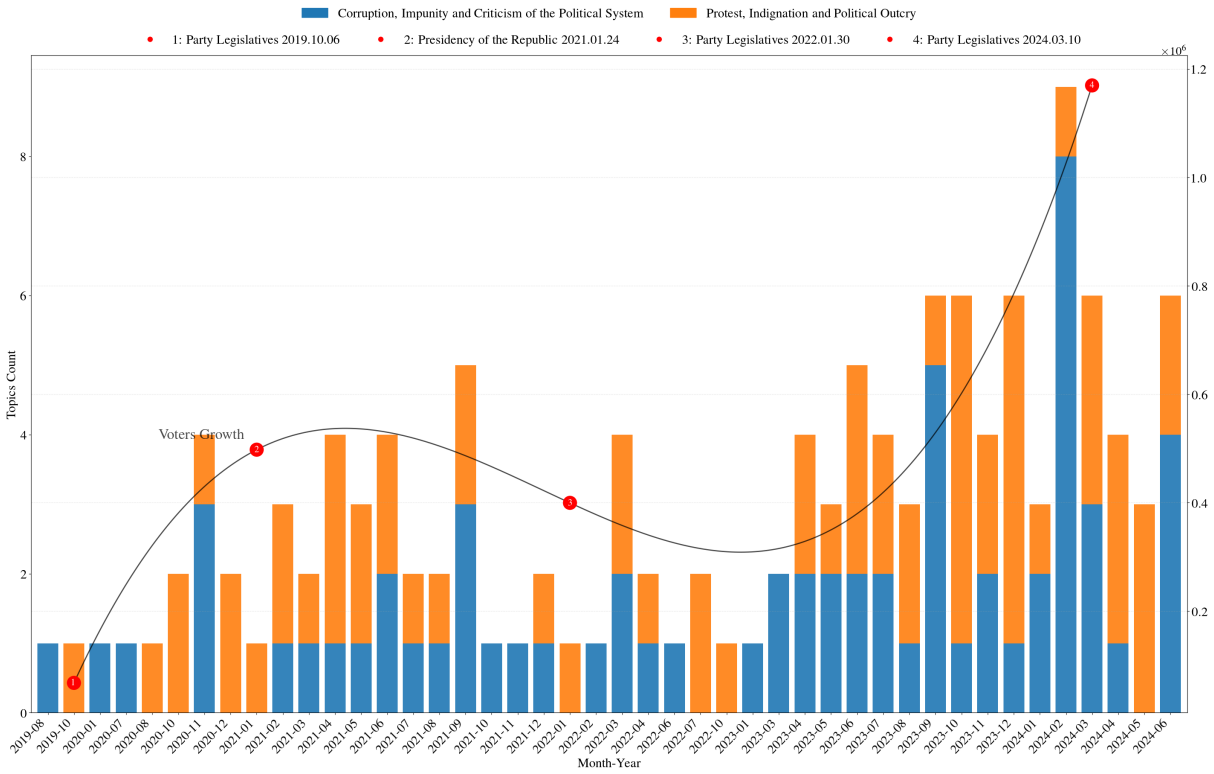


FIGURE C.6. Monthly evolution of 2.1.1.1: *'Social, Political Accountability and Dissatisfaction'* Topics.

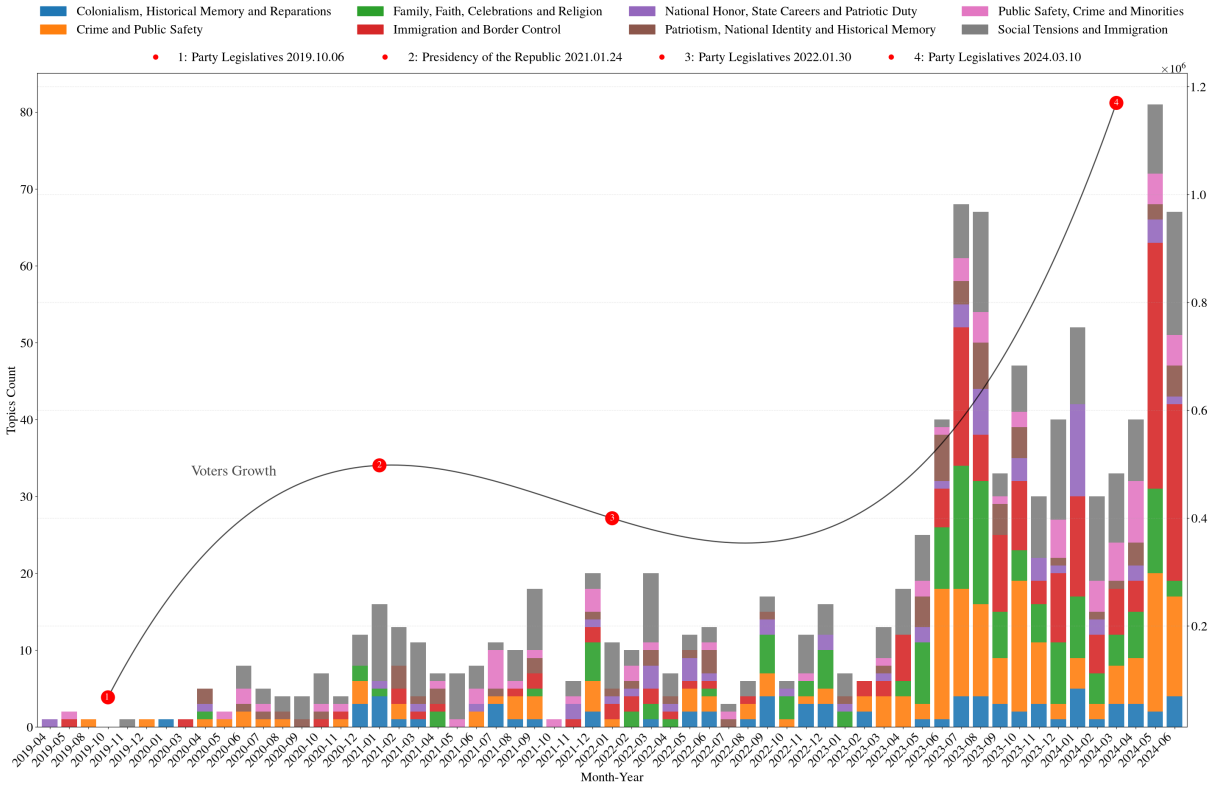


FIGURE C.7. Monthly evolution of Topics Related to 2.1.1.2: *'National Identity, Tradition, Ethnic Issues, Immigration, and Security Topics'*.

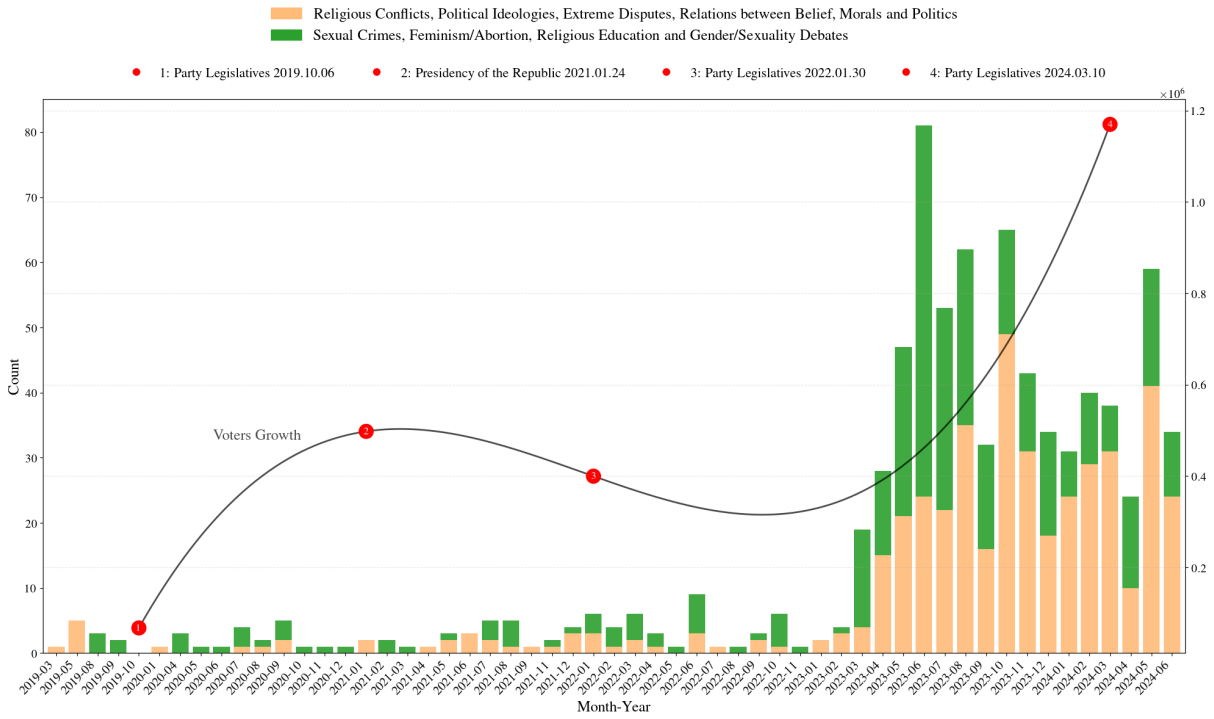


FIGURE C.8. Monthly evolution of 2.1.2: *'Ideological Positions, Moral, Sexual and Religious Debates'.*

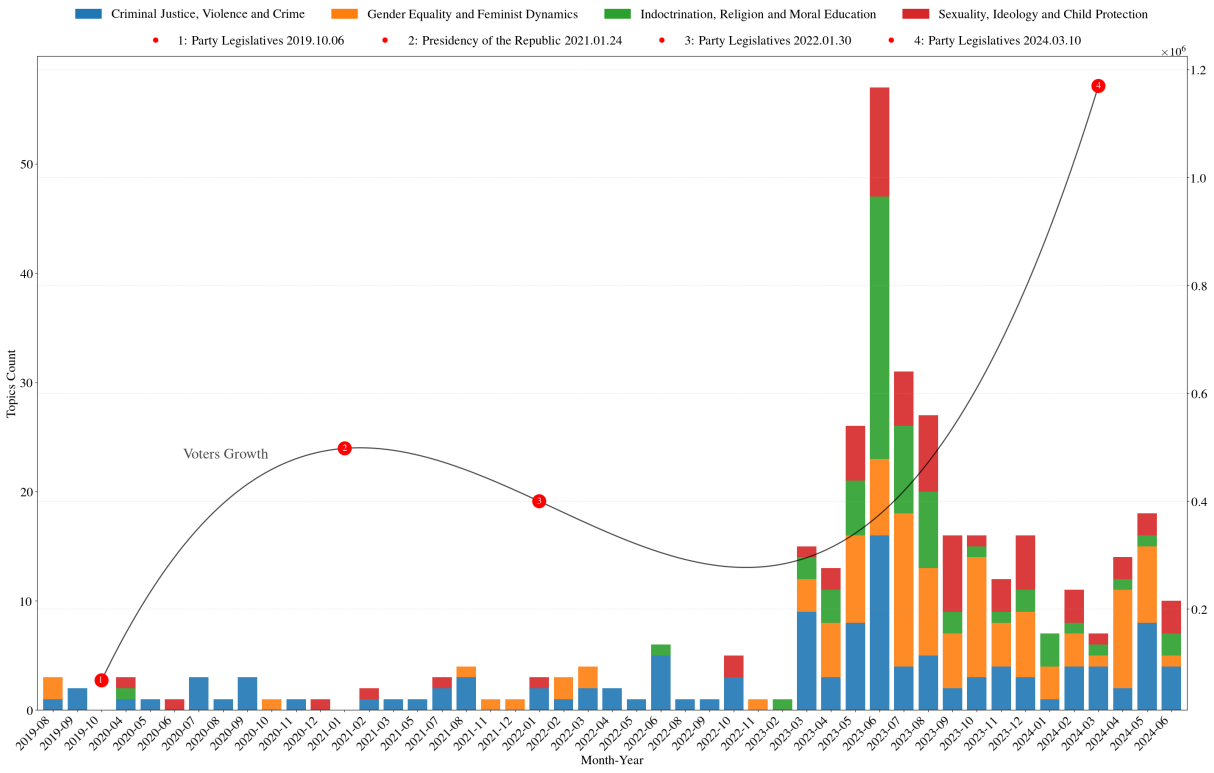


FIGURE C.9. Monthly evolution of 2.1.2.1: *'Sexual Crimes, Feminism/Abortion, Religious Education and Gender/Sexuality Debates'* Topics.

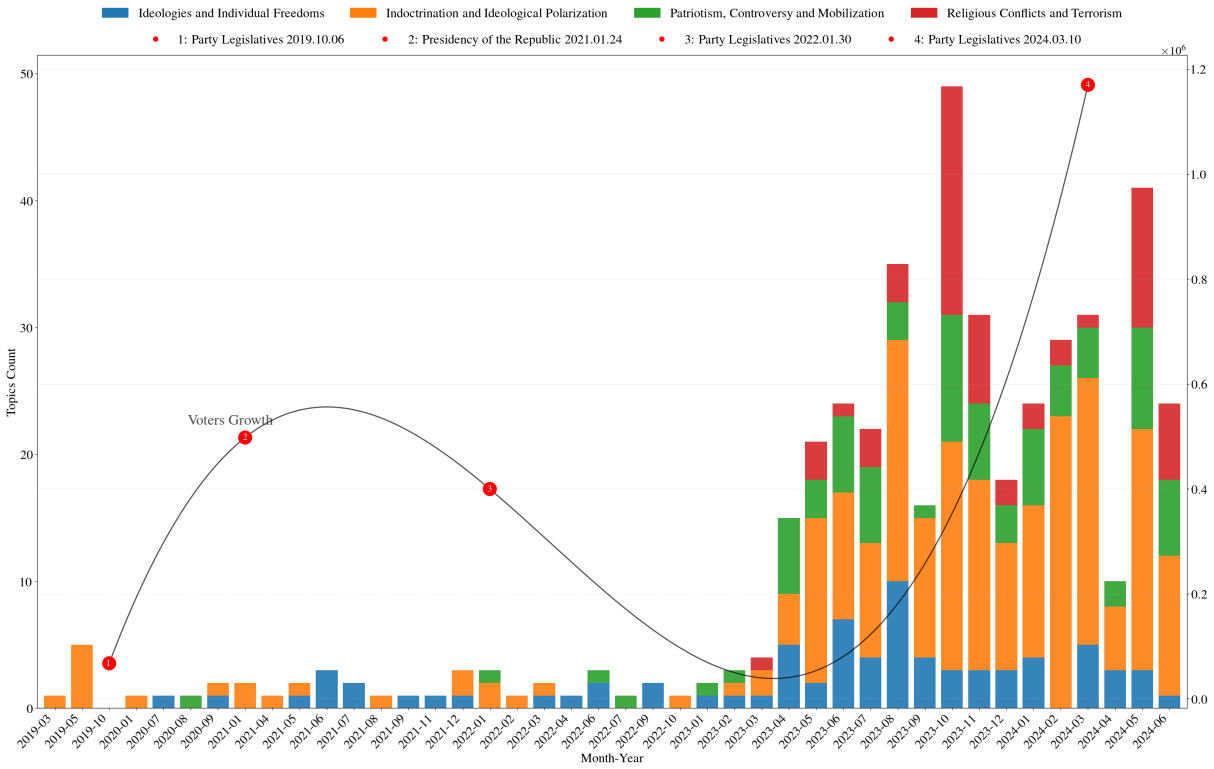


FIGURE C.10. Monthly evolution of 2.1.2.2: *'Religious Conflits, Political Ideologies, Extreme Disputes, Relations between Belief, Morals and Politics'* Topics.

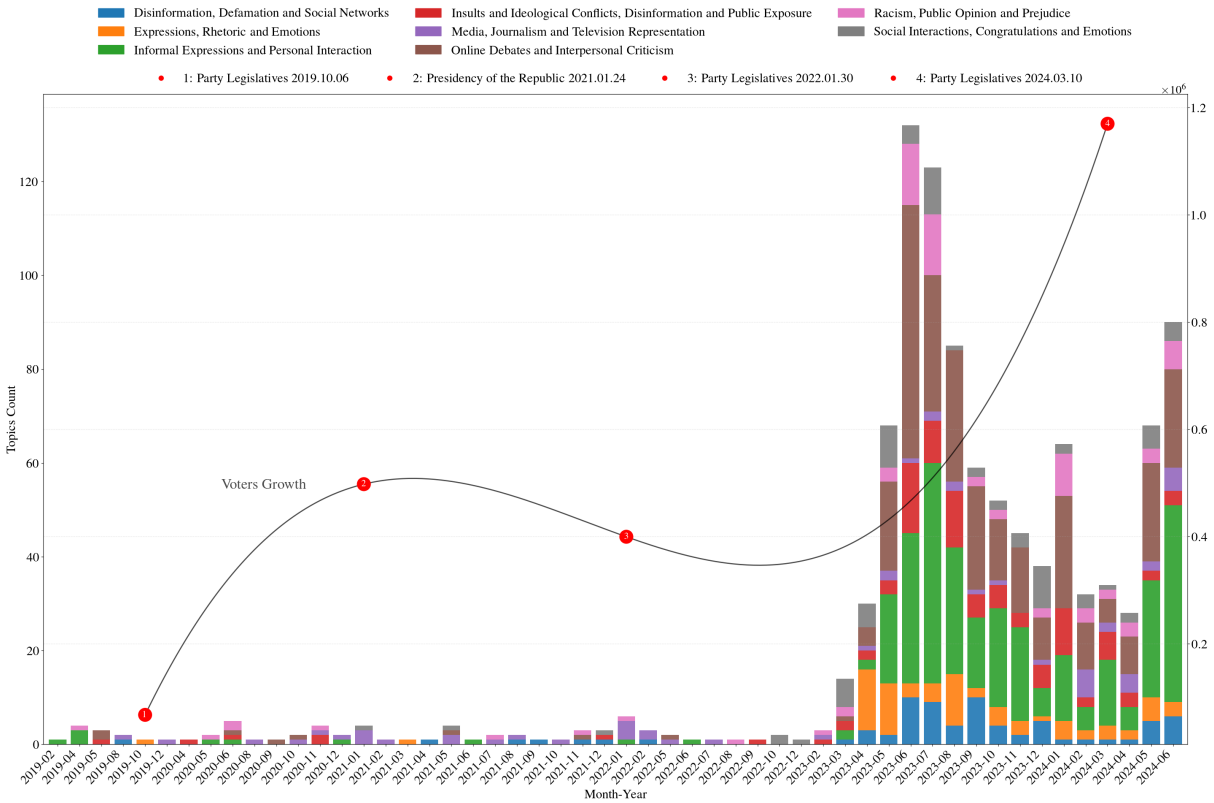


FIGURE C.11. Monthly evolution of 2.2: Discourse of *'Social and Digital Interaction, Internet Colloquial Tone'* Topics.

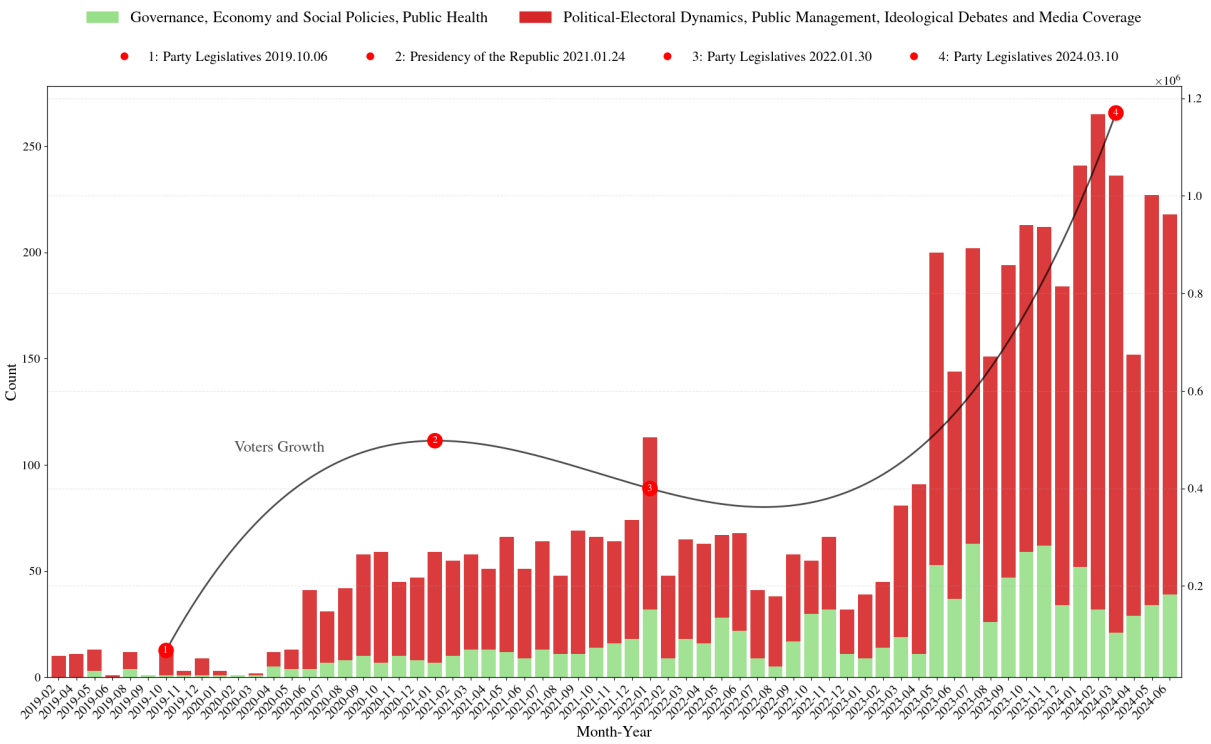


FIGURE C.12. Monthly evolution of 3: *'Party, Electoral and Parliamentary Dynamics'* Dimension.

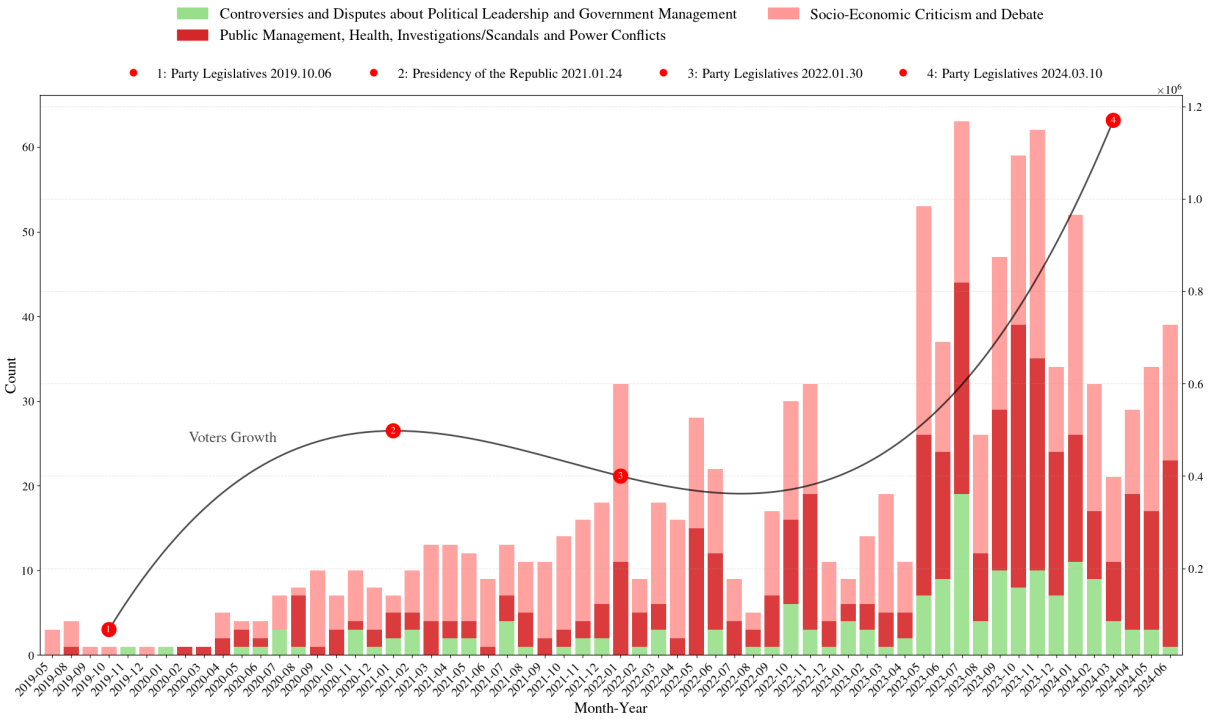


FIGURE C.13. Monthly evolution of 3.1: *'Governance, Economy and Social Policies, Public Health'.*

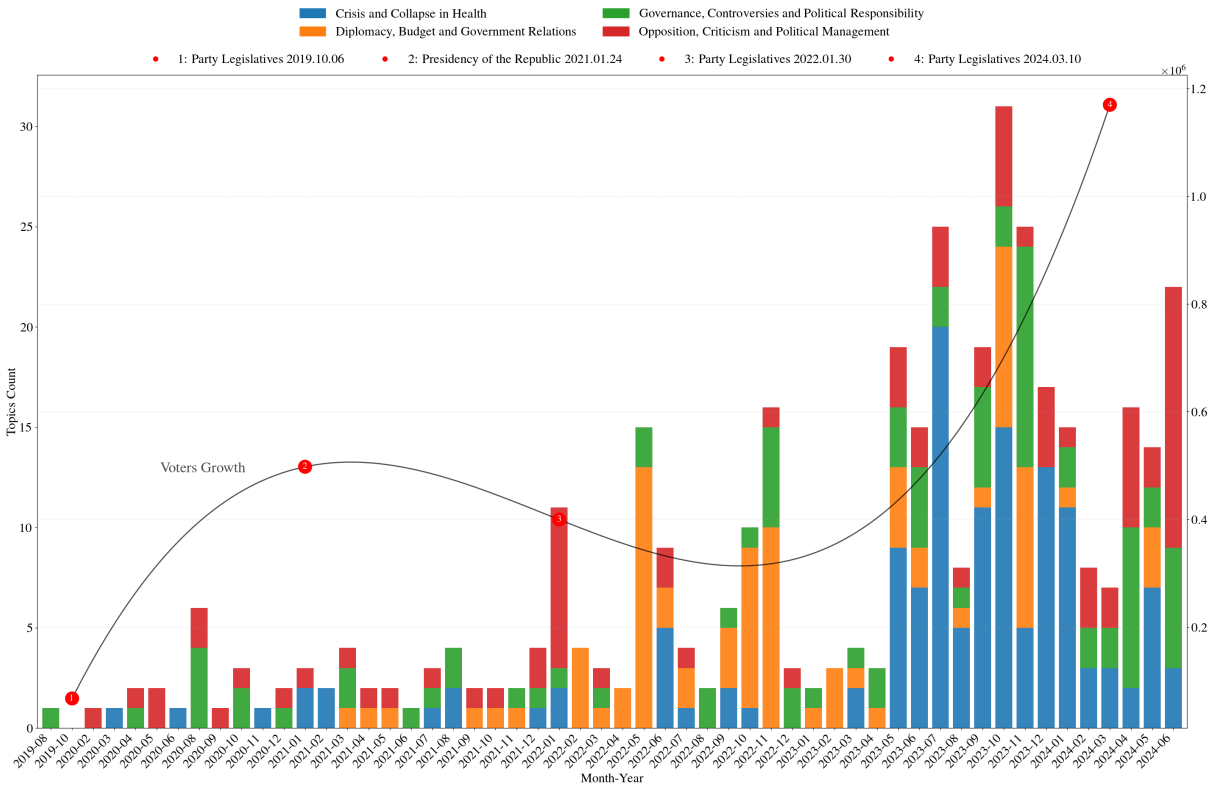


FIGURE C.14. Monthly evolution of 3.1.1: *'Public Management, Health, Investigations/Scandals and Power Conflicts'* Topics.

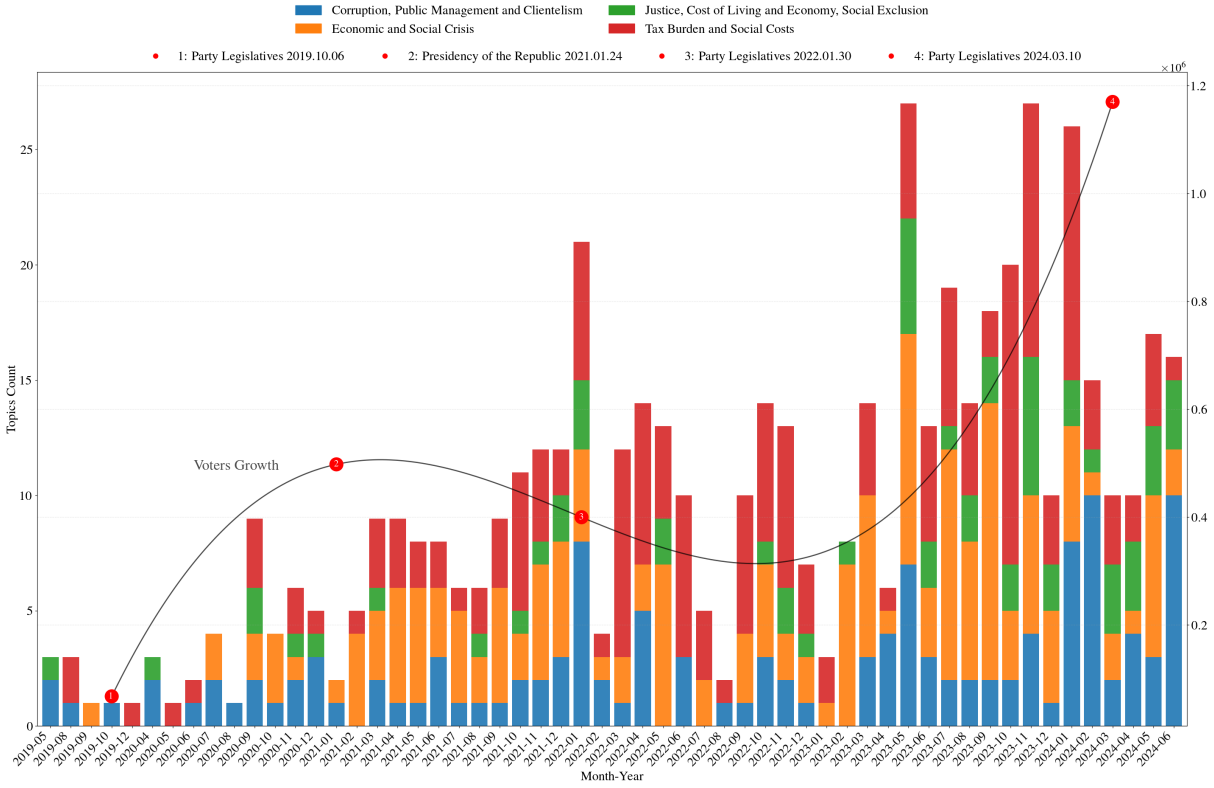


FIGURE C.15. Monthly evolution of 3.1.2: *'Socio-Economic Criticism and Debate'* Topics.

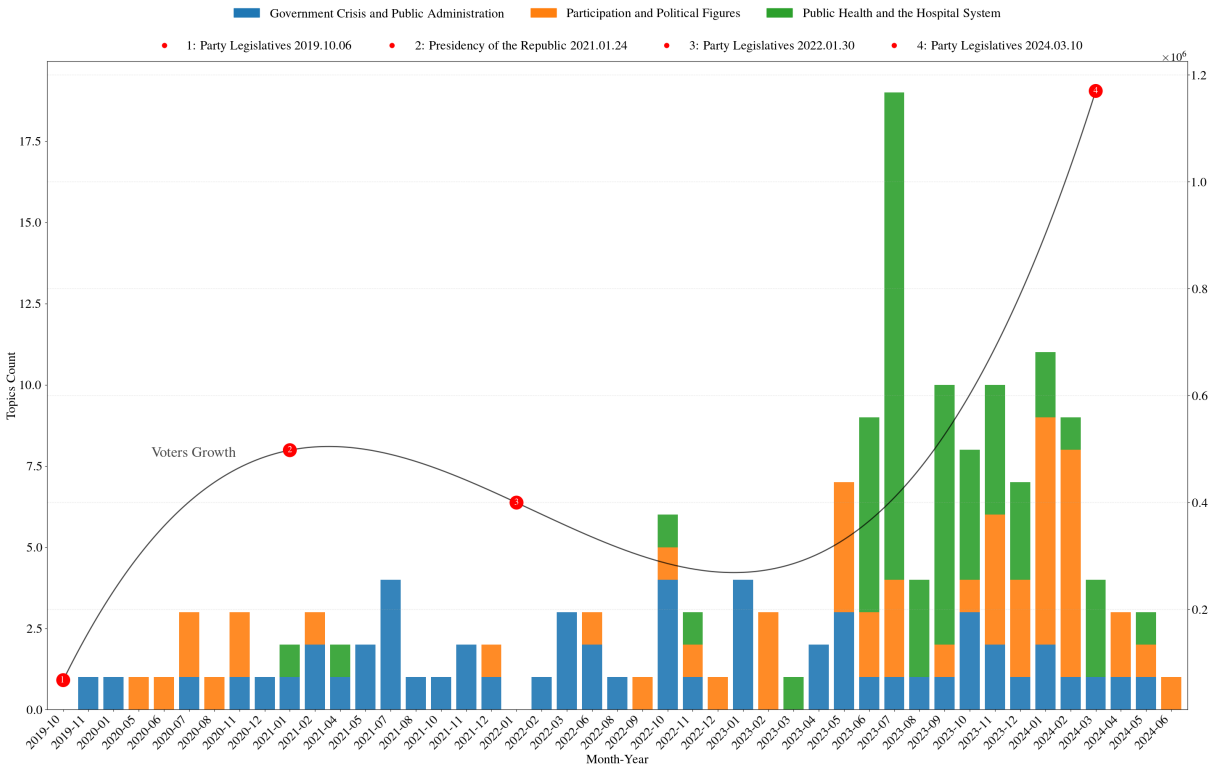


FIGURE C.16. Monthly evolution of 3.1.3: *'Controversies and Disputes about Political Leadership and Government Management'* Topics.

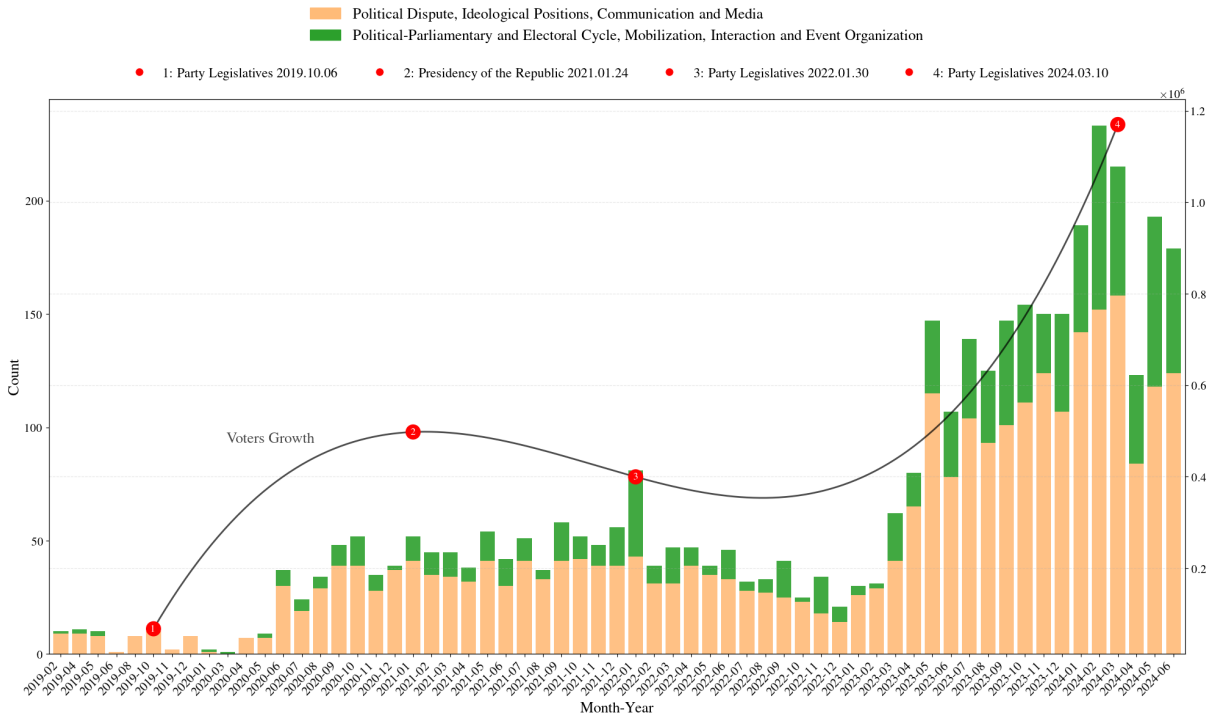


FIGURE C.17. Monthly evolution of 3.2: *'Political-Electoral Dynamics, Public Management, Ideological Debates and Media Coverage'* Dimension.

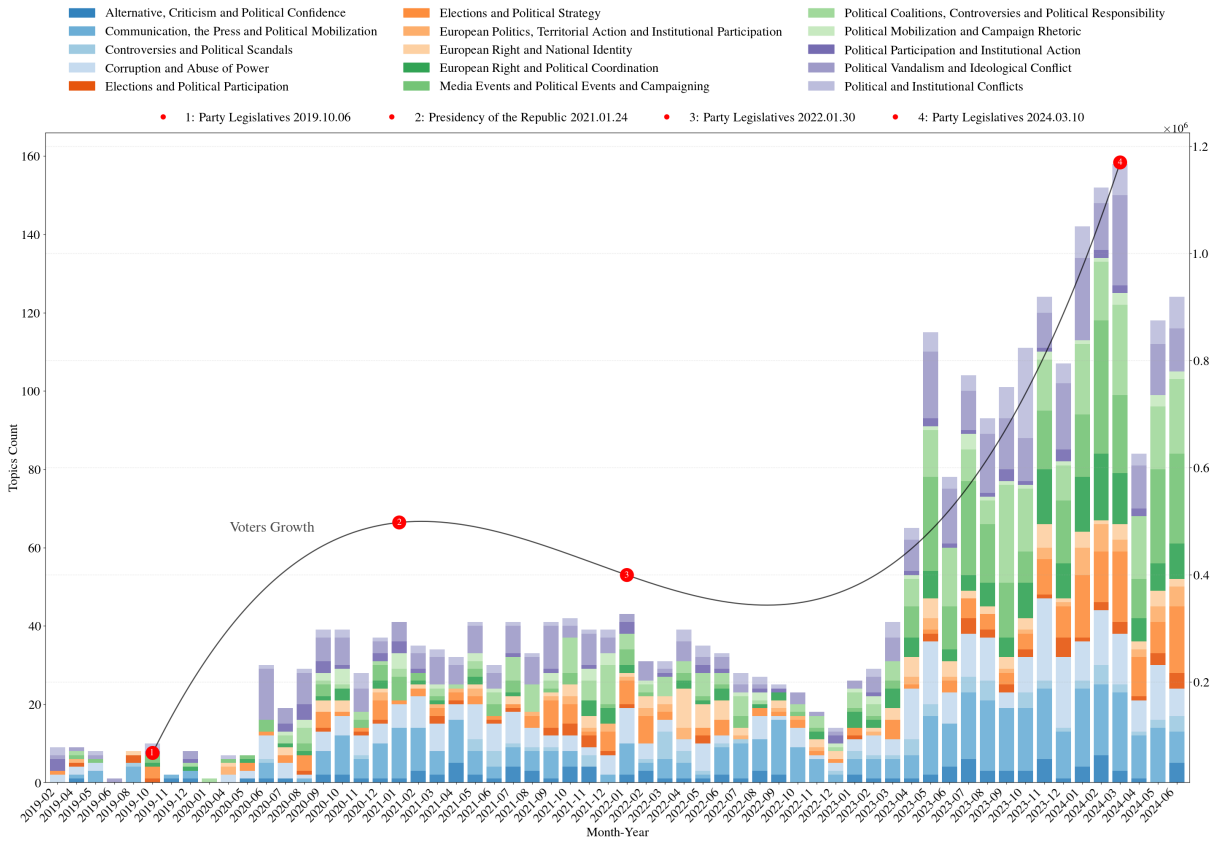


FIGURE C.18. Monthly evolution of 3.2.1: *'Political Dispute, Ideological Positions, Communication and Media'* Topics.

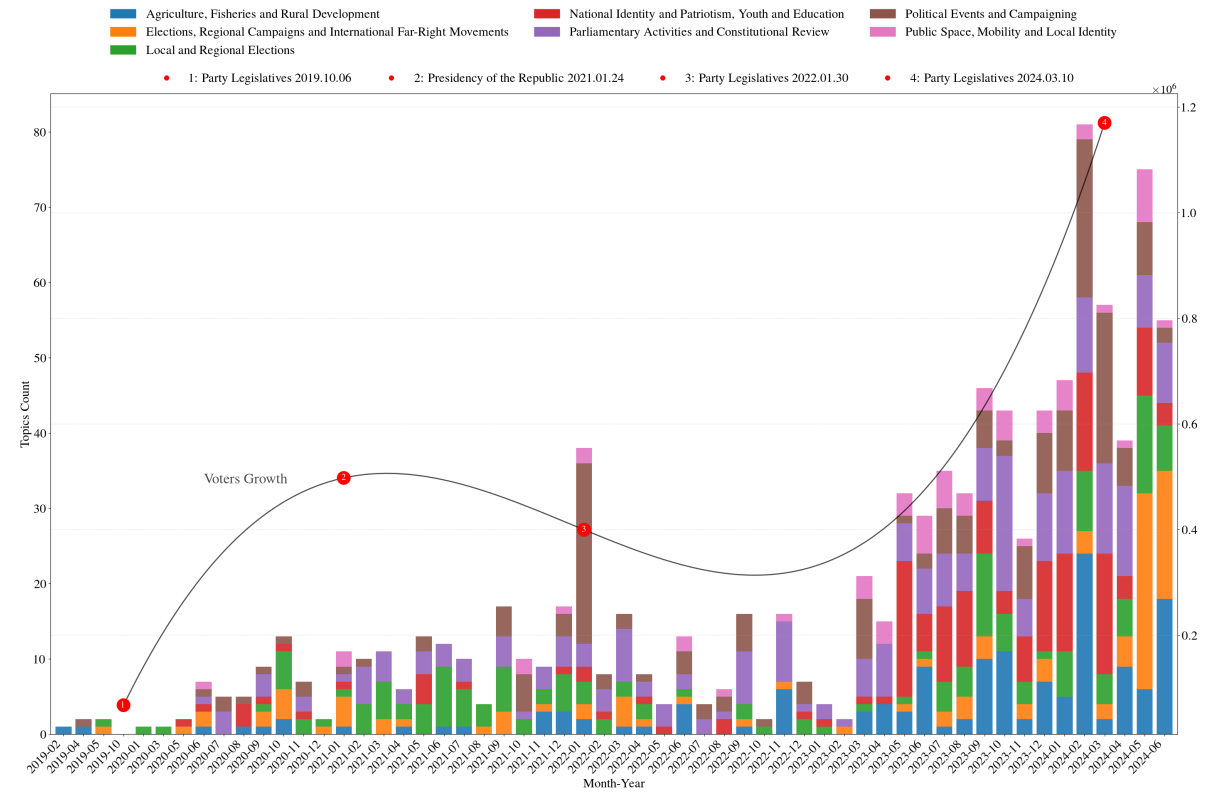


FIGURE C.19. Monthly evolution of 3.2.2: *'Political-Parliamentary and Electoral Cycle, Mobilization, Interaction and Event Organization'* Topics.

APPENDIX D

Zero Shot Classification Figures

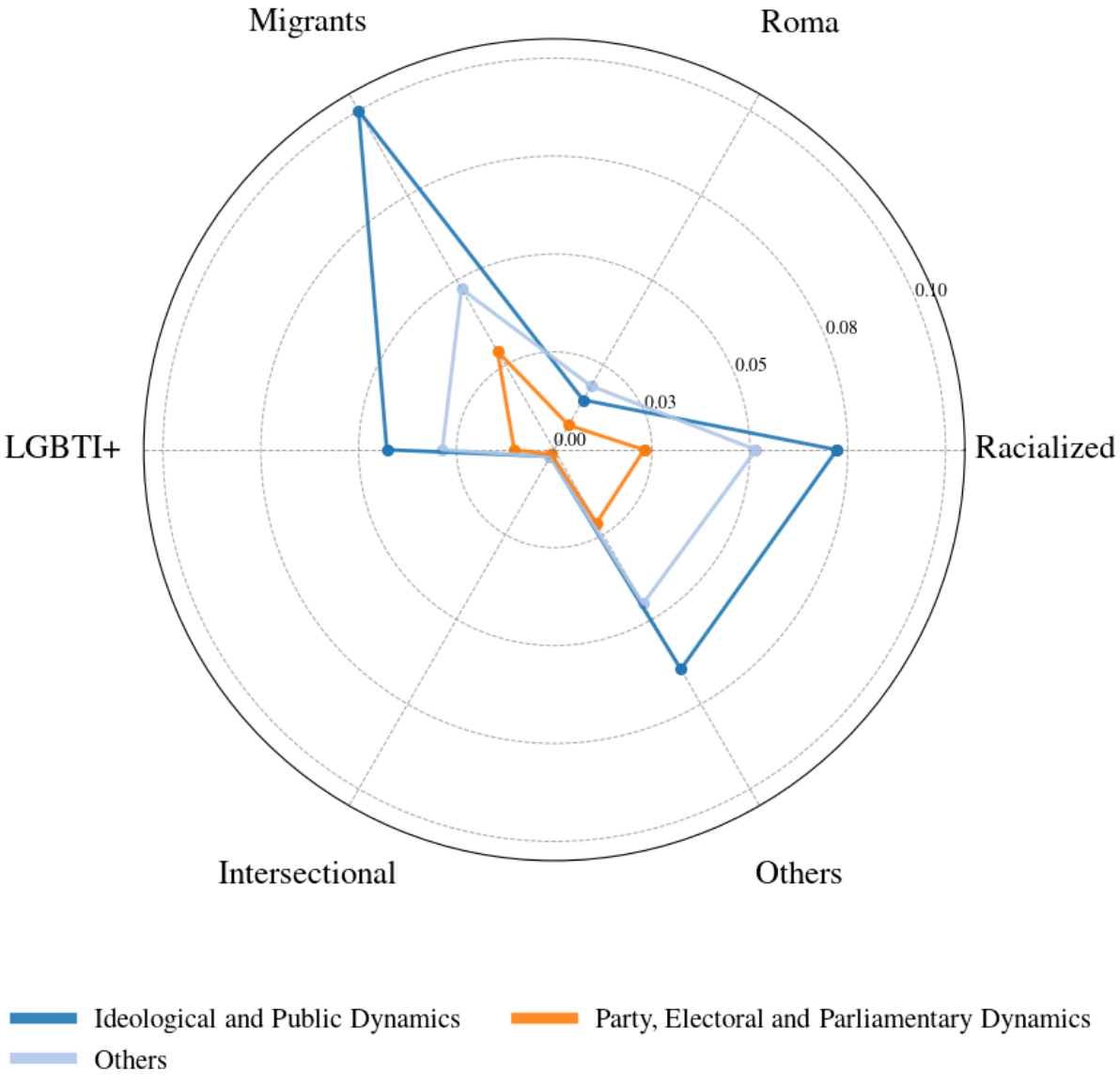


FIGURE D.1. HS targets probability distribution across *'Political Discourse'*.

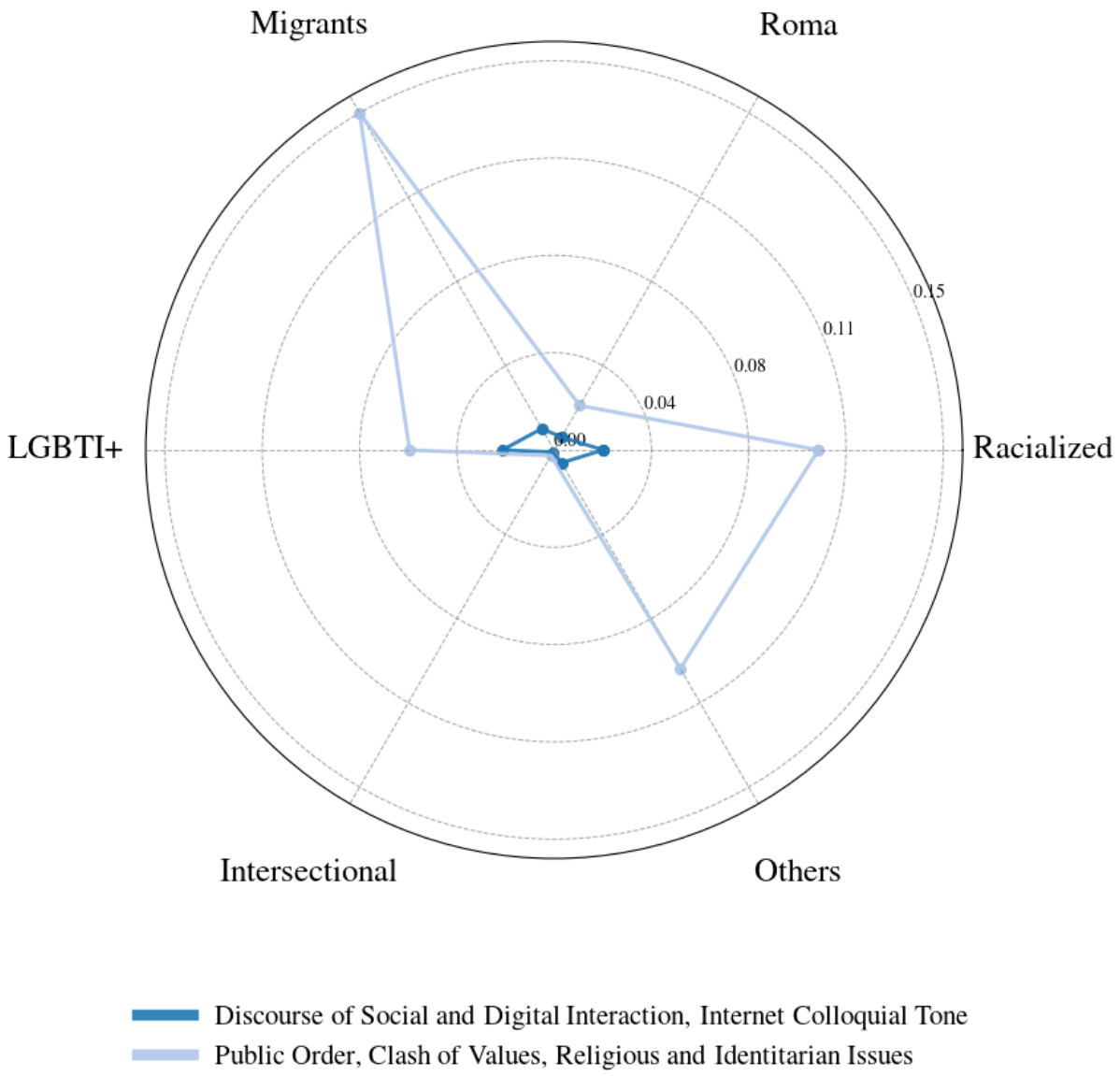


FIGURE D.2. HS targets probability distribution across *'Ideological and Public Dynamics'* Dimension.

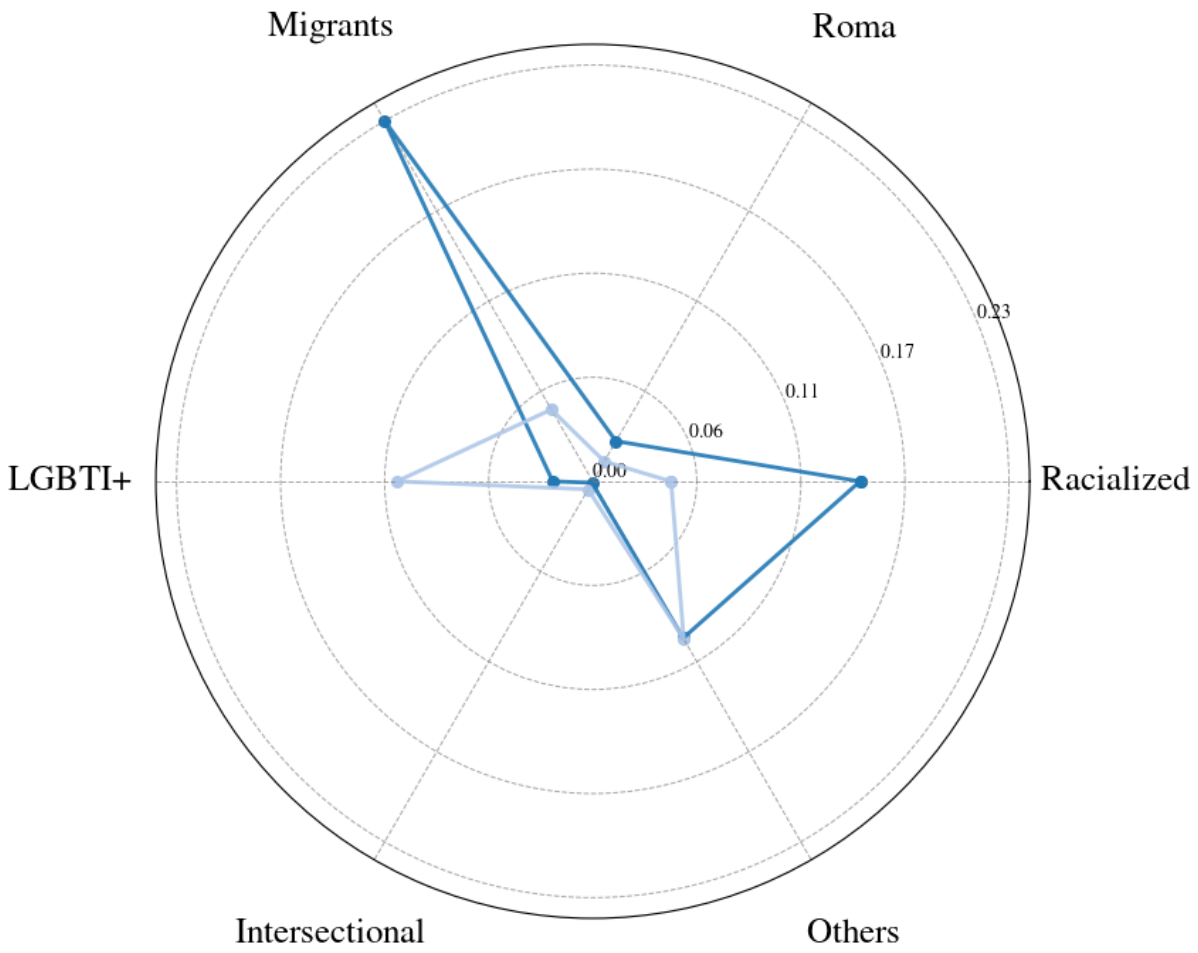


FIGURE D.3. HS targets probability distribution across *'Public Order, Clash of Values, Religious and Identitarian Issues'*.

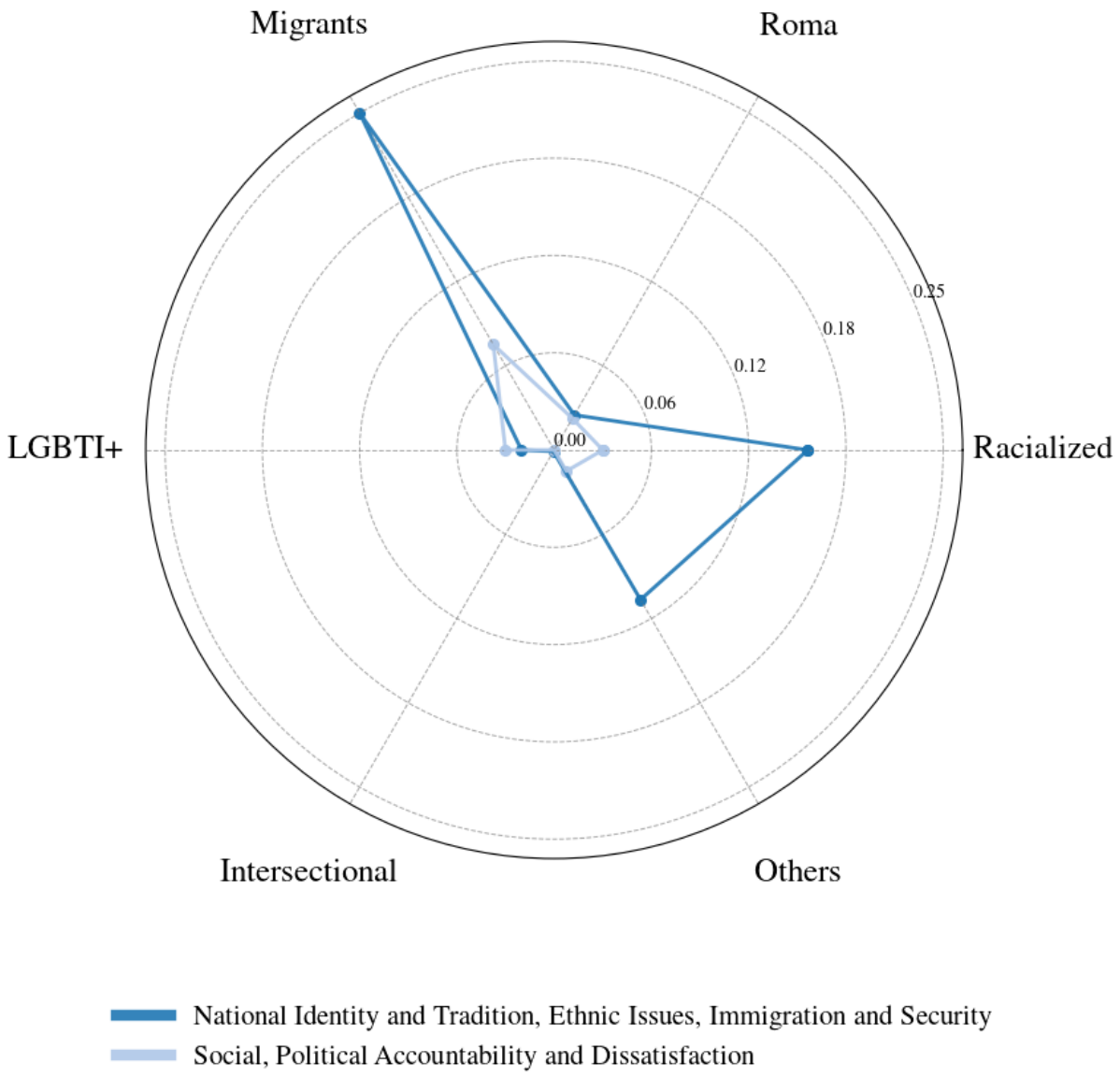


FIGURE D.4. HS targets probability distribution across *'Conservative Approach, Public Order, Nationalism/Patriotism and Traditional Values'*.

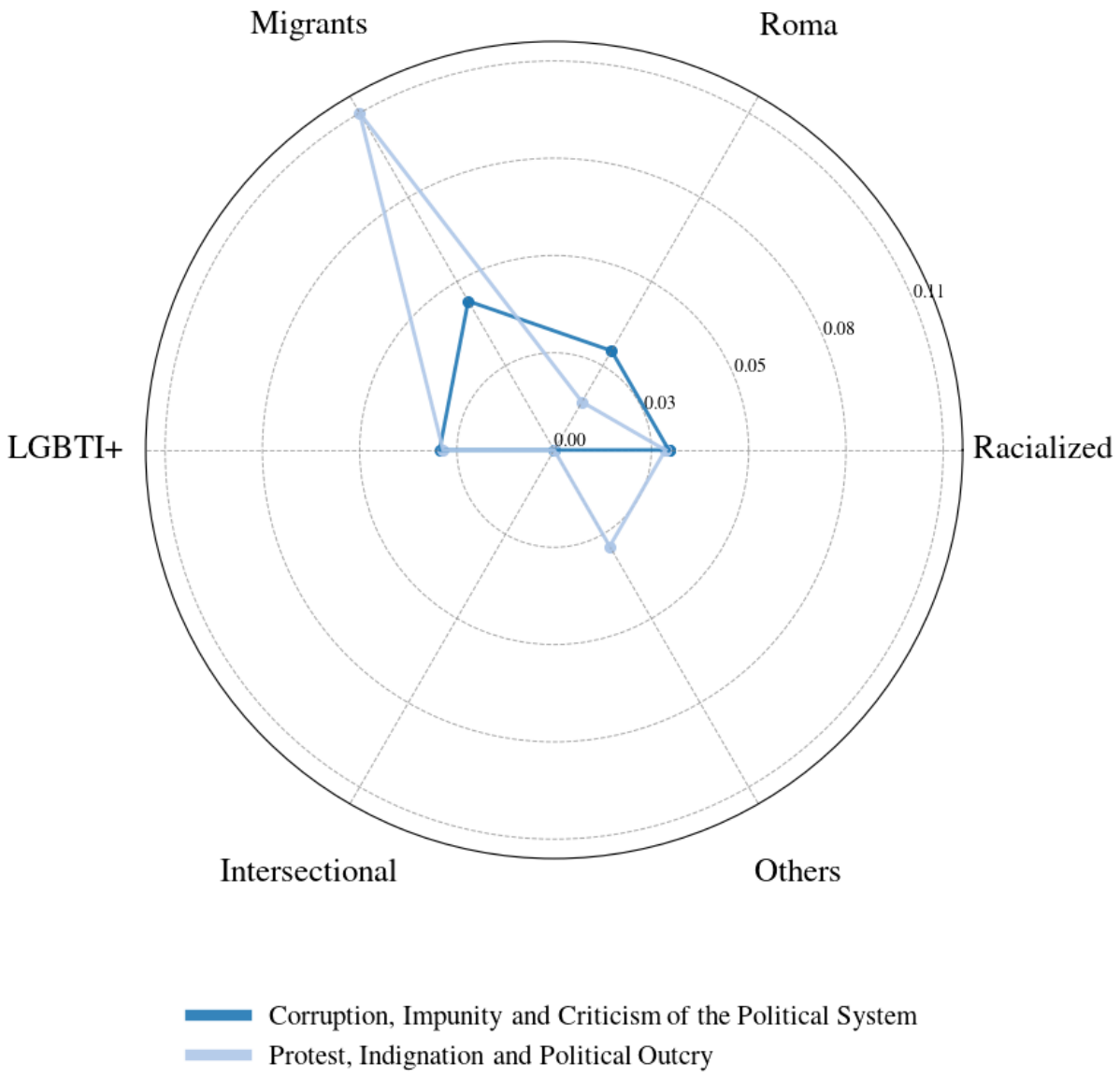


FIGURE D.5. HS targets probability distribution across *'Social, Political Accountability and Dissatisfaction'*.

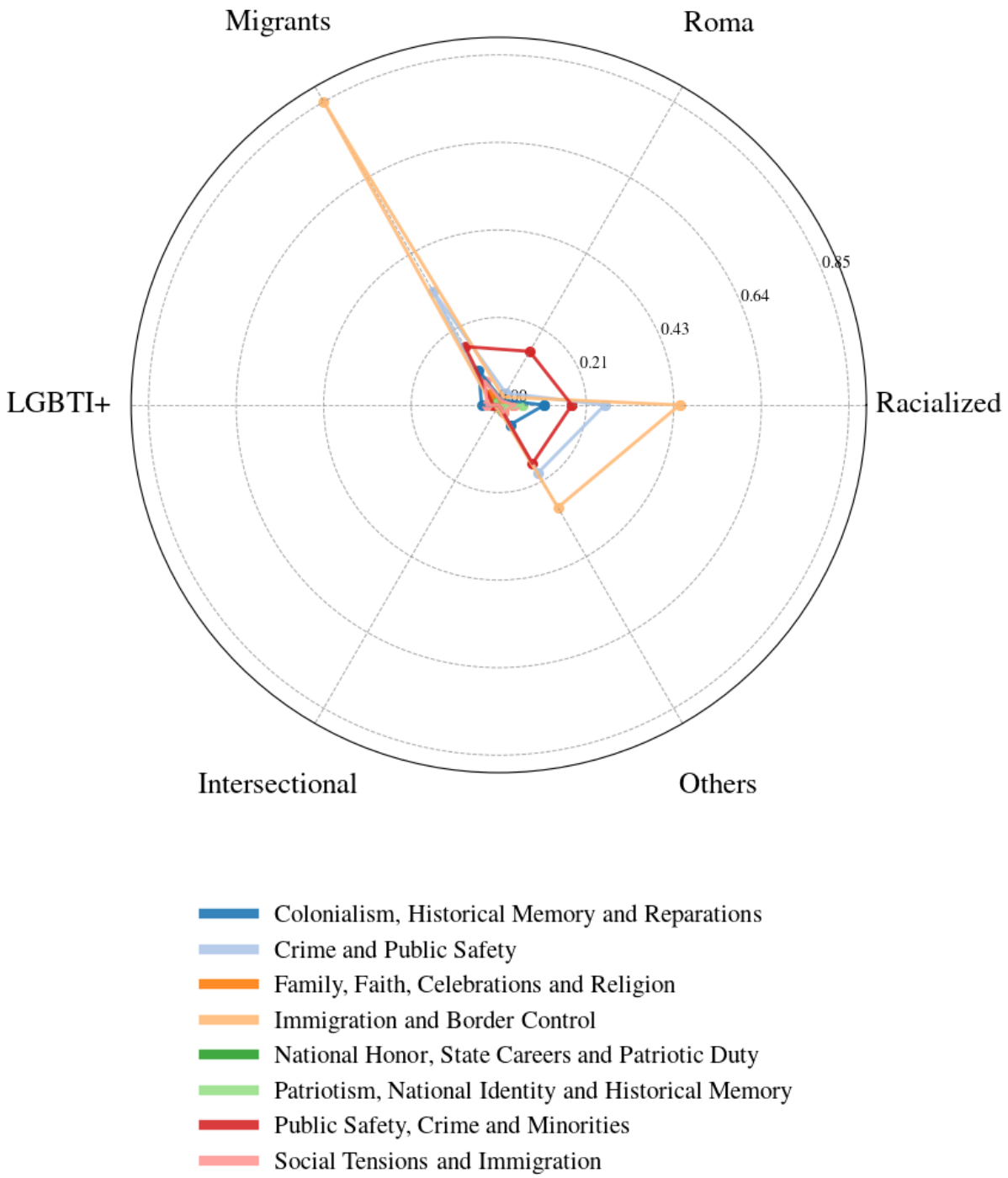


FIGURE D.6. HS targets probability distribution across *'National Identity and Tradition, Ethnic Issues, Immigration and Security'*.

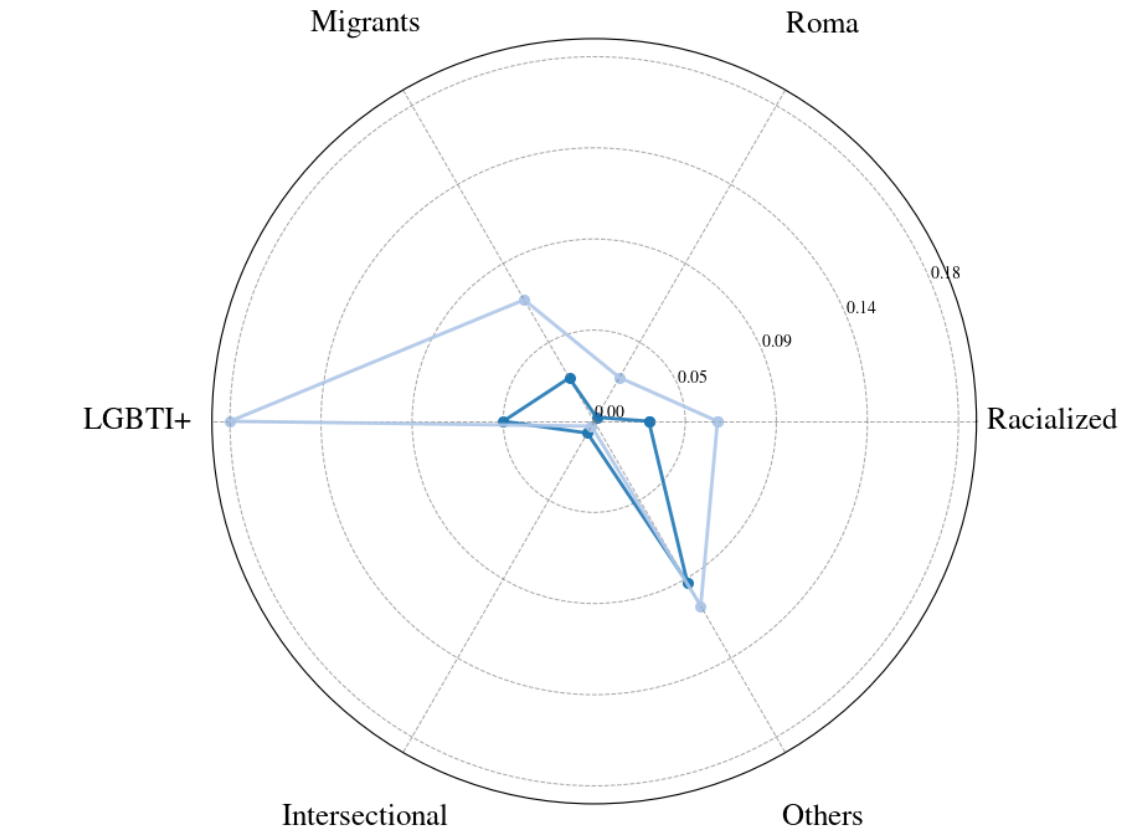


FIGURE D.7. HS targets probability distribution across *'Ideological Positions, Moral, Sexual and Religious Debates'*.

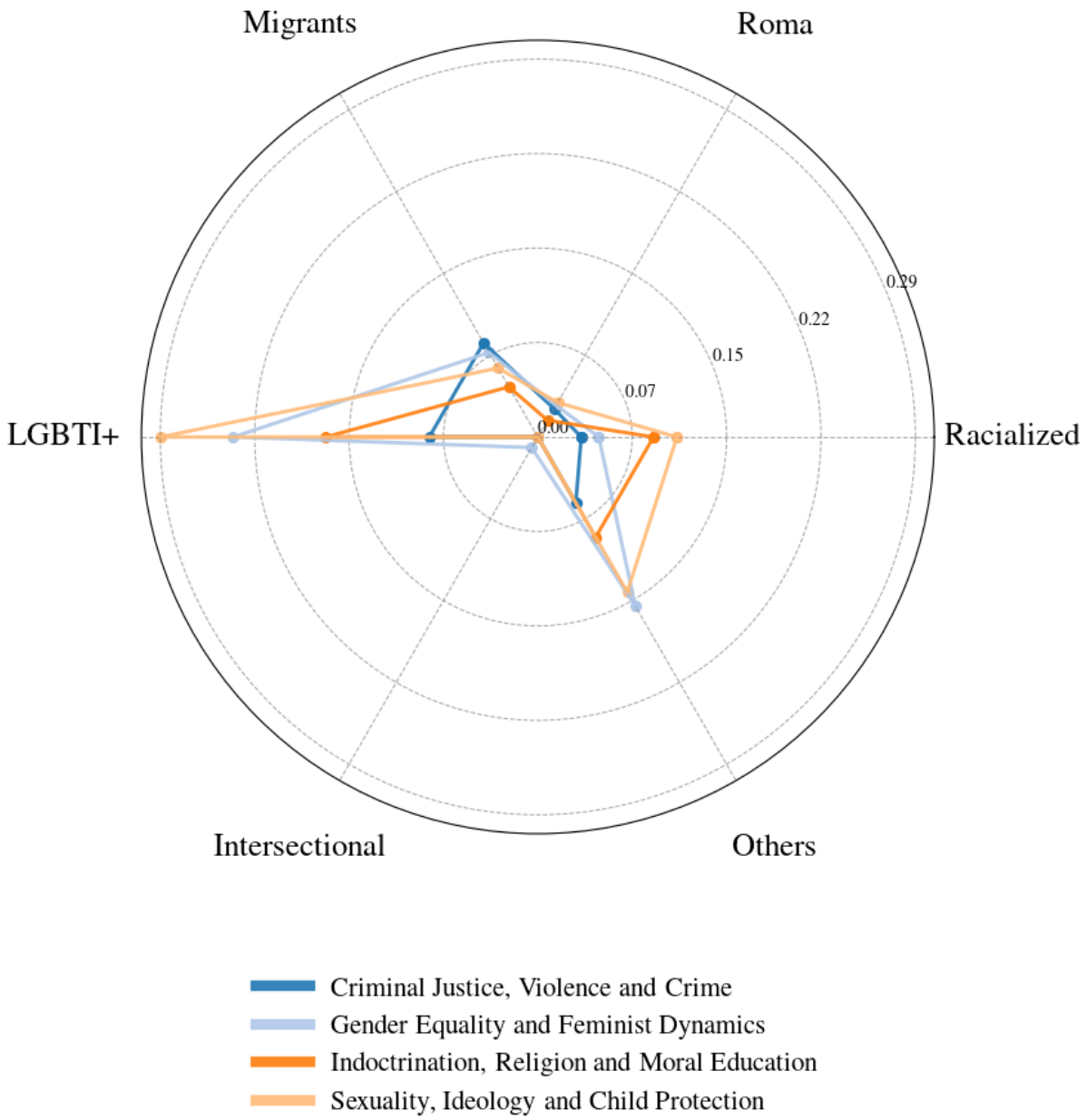


FIGURE D.8. HS targets probability distribution across *'Sexual Crimes, Feminism/Abortion, Religious Education and Gender/Sexuality Debates'*.

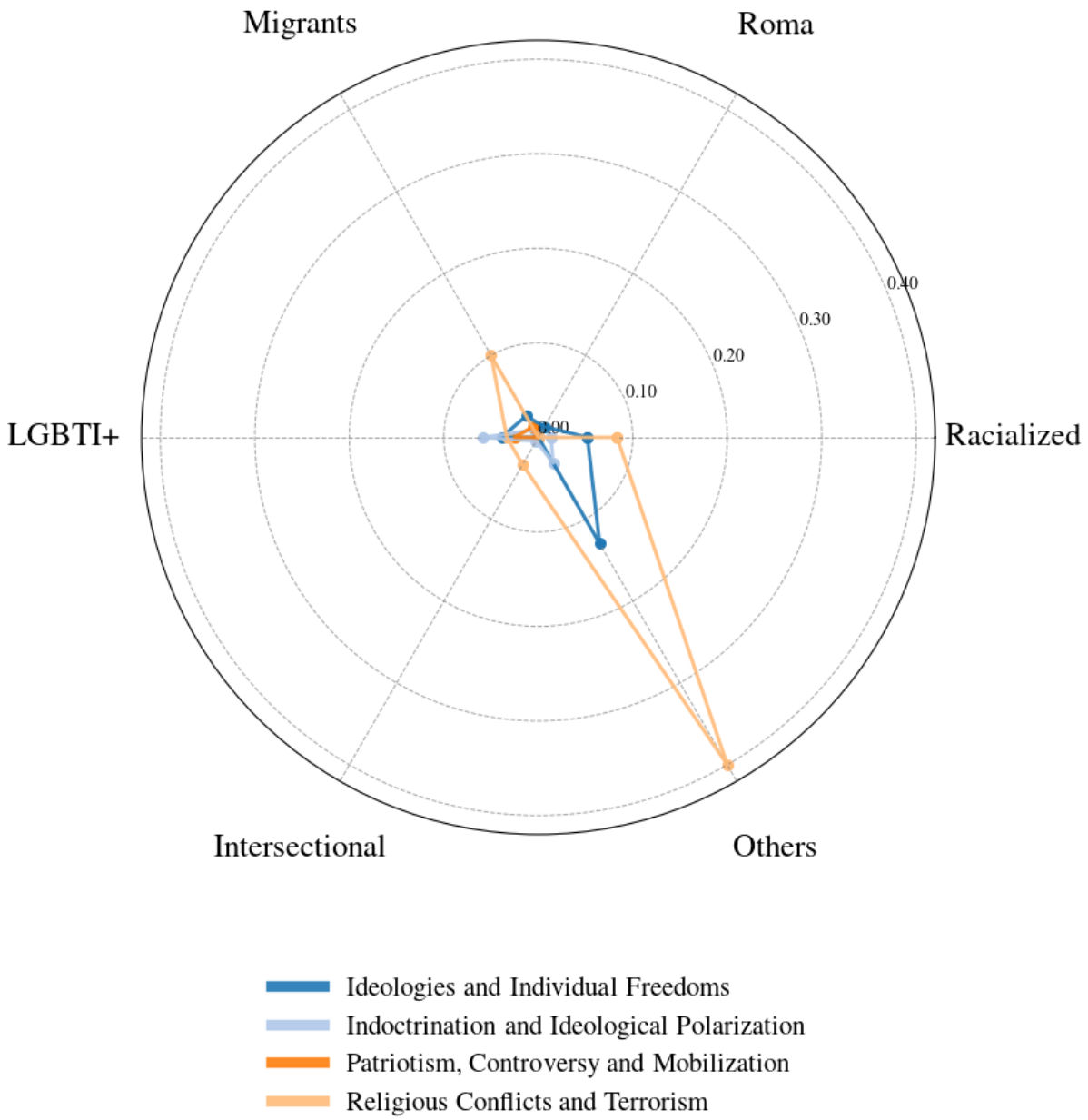


FIGURE D.9. HS targets probability distribution across *'Religious Conflicts, Political Ideologies, Extreme Disputes, and Relations between Belief, Morals and Politics'*.

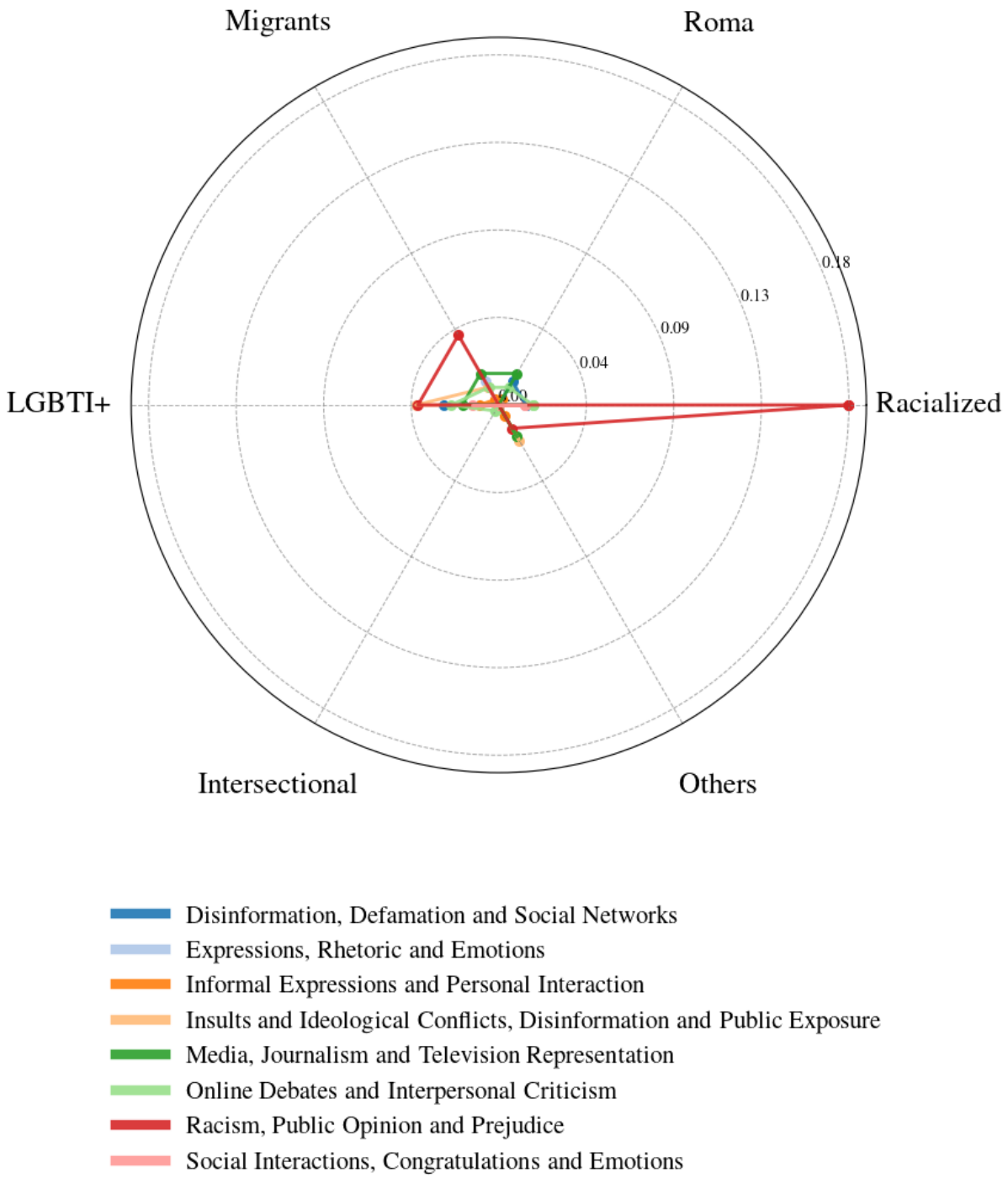


FIGURE D.10. HS targets probability distribution across *'Discourse of Social and Digital Interaction, Internet Colloquial Tone'*.

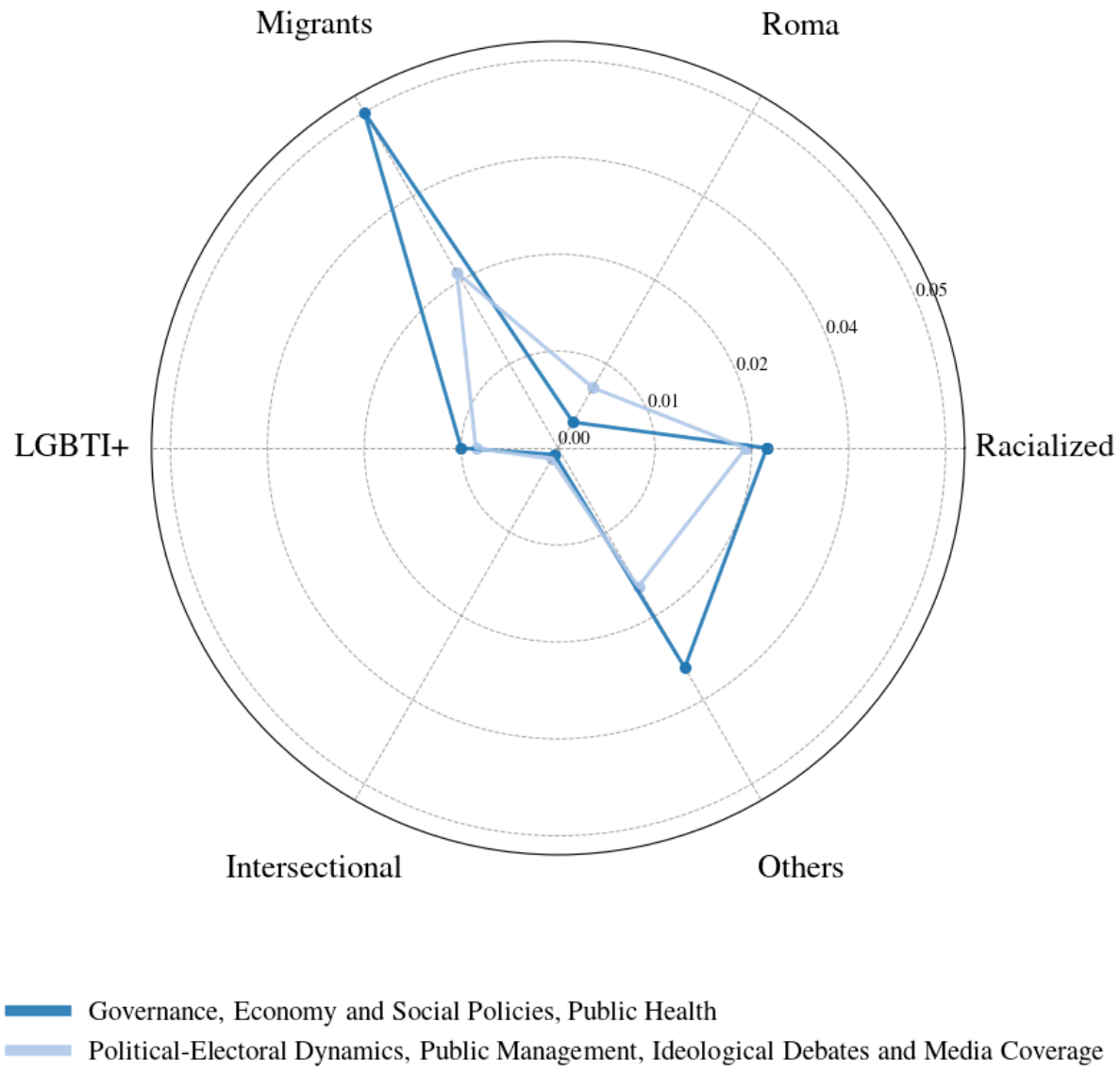


FIGURE D.11. HS targets probability distribution across *'Party, Electoral and Parliamentary Dynamics'* Dimension.

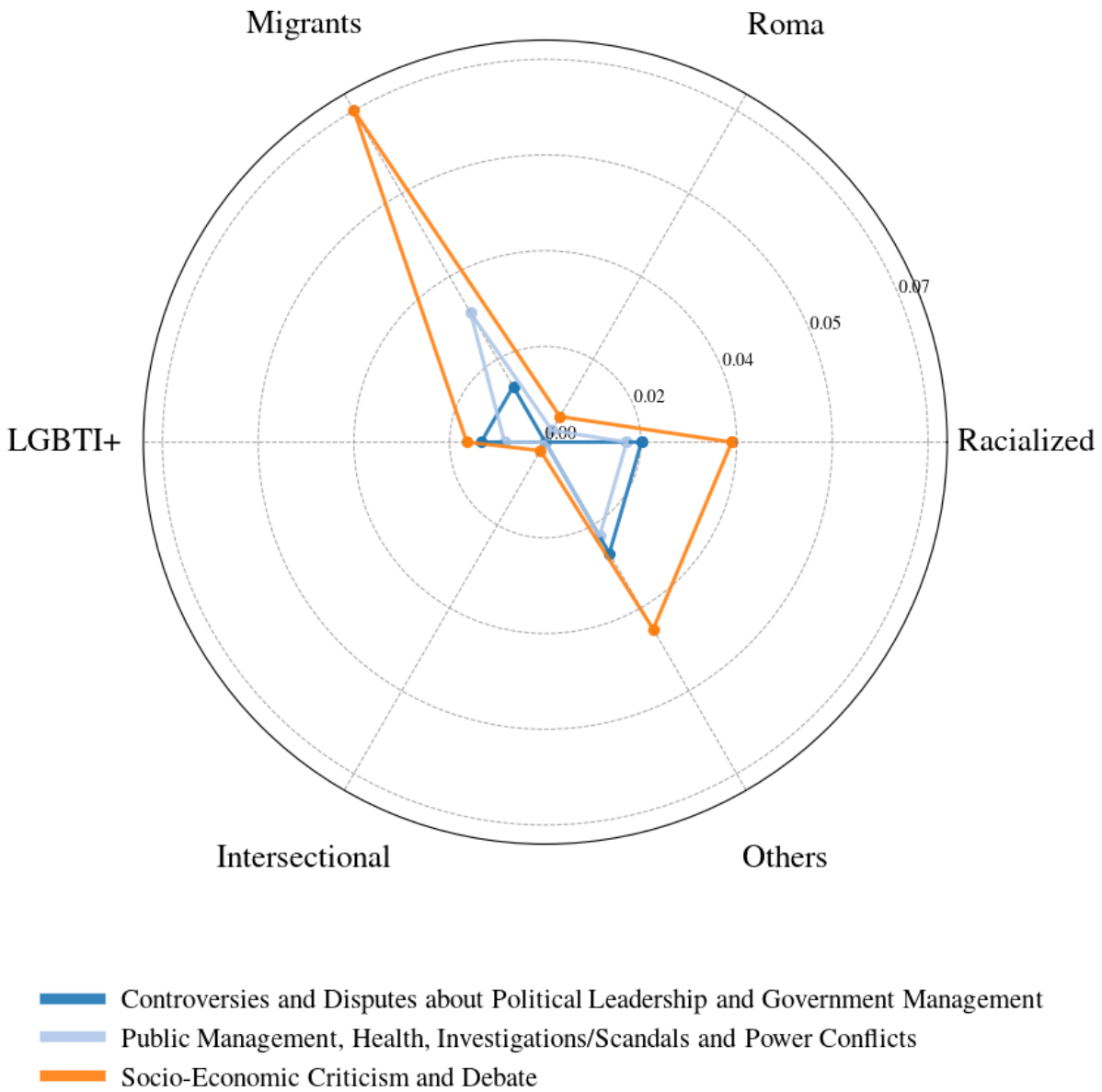


FIGURE D.12. HS targets probability distribution across *'Governance, Economy and Social Policies, Public Health'*.

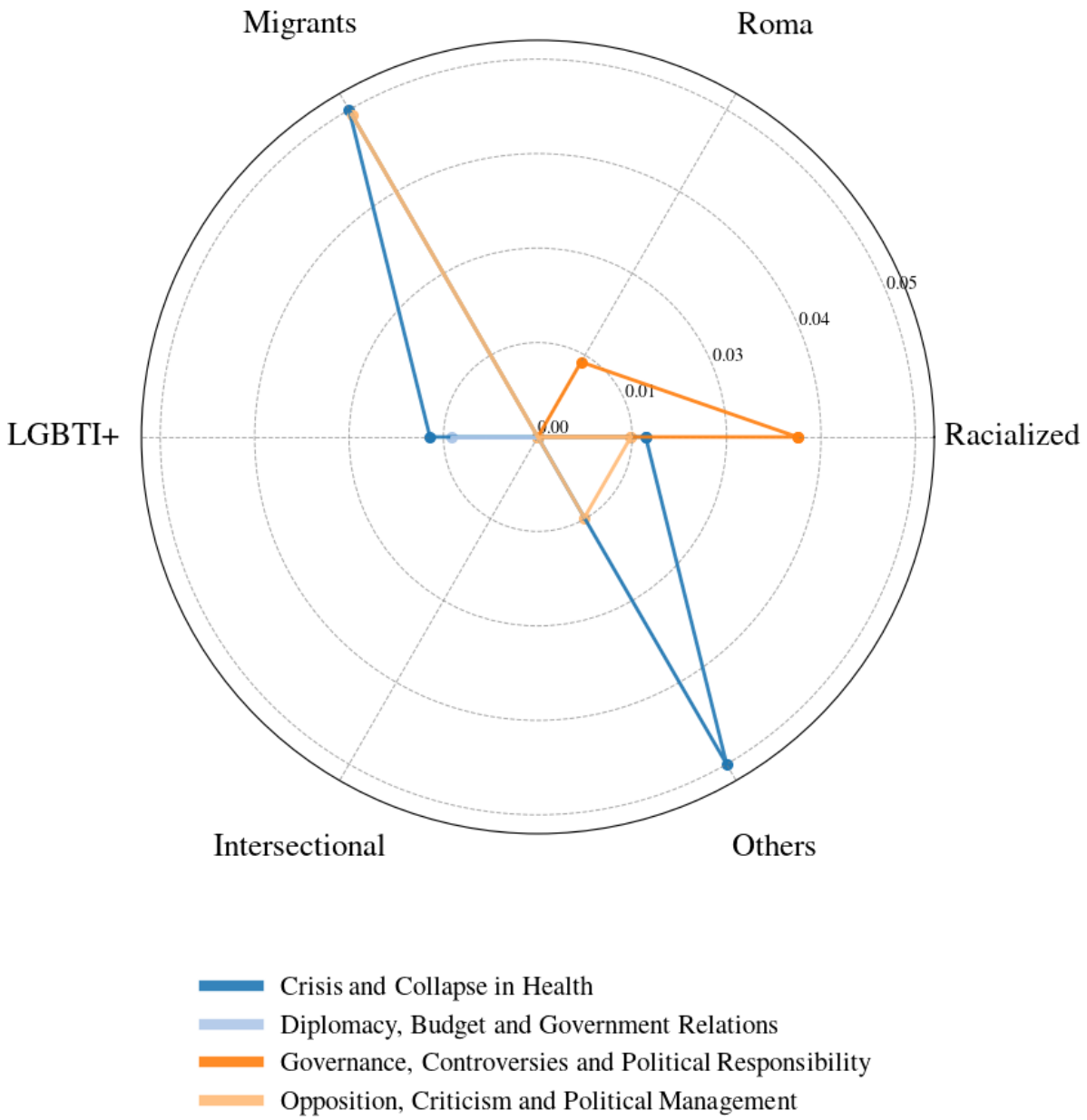


FIGURE D.13. HS targets probability distribution across *'Public Management, Health, Investigations/Scandals and Power Conflicts'*.

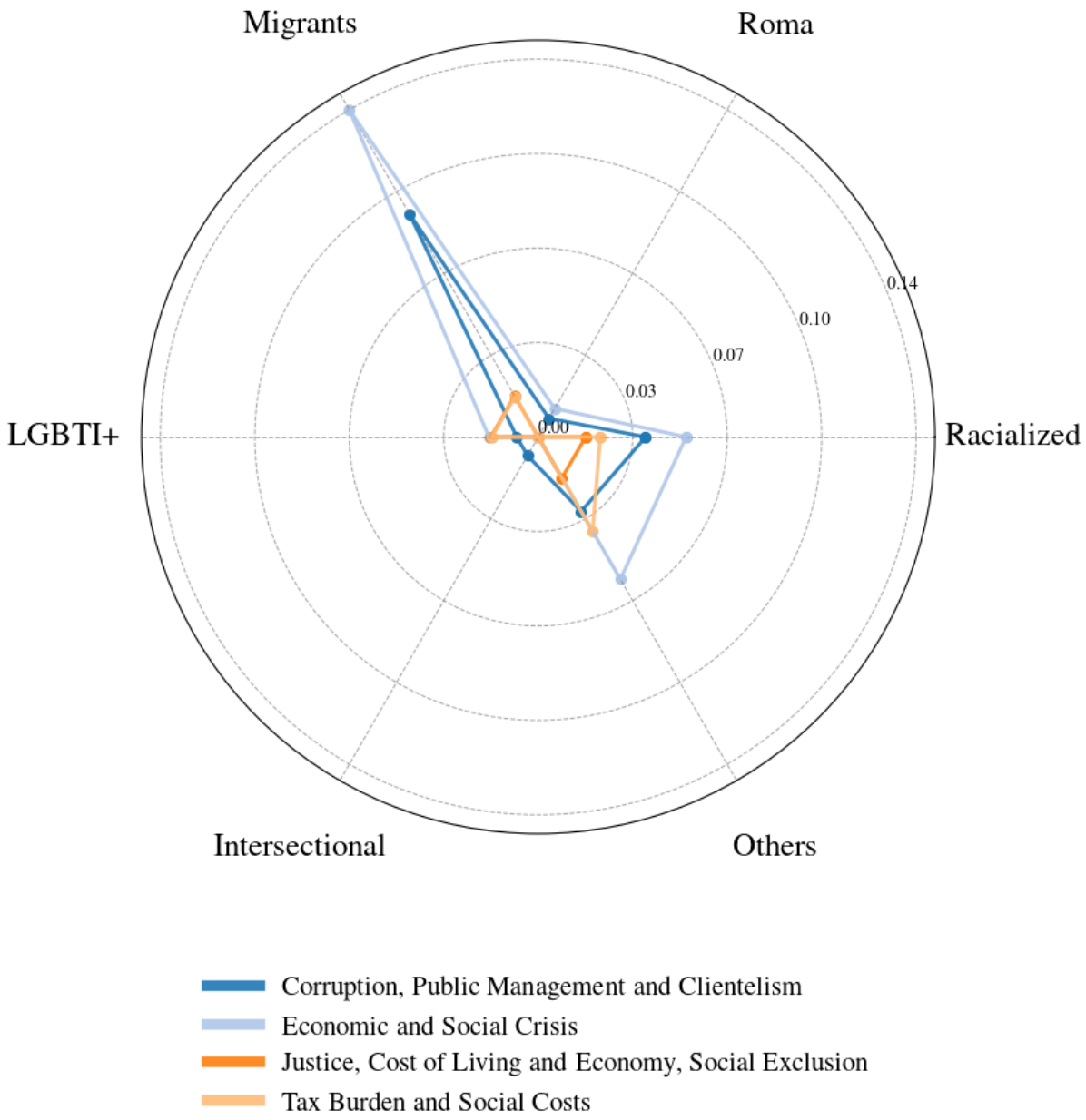


FIGURE D.14. HS targets probability distribution across *'Socio-Economic Criticism and Debate'*.

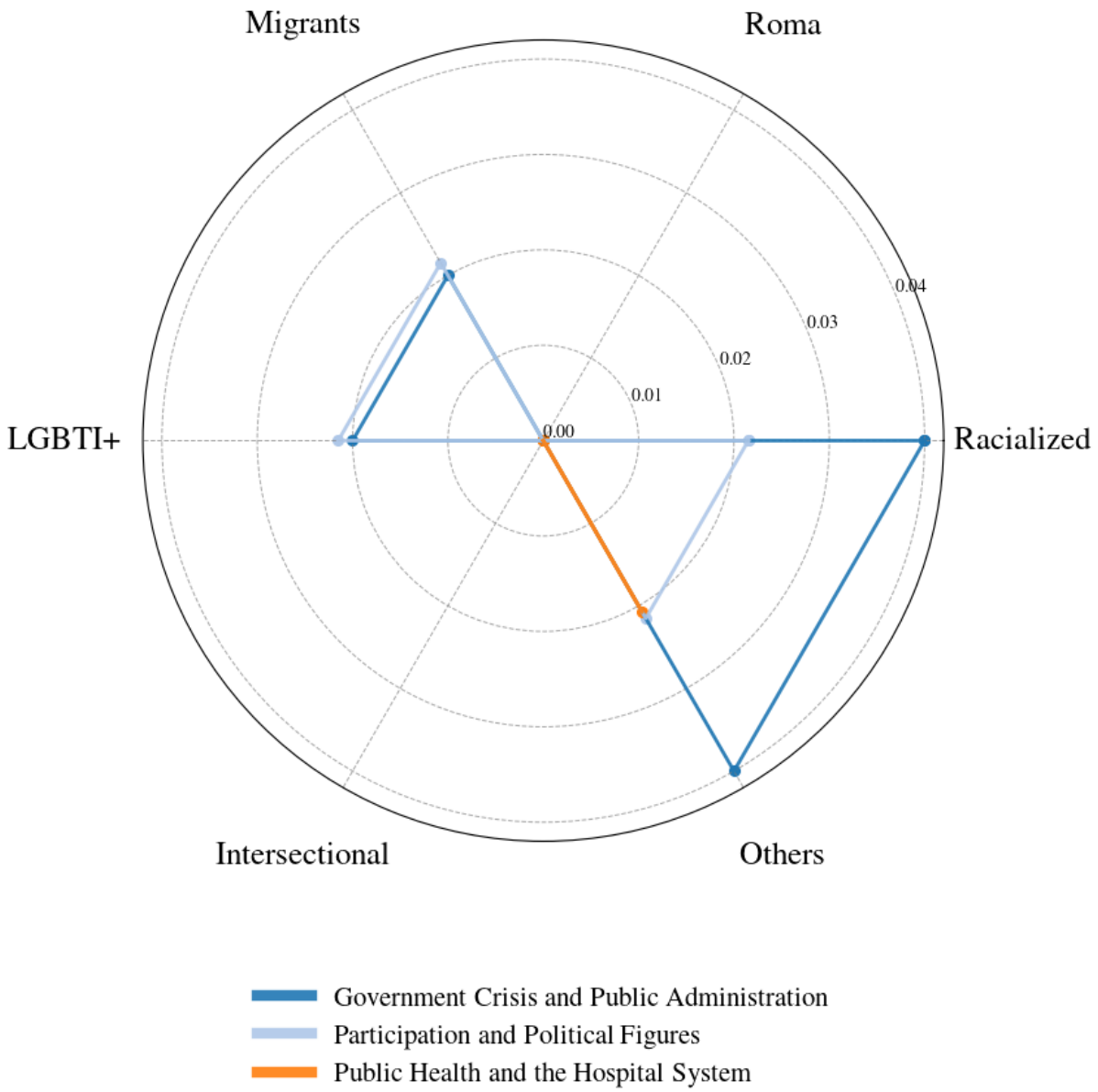


FIGURE D.15. HS targets probability distribution across *'Controversies and Disputes about Political Leadership and Government Management'*.

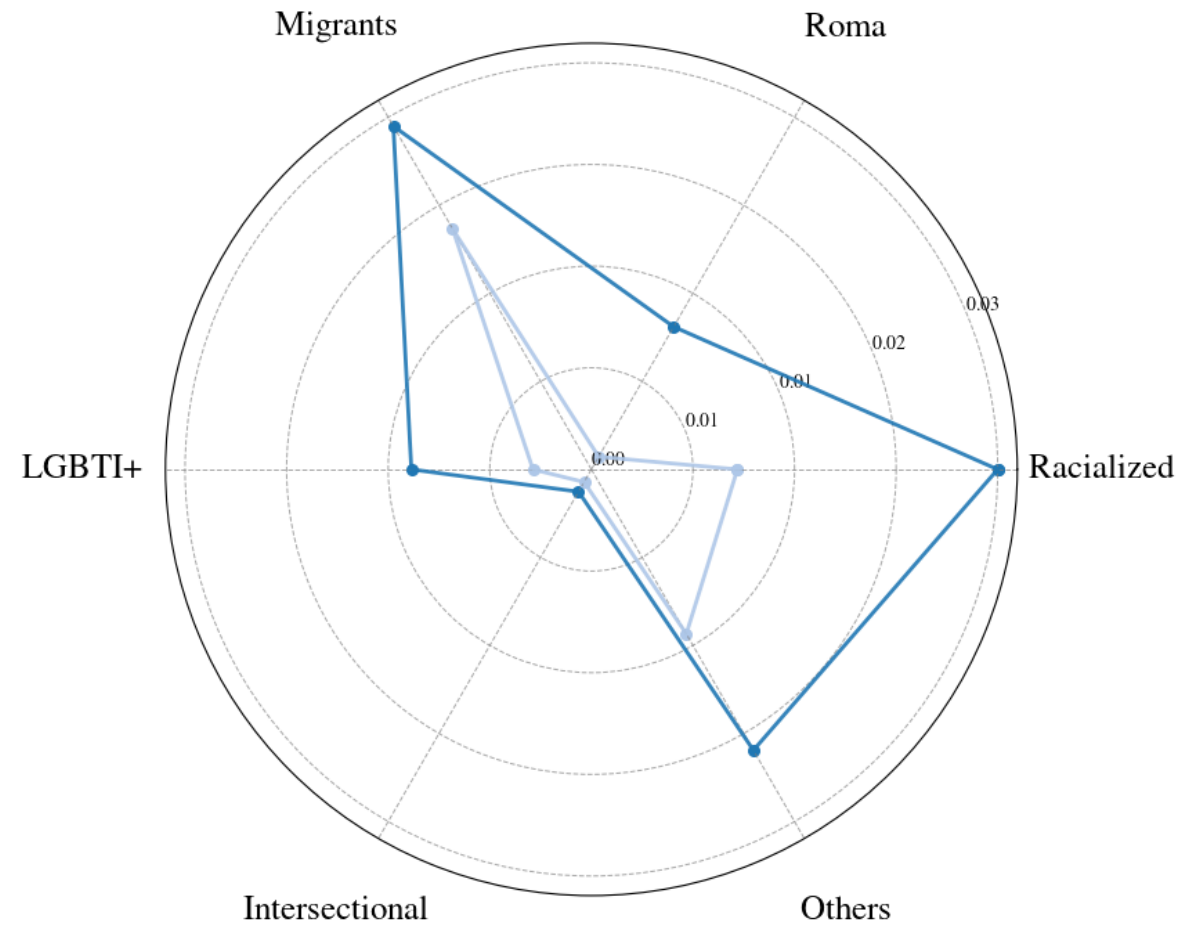


FIGURE D.16. HS targets probability distribution across *'Political-Electoral Dynamics, Public Management, Ideological Debates and Media Coverage'*.

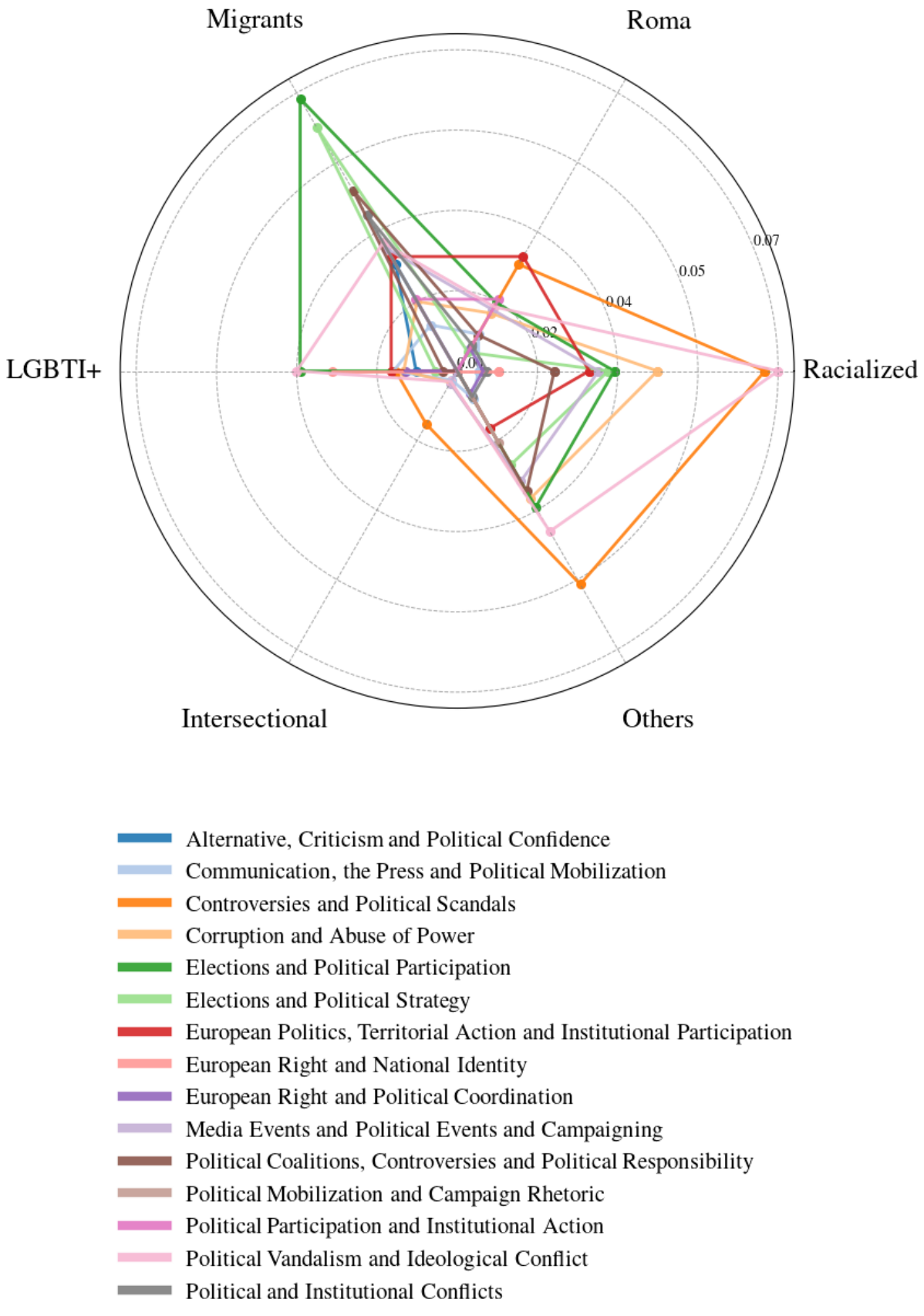


FIGURE D.17. HS targets probability distribution across *'Political Dispute, Ideological Positions, Communication and Media'*.

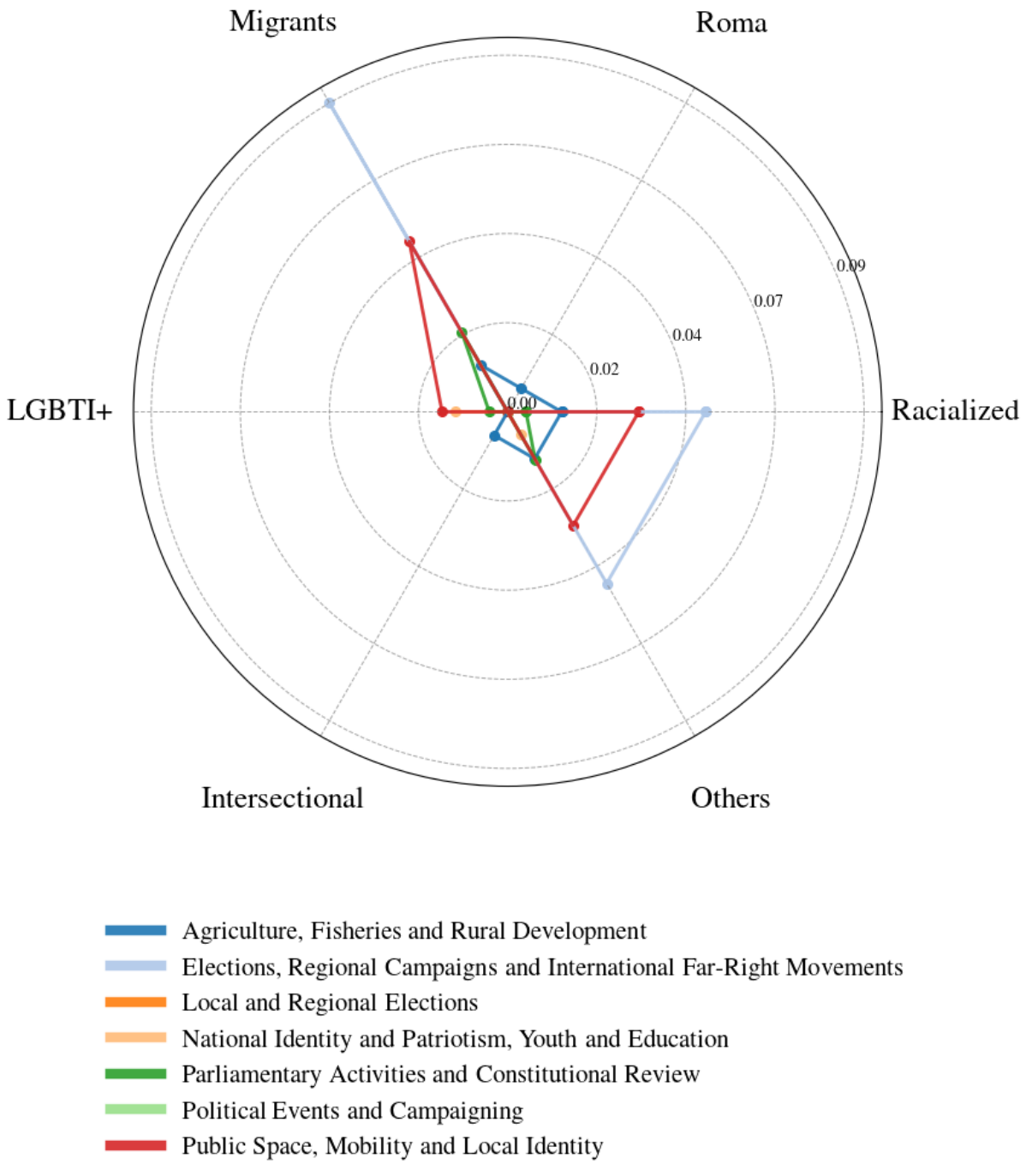


FIGURE D.18. HS targets probability distribution across *'Political-Parliamentary and Electoral Cycle, Mobilization, Interaction and Event Organization'*.

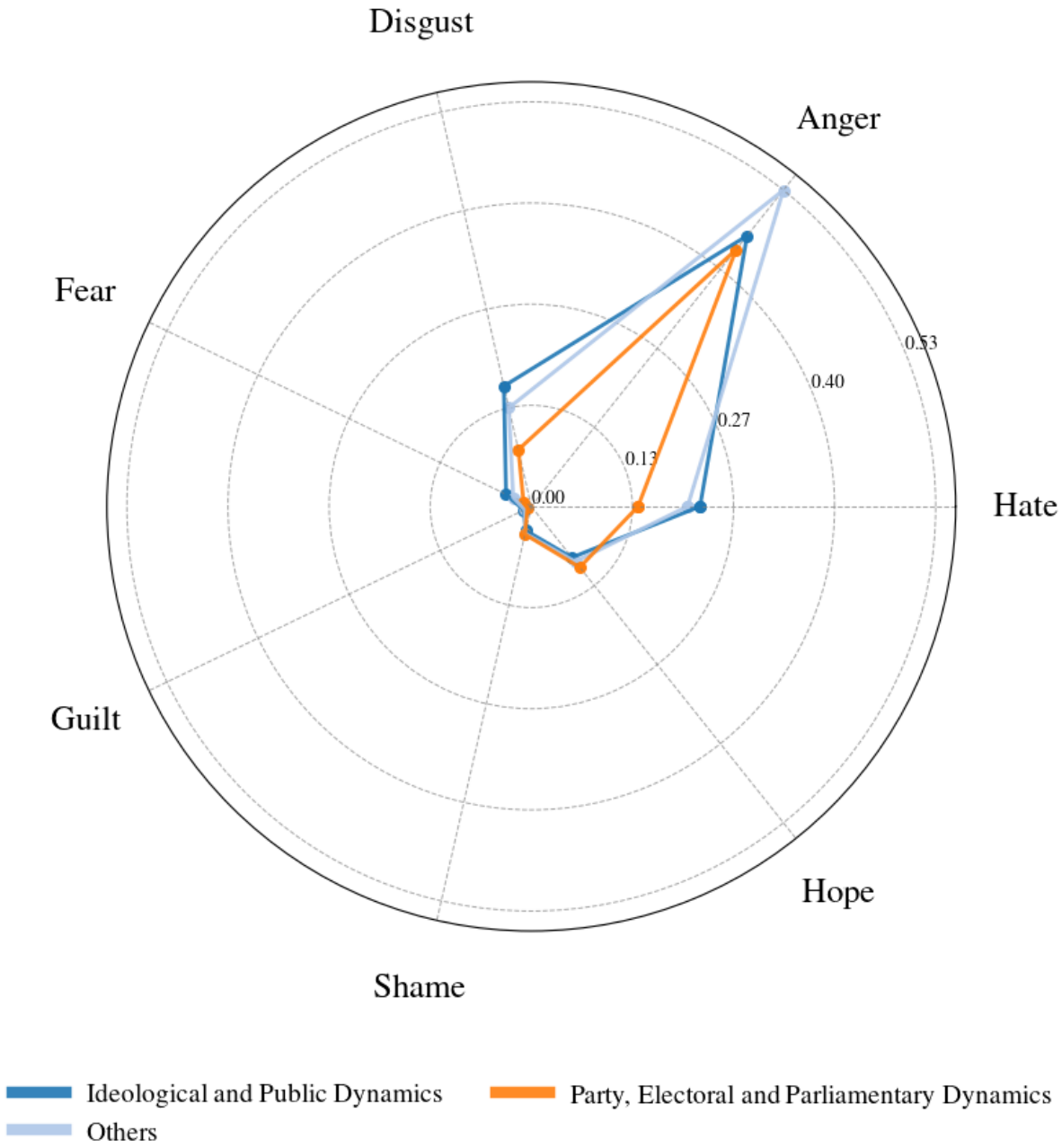


FIGURE D.19. Emotions probability distribution across *'Political Discourse'* dimension.

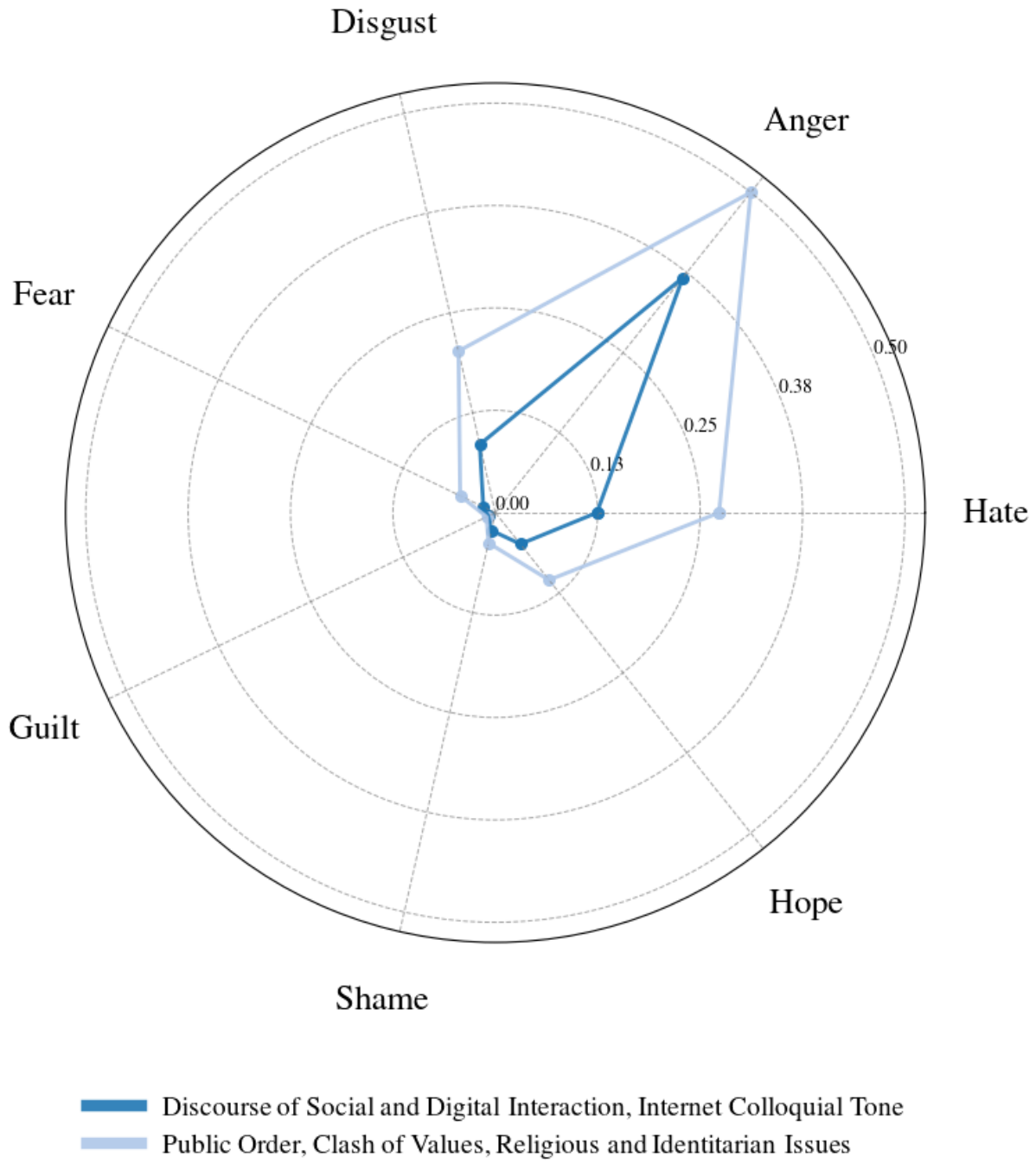


FIGURE D.20. Emotions probability distribution across *'Ideological and Public Dynamics'* Dimension.

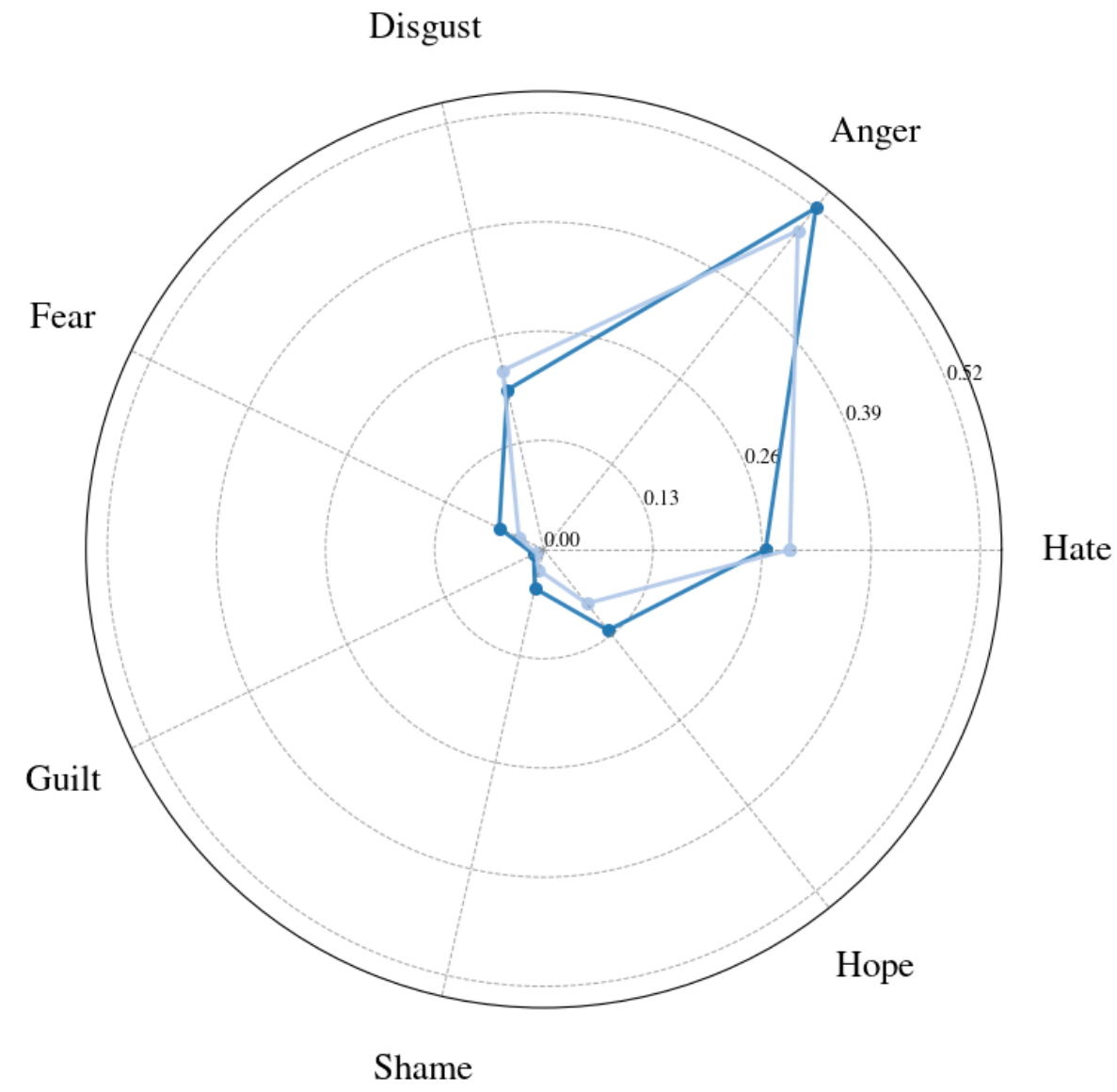


FIGURE D.21. Emotions probability distribution across *'Public Order, Clash of Values, Religious and Identitarian Issues'*.

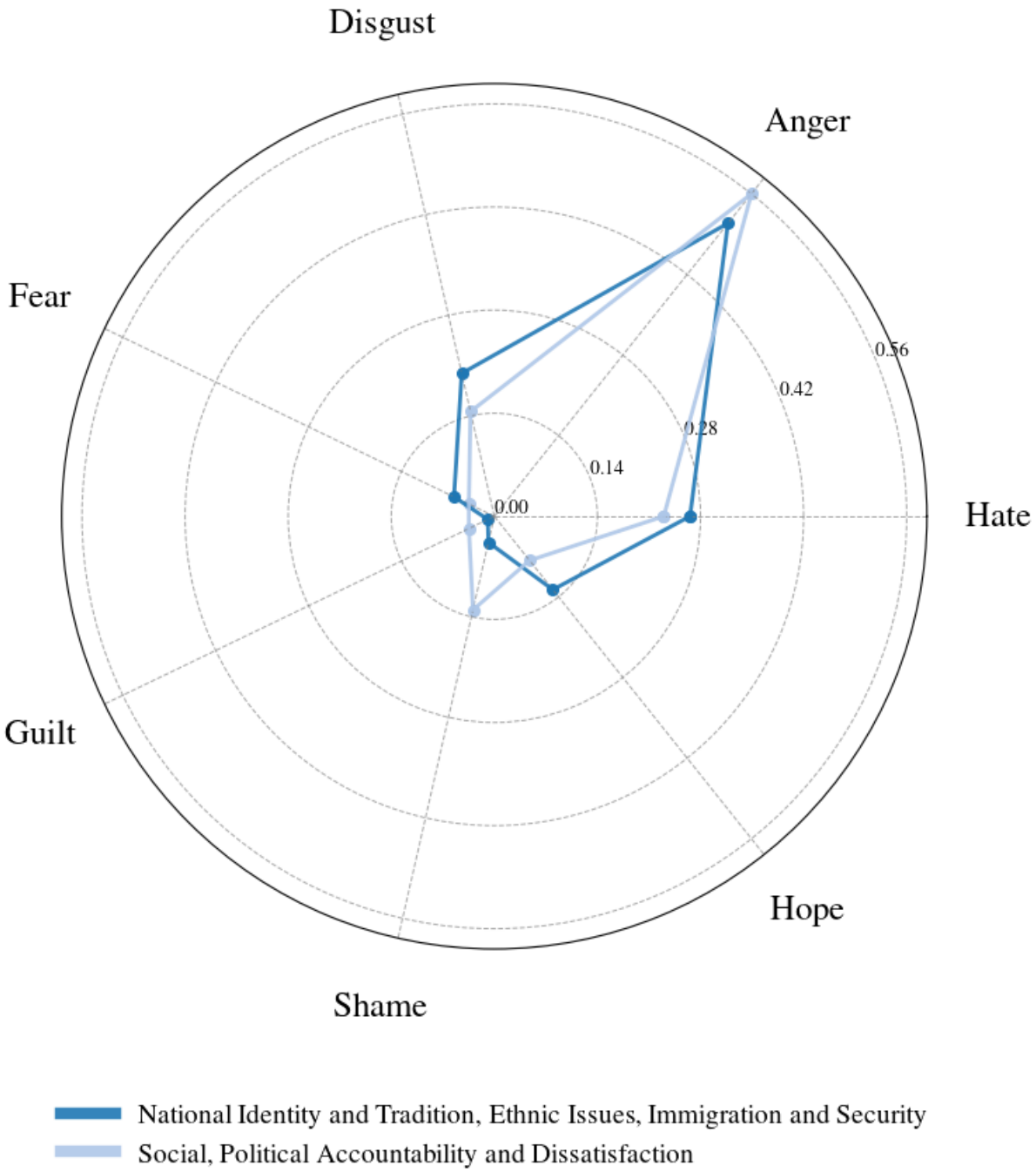


FIGURE D.22. Emotions probability distribution across *'Conservative Approach, Public Order, Nationalism/Patriotism and Traditional Values'*.

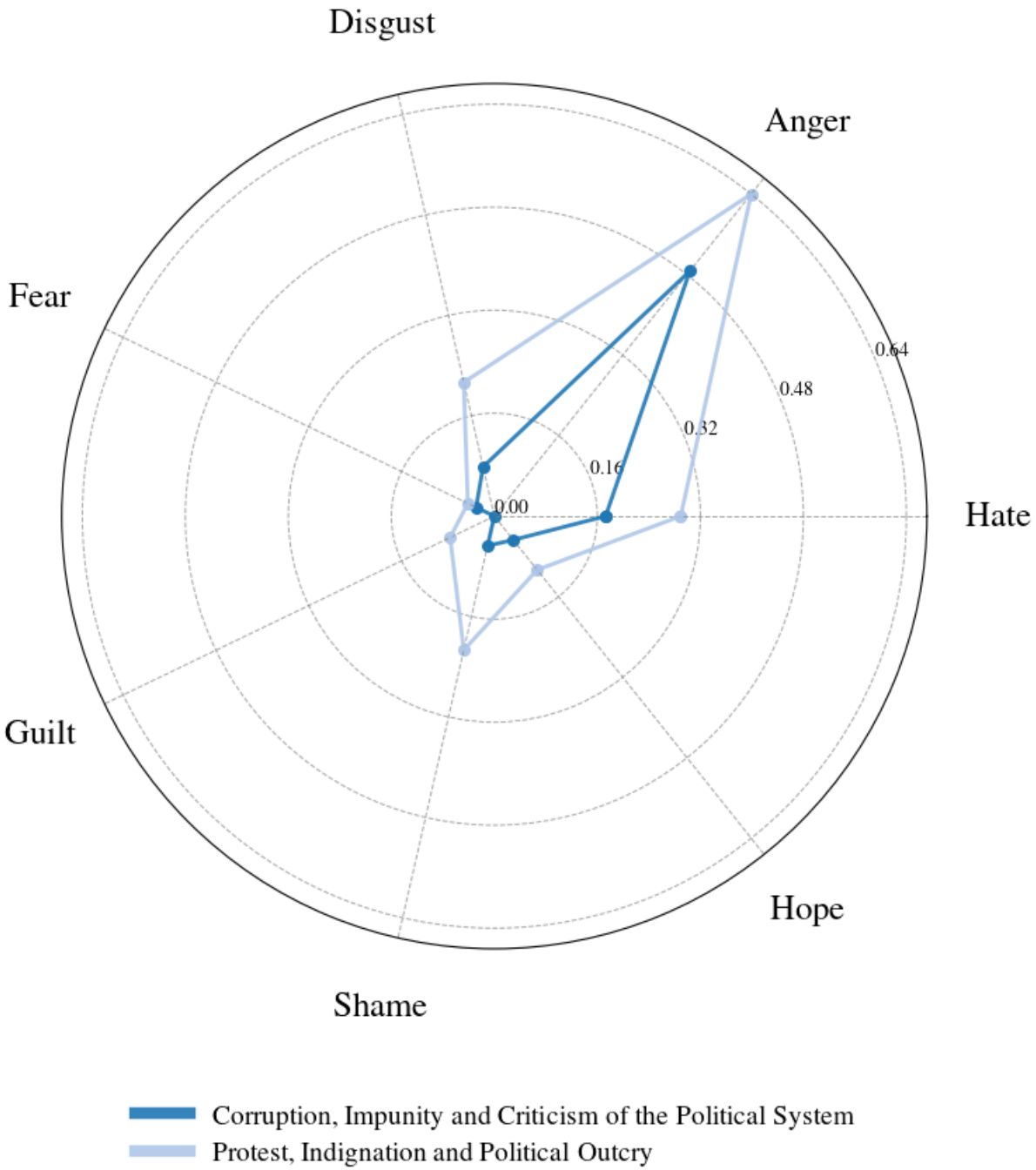


FIGURE D.23. Emotions probability distribution across *'Social, Political Accountability and Dissatisfaction'*.

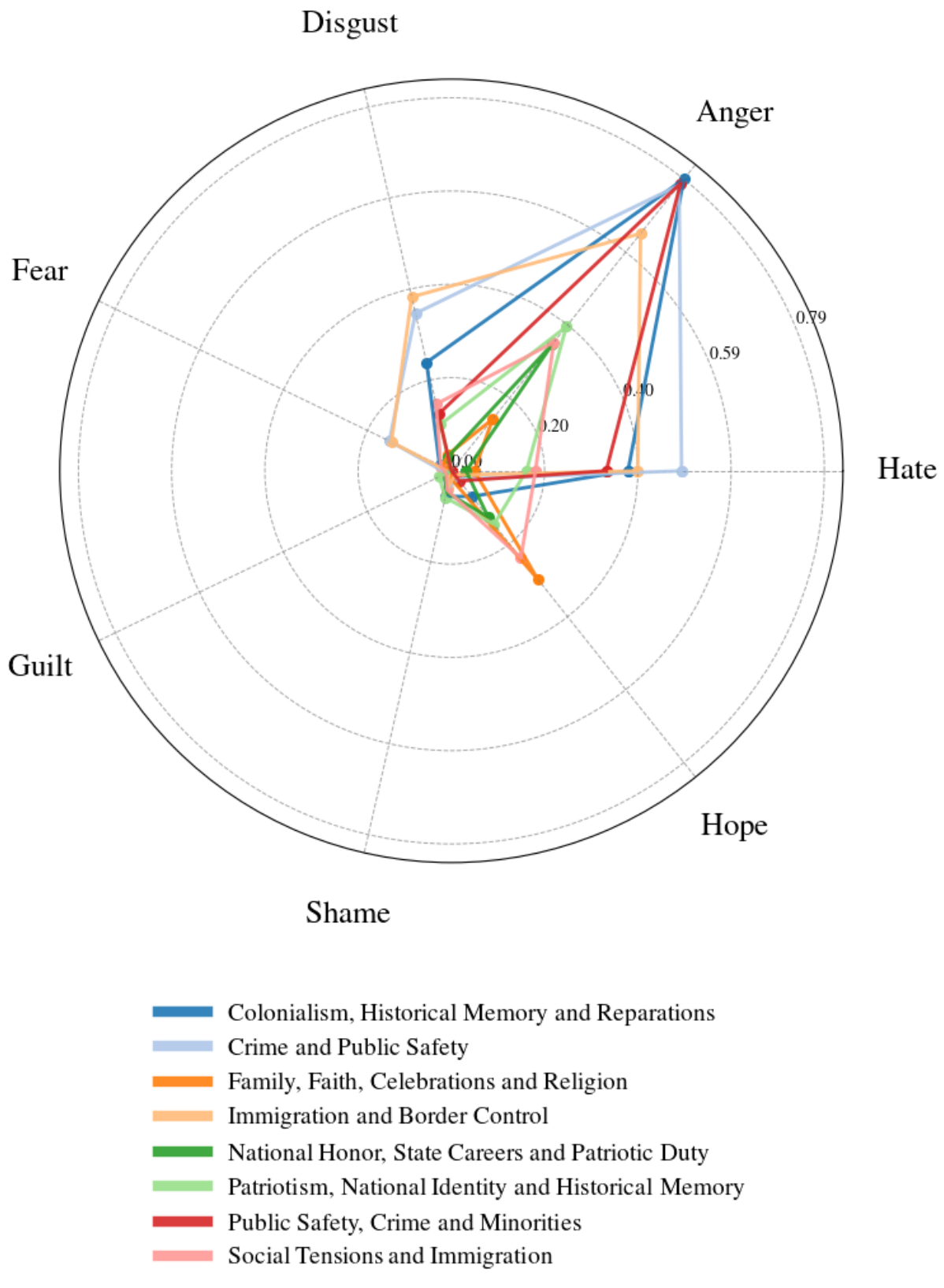


FIGURE D.24. Emotions probability distribution across *'National Identity and Tradition, Ethnic Issues, Immigration and Security'*.

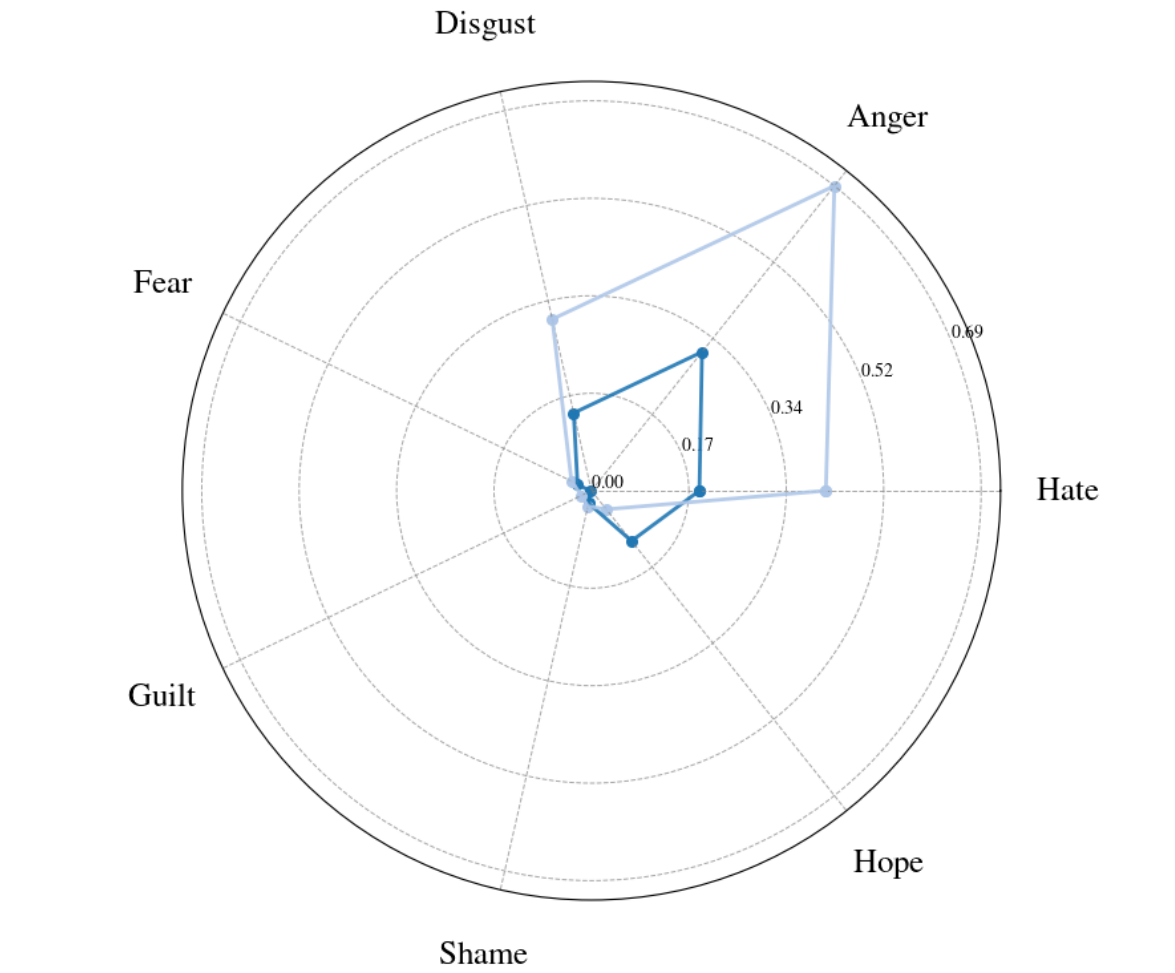


FIGURE D.25. Emotions probability distribution across *'Ideological Positions, Moral, Sexual and Religious Debates'*.

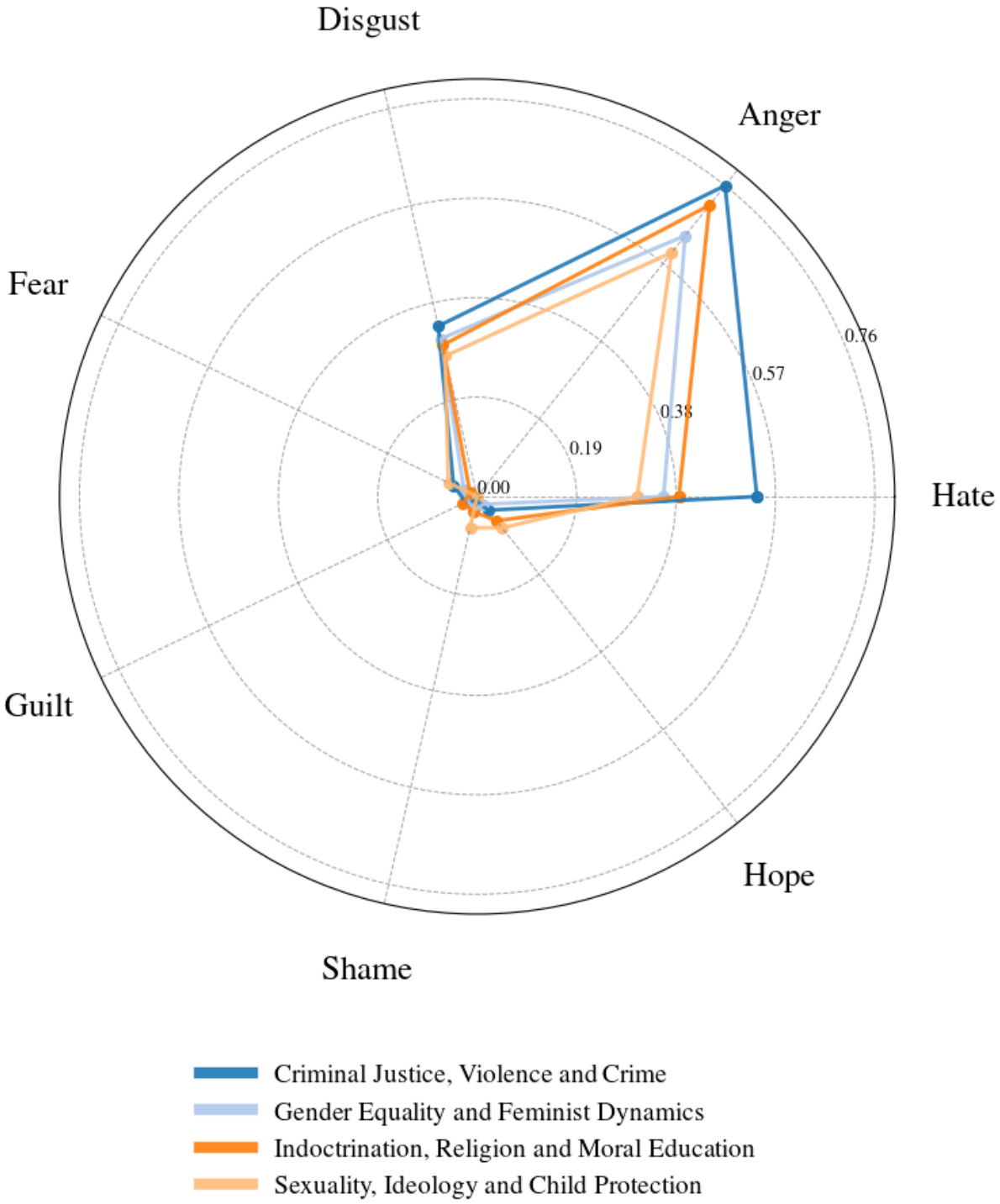


FIGURE D.26. Emotions probability distribution across *'Sexual Crimes, Feminism/Abortion, Religious Education and Gender/Sexuality Debates'*.

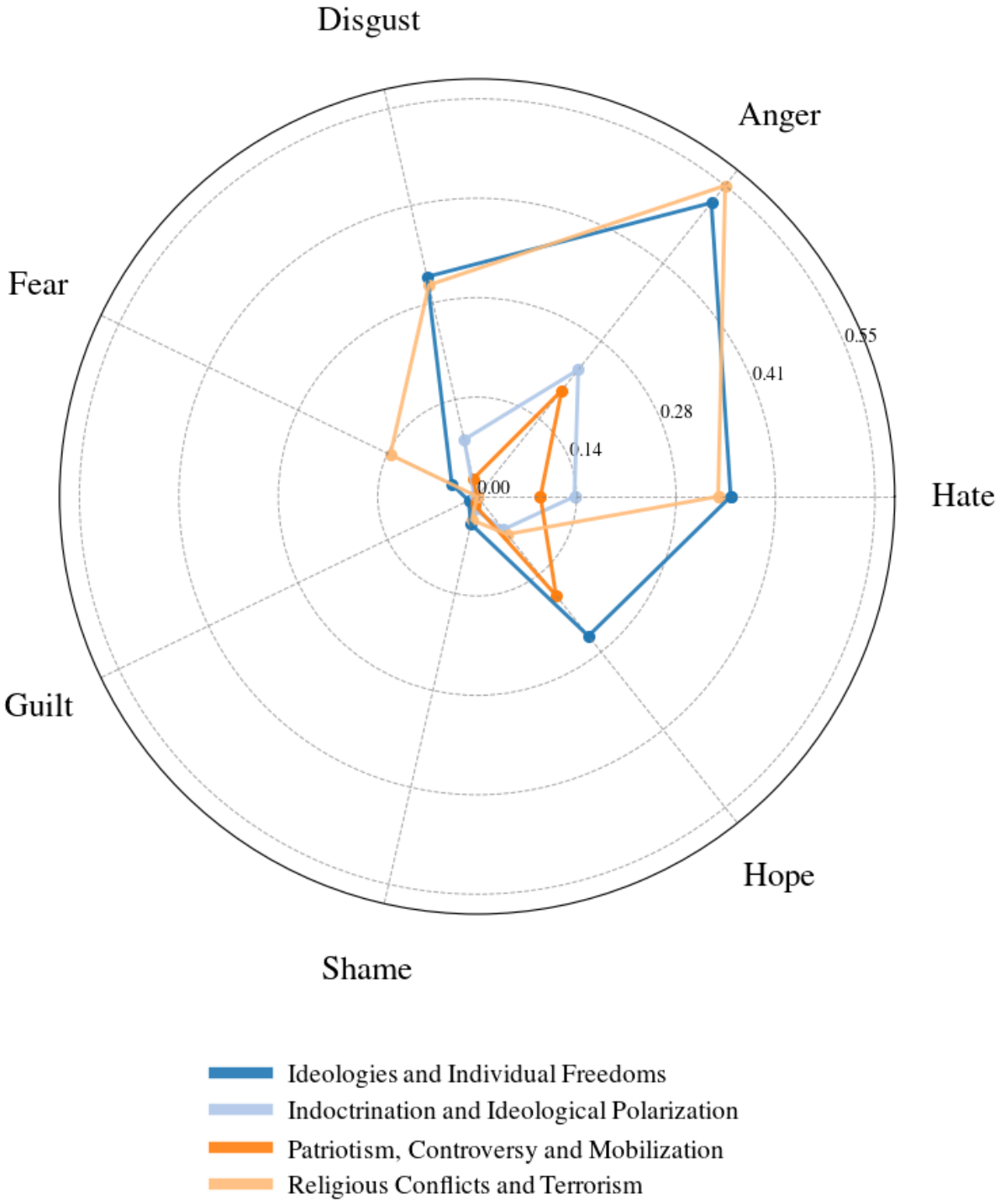


FIGURE D.27. Emotions probability distribution across *'Religious Conflicts, Political Ideologies, Extreme Disputes, and Relations between Belief, Morals and Politics'*.

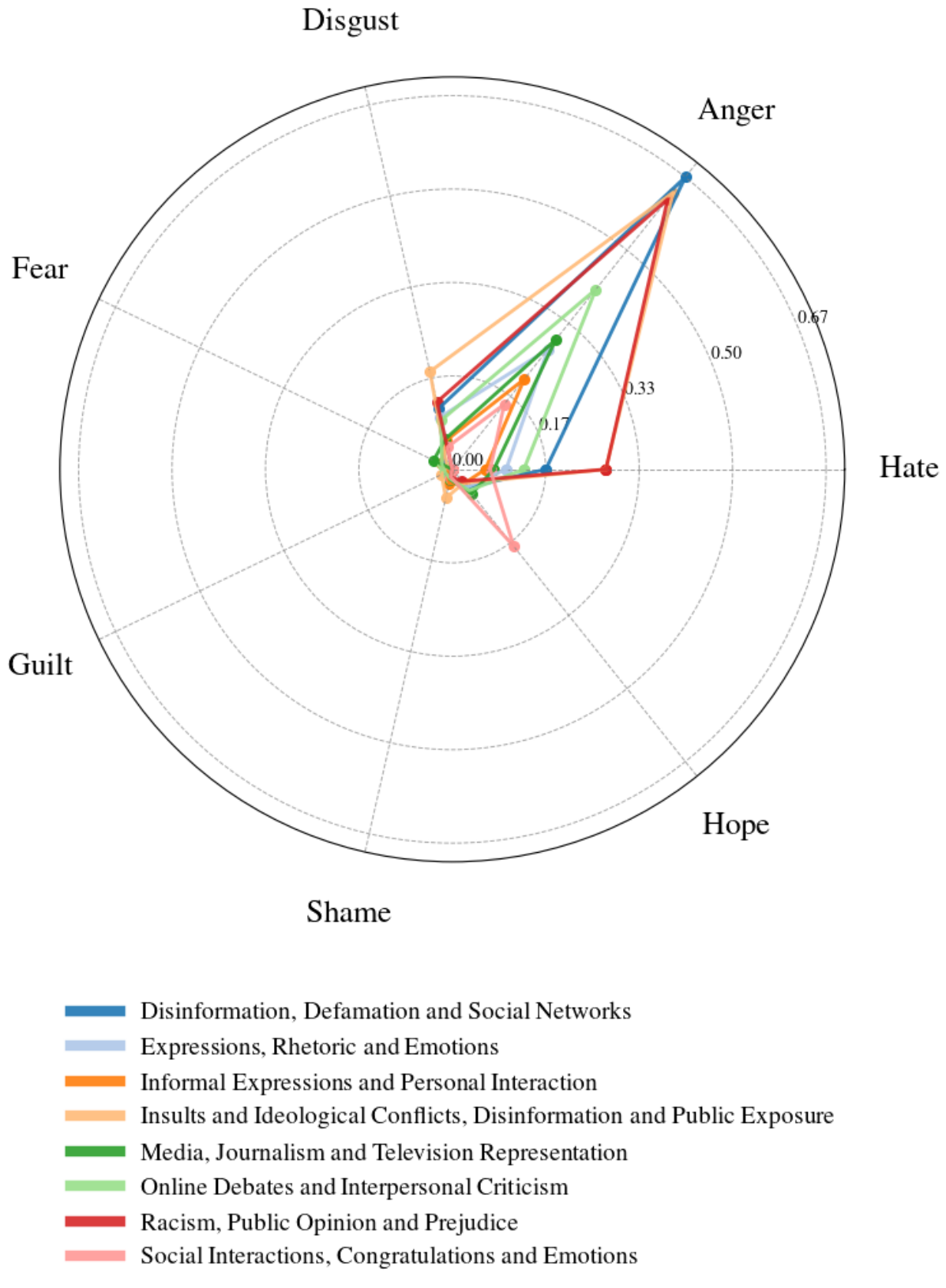


FIGURE D.28. Emotions probability distribution across *'Discourse of Social and Digital Interaction, Internet Colloquial Tone'*.

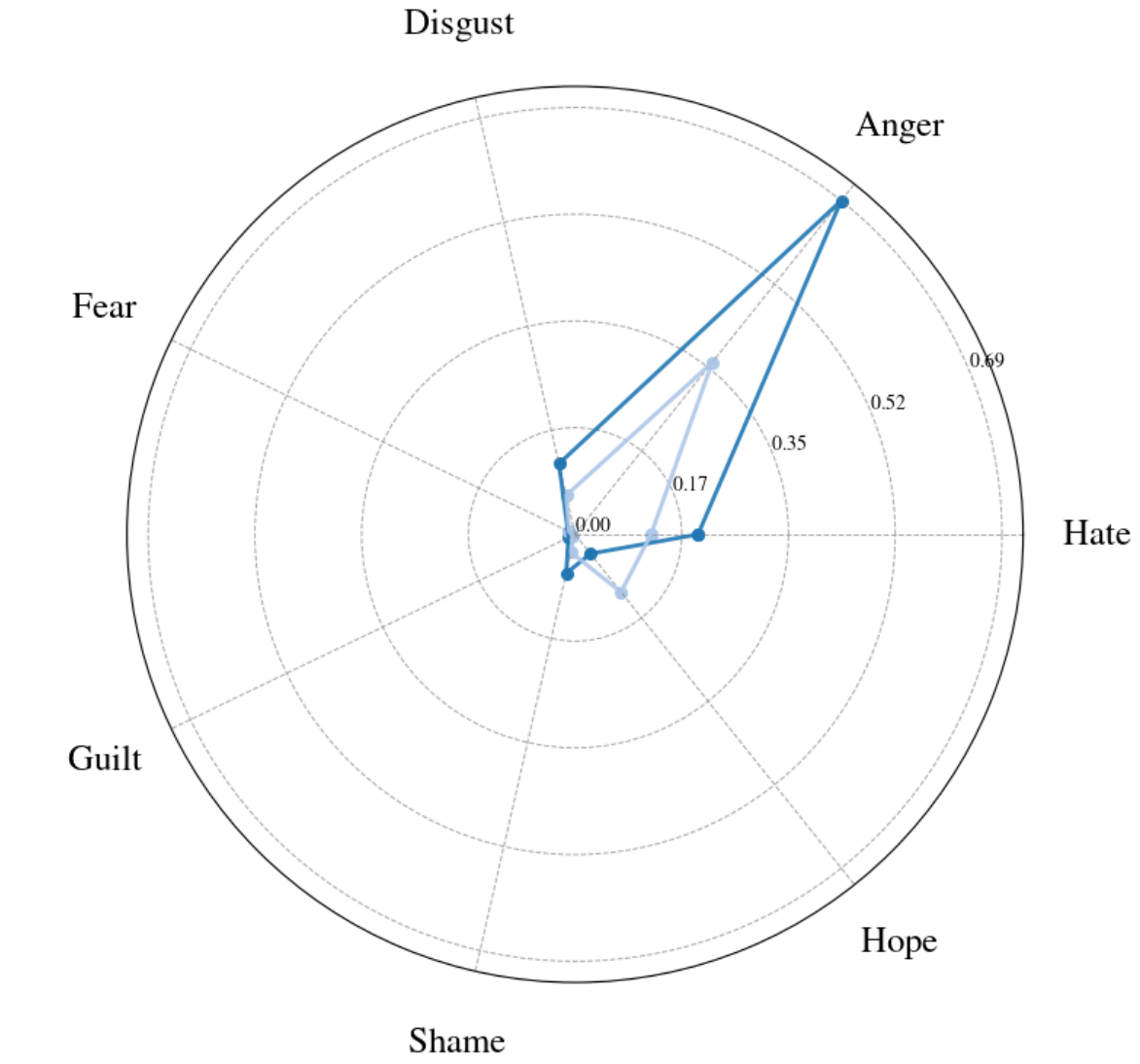


FIGURE D.29. Emotions probability distribution across *'Party, Electoral and Parliamentary Dynamics'* Dimension.

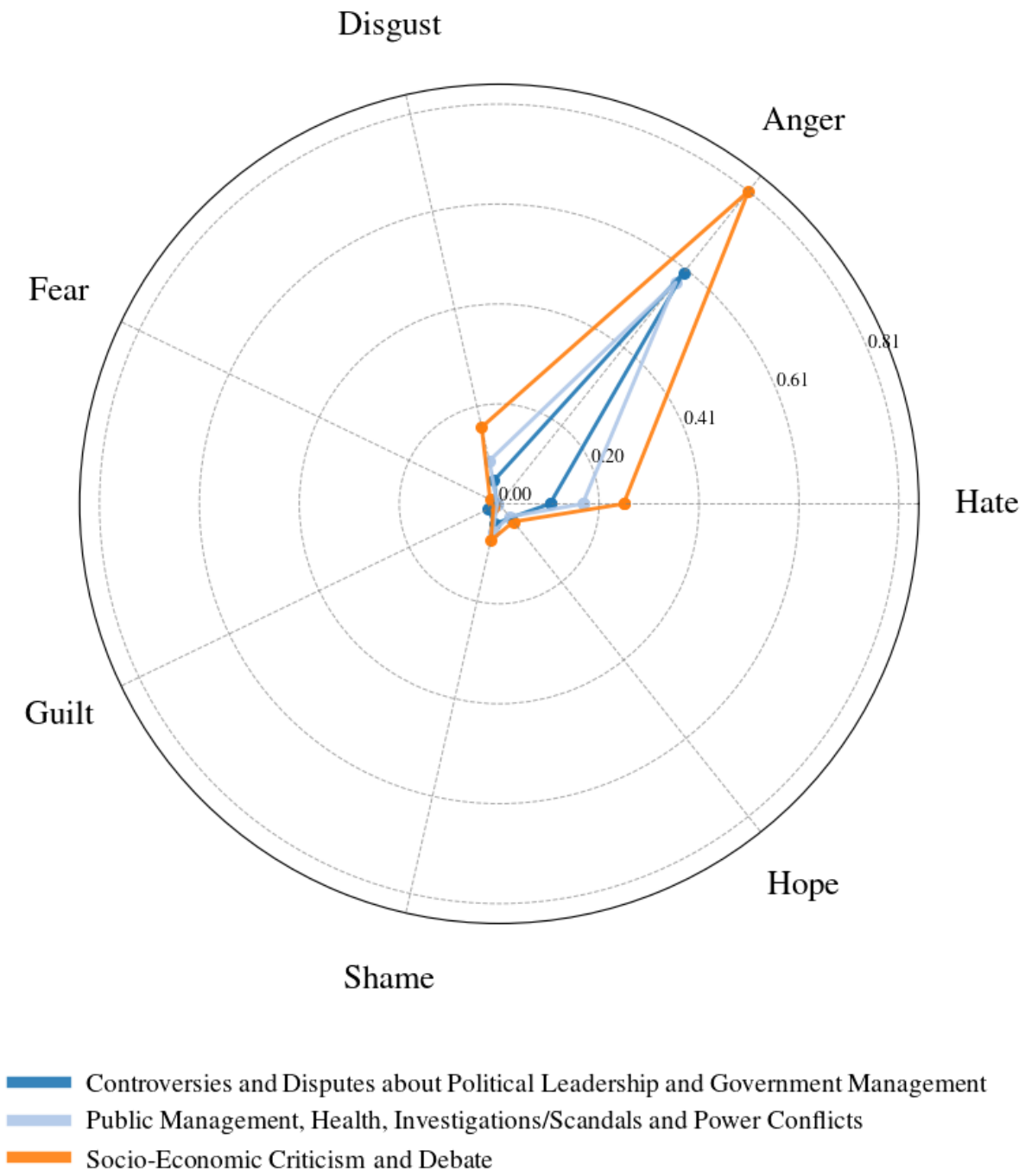


FIGURE D.30. Emotions probability distribution across *'Governance, Economy and Social Policies, Public Health'*.

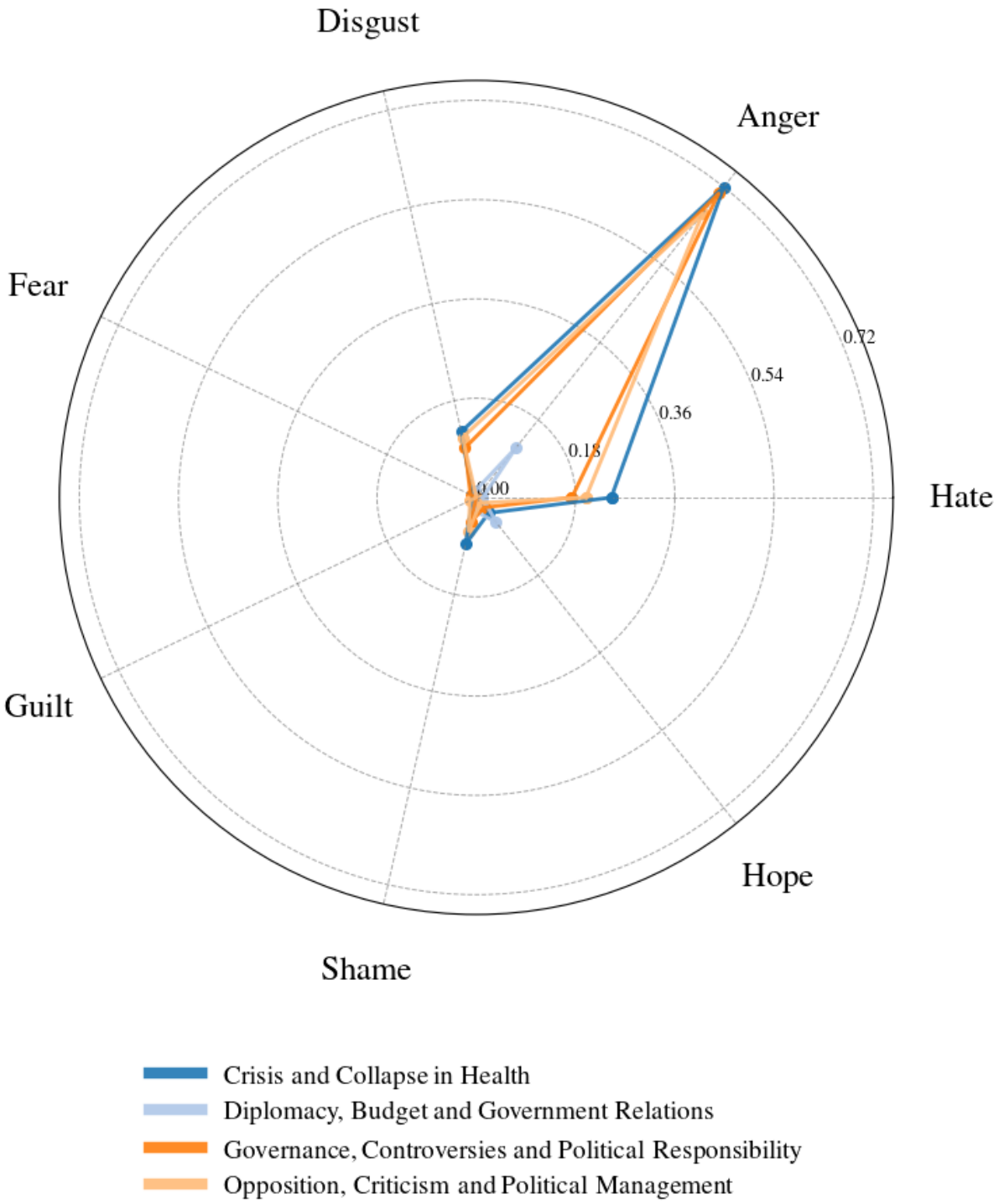


FIGURE D.31. Emotions probability distribution across *'Public Management, Health, Investigations/Scandals and Power Conflicts'*.

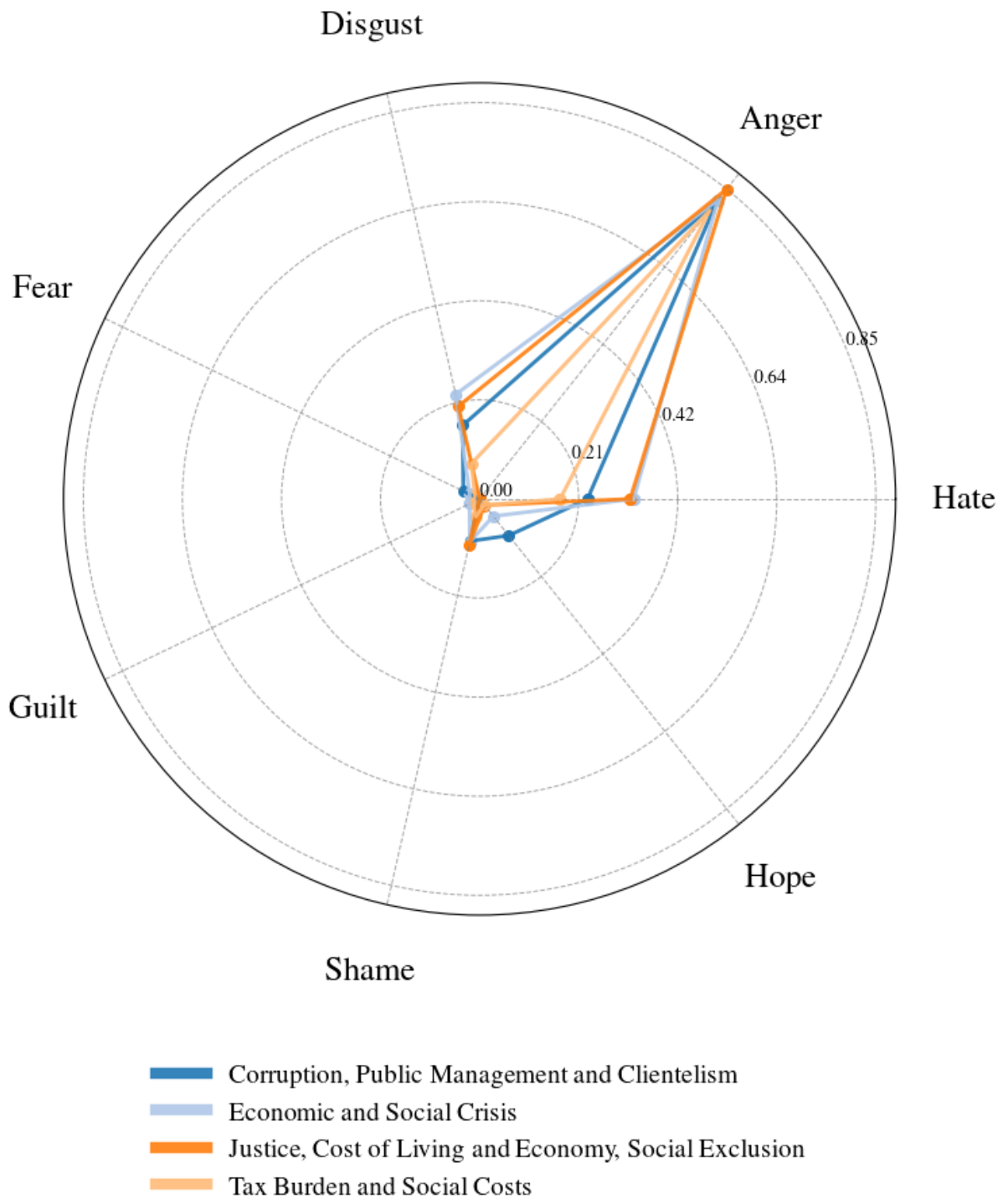


FIGURE D.32. Emotions probability distribution across *'Socio-Economic Criticism and Debate'*.

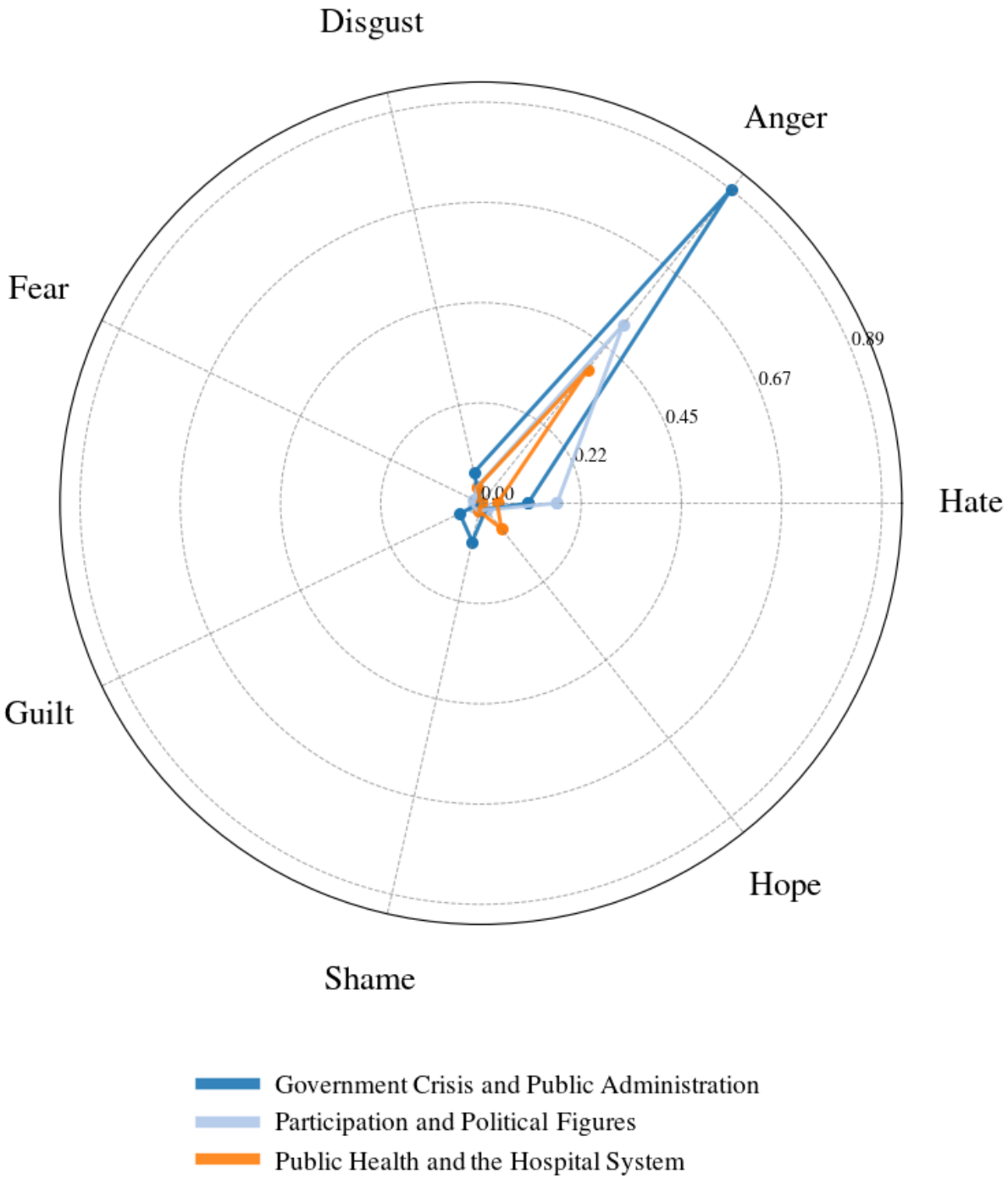


FIGURE D.33. Emotions probability distribution across *'Controversies and Disputes about Political Leadership and Government Management'*.

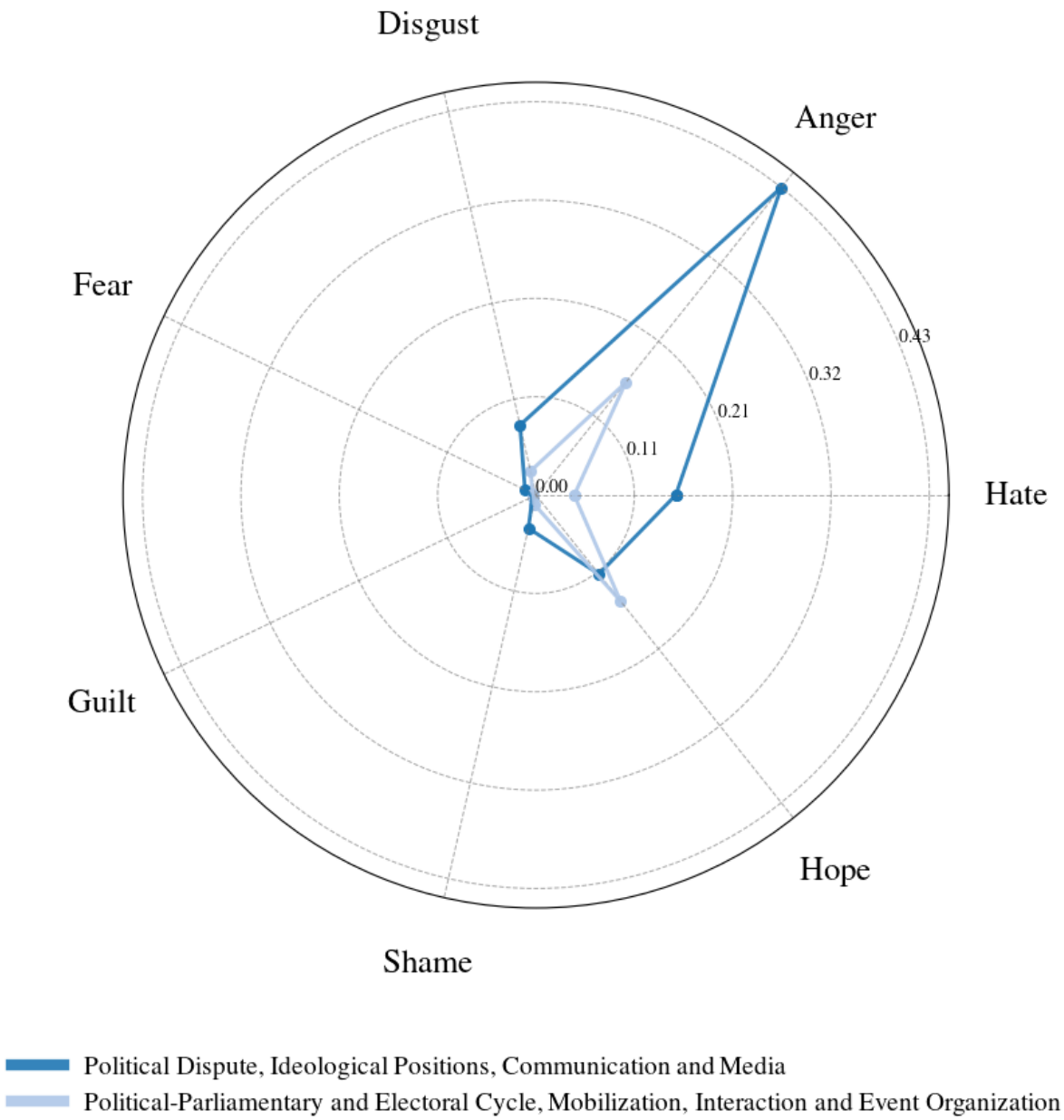


FIGURE D.34. Emotions probability distribution across *'Political-Electoral Dynamics, Public Management, Ideological Debates and Media Coverage'*.

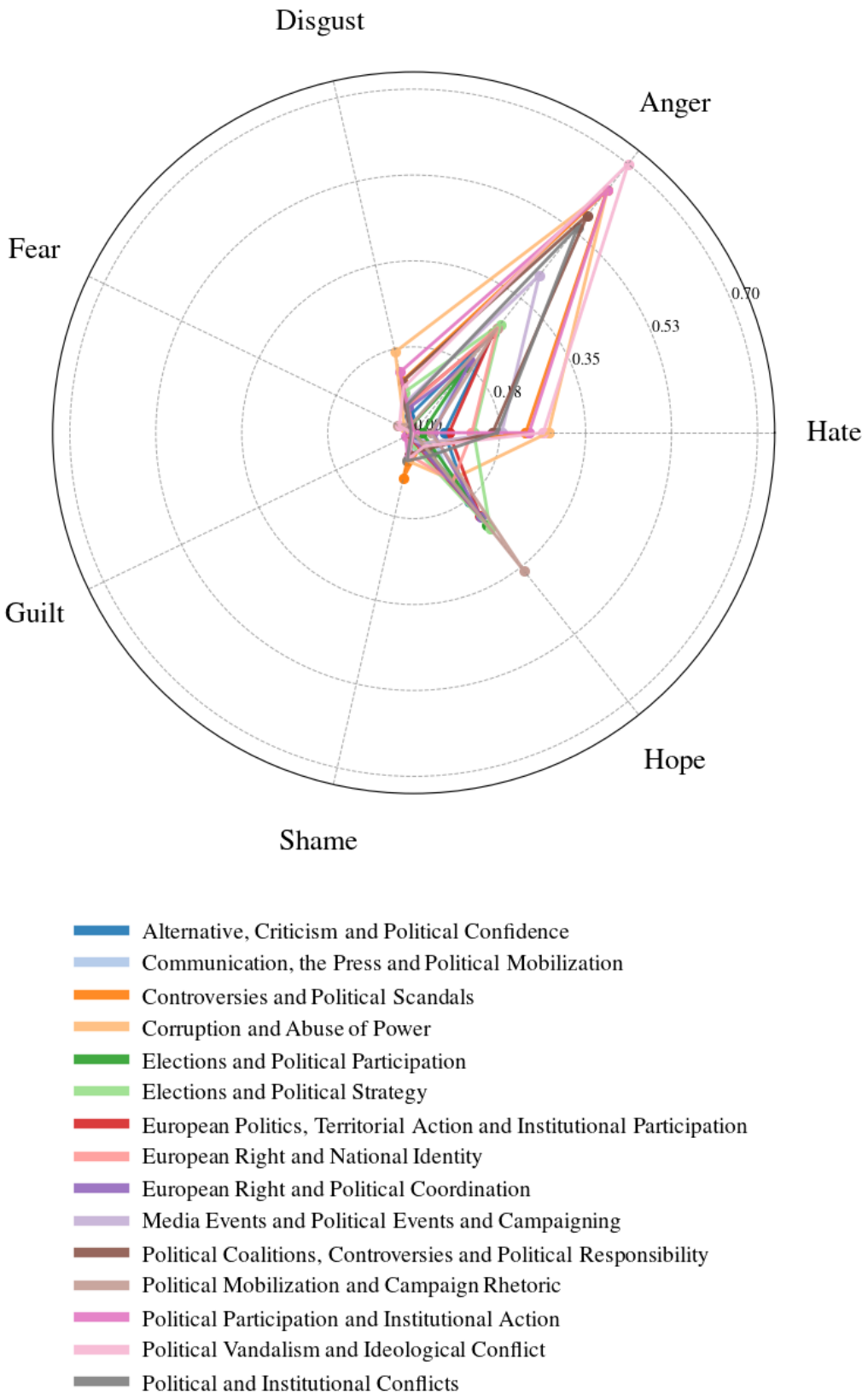


FIGURE D.35. Emotions probability distribution across *'Political Dispute, Ideological Positions, Communication and Media'*.

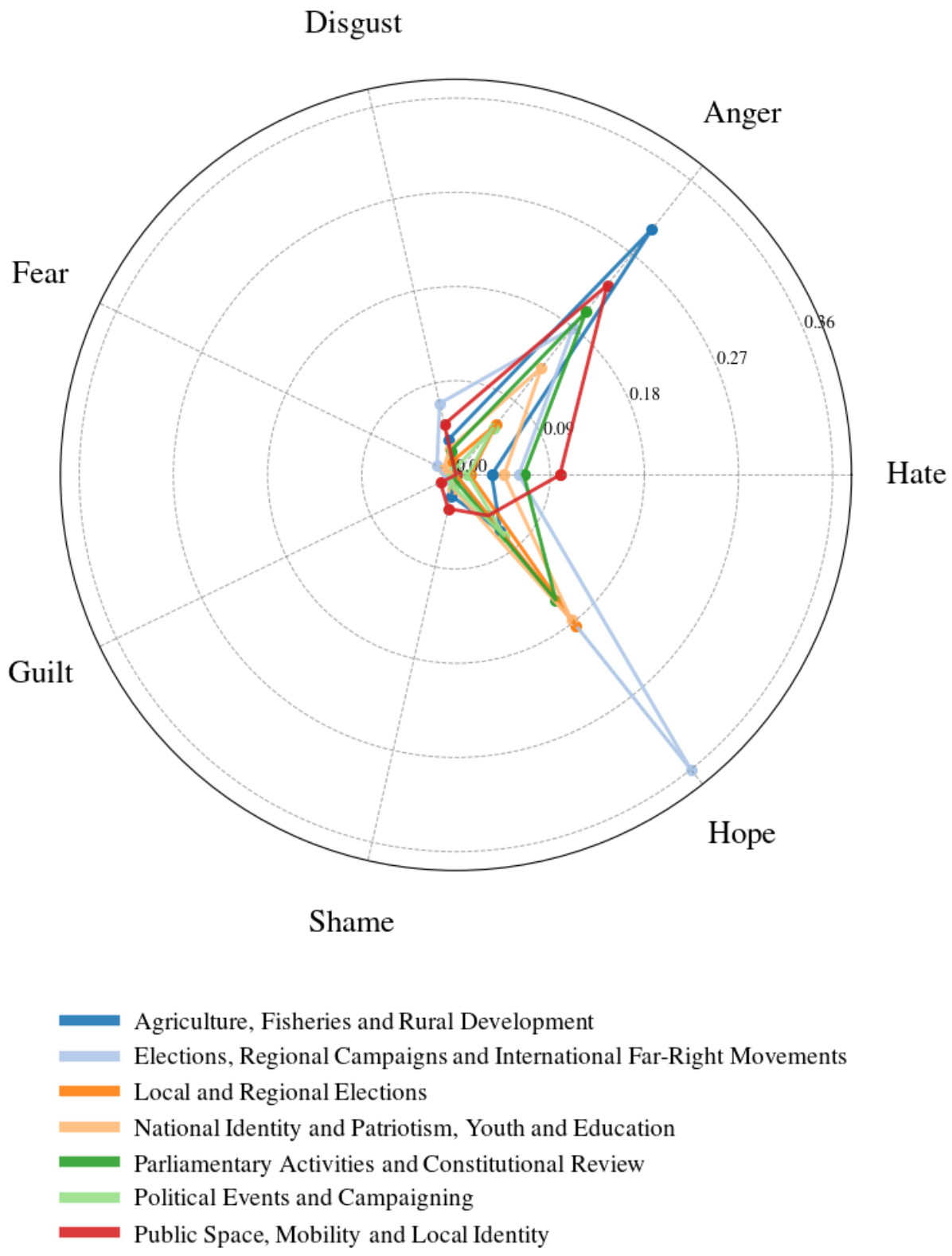


FIGURE D.36. Emotions probability distribution across *'Political-Parliamentary and Electoral Cycle, Mobilization, Interaction and Event Organization'*.

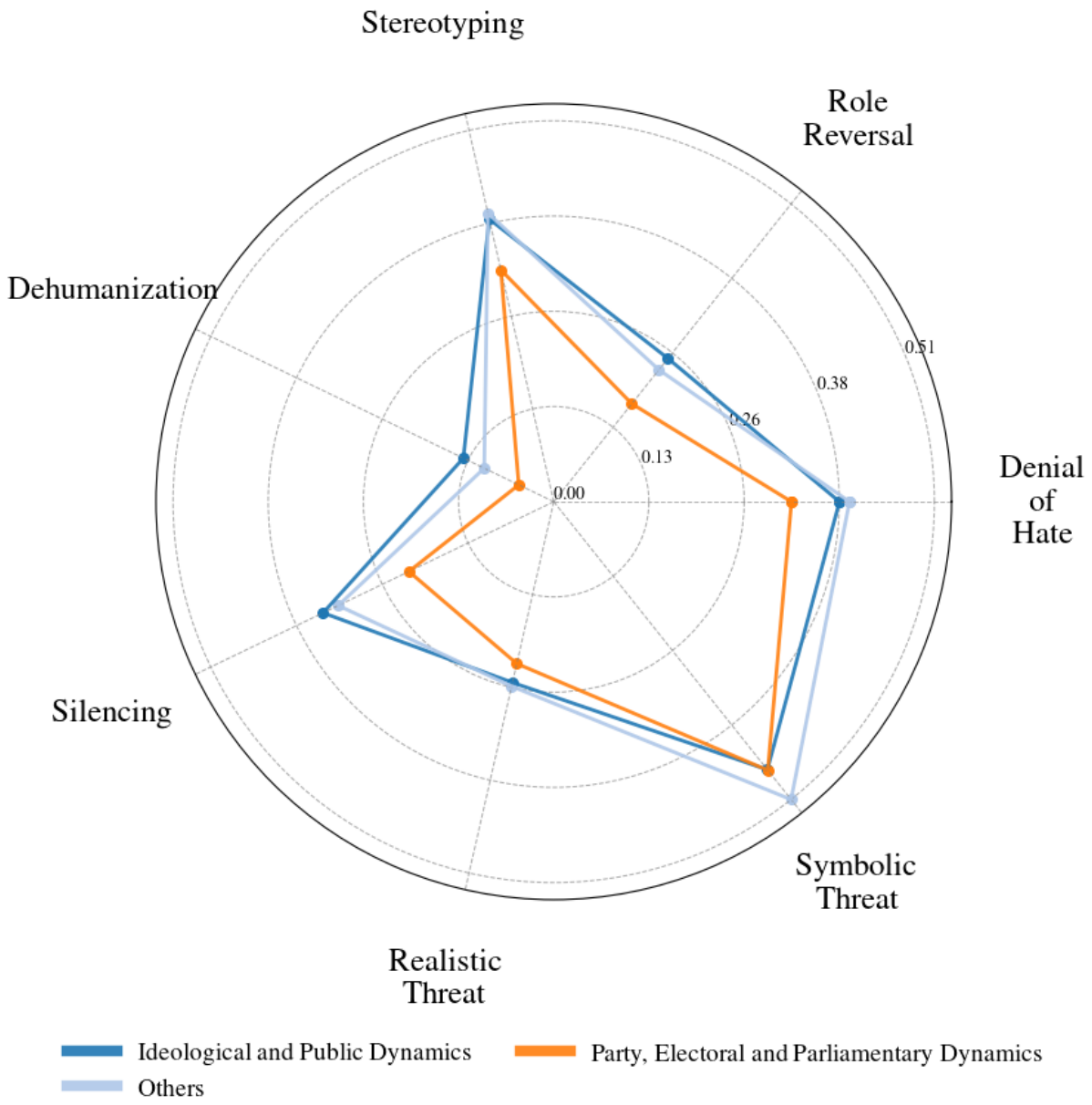


FIGURE D.37. HS strategies probability distribution across *'Political Discourse'* dimension.

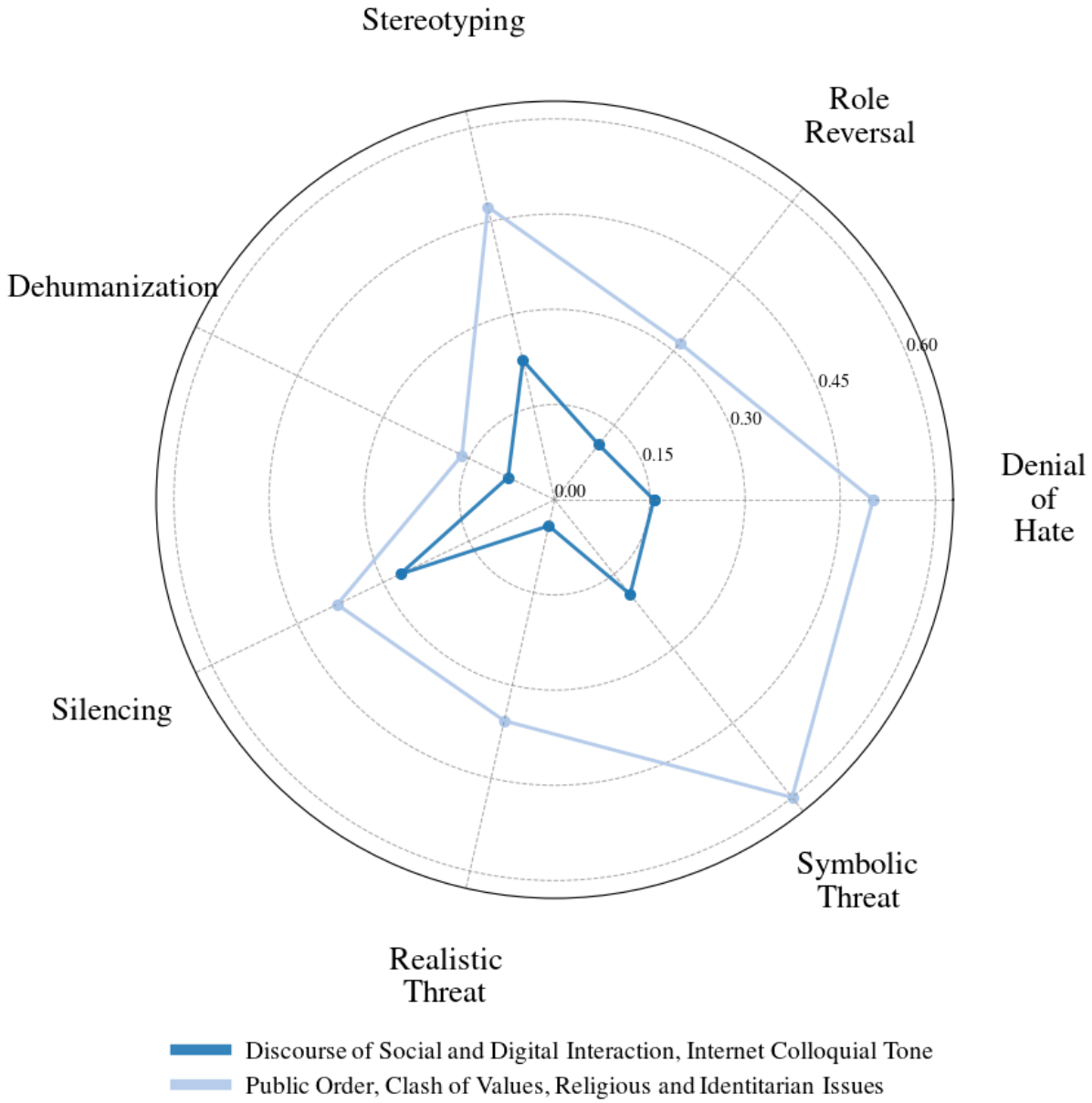


FIGURE D.38. HS strategies probability distribution across *'Ideological and Public Dynamics'* Dimension.

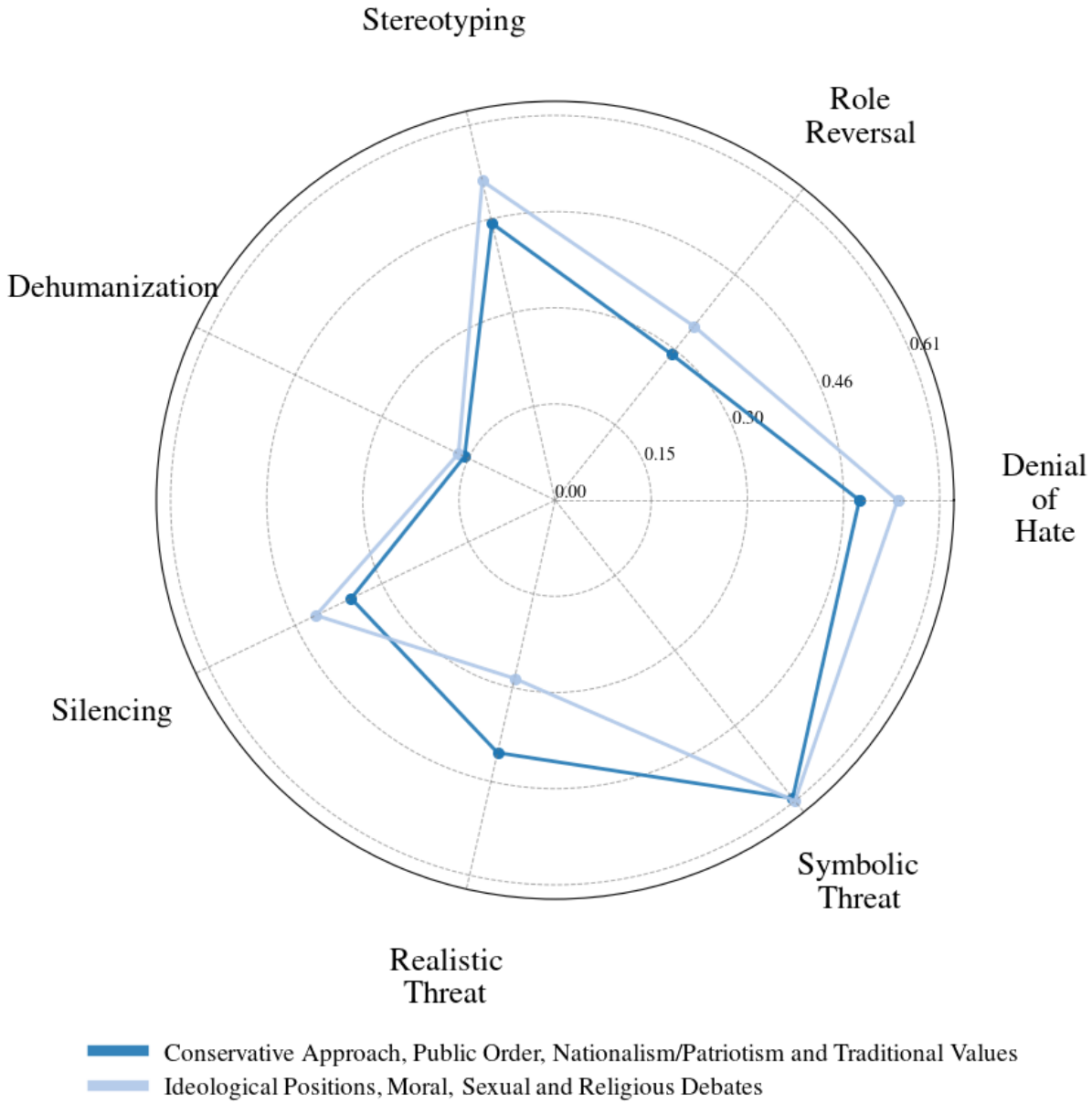


FIGURE D.39. HS strategies probability distribution across *'Public Order, Clash of Values, Religious and Identitarian Issues'*.

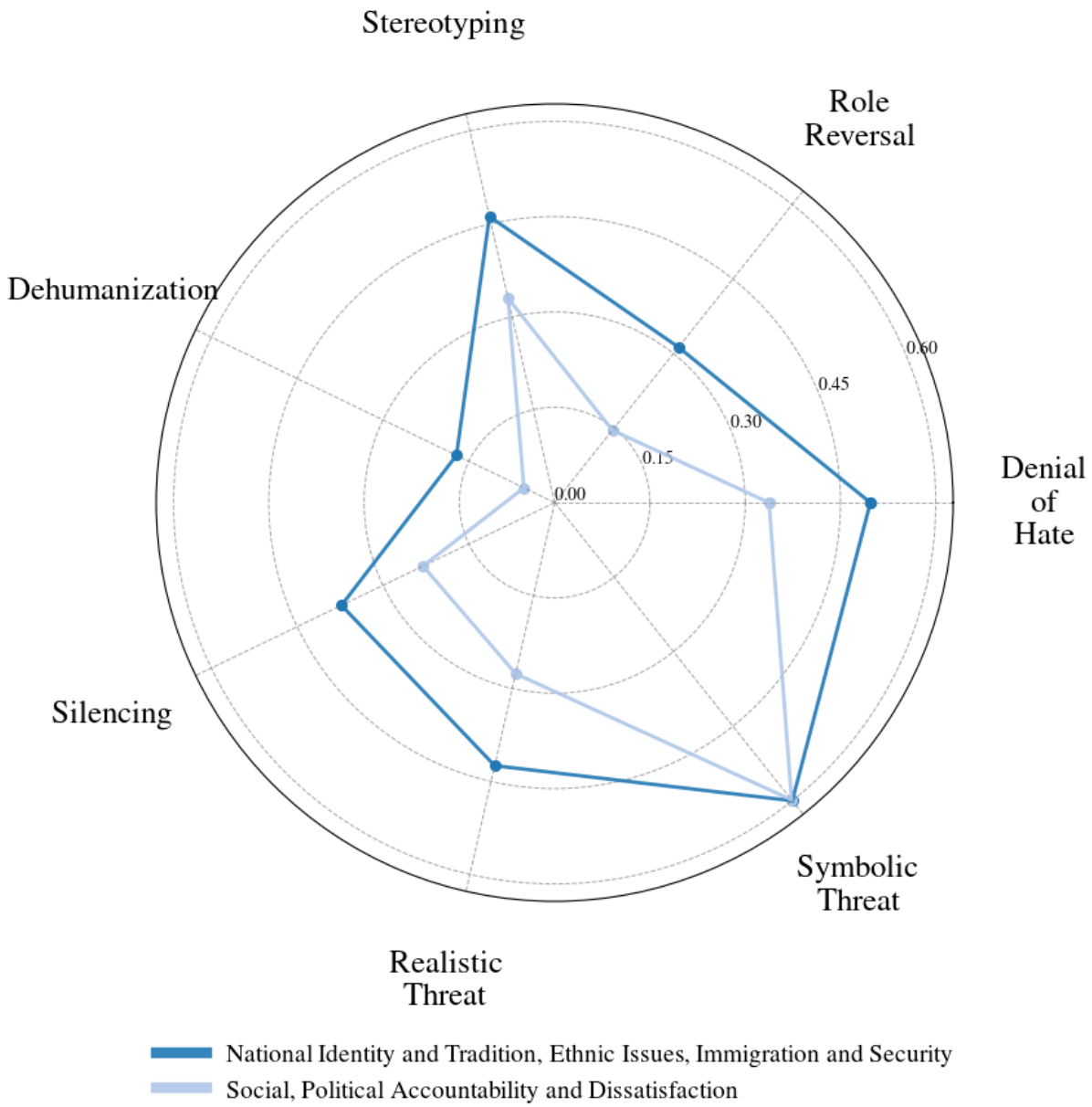


FIGURE D.40. HS strategies probability distribution across *'Conservative Approach, Public Order, Nationalism/Patriotism and Traditional Values'*.

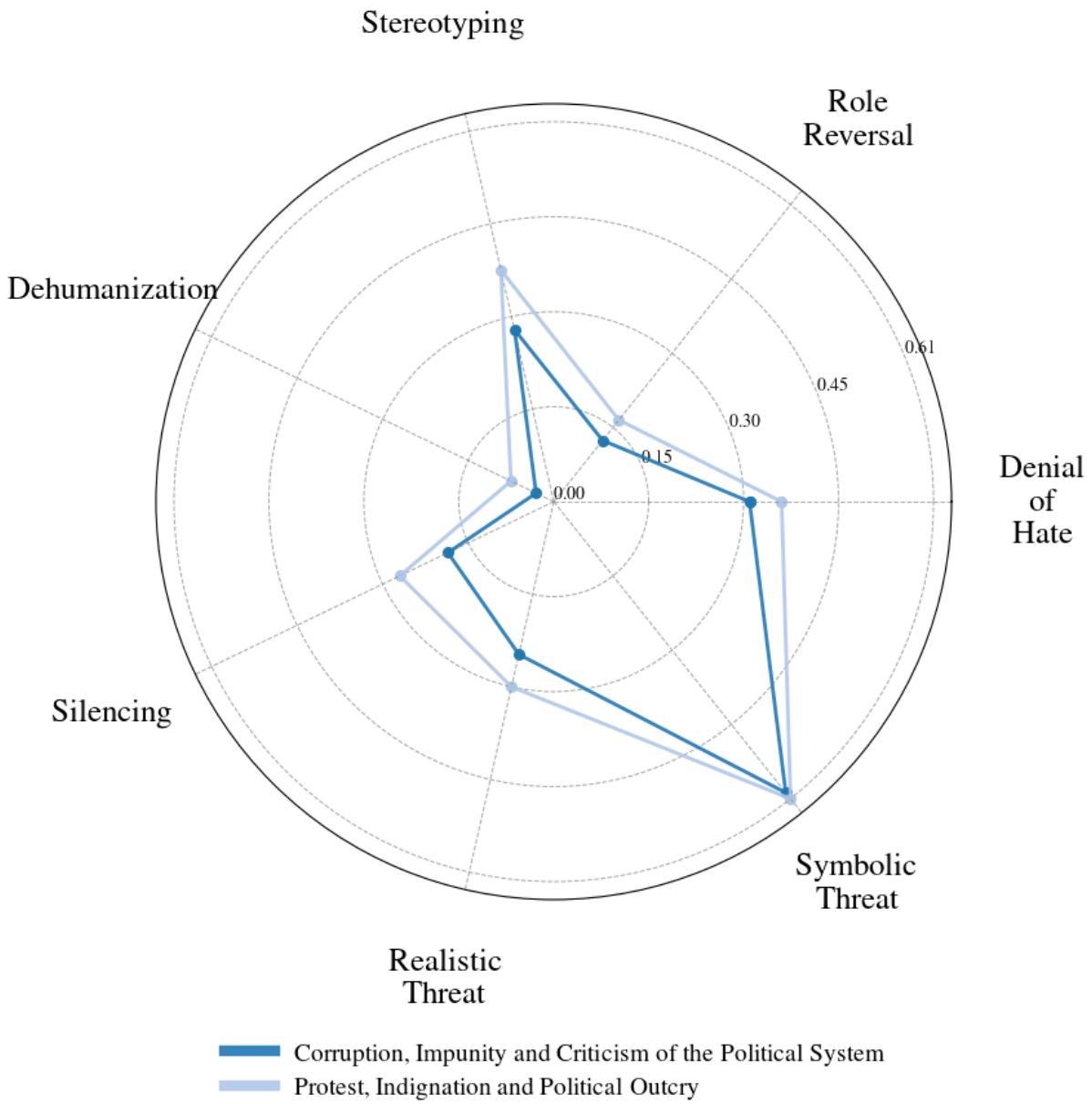


FIGURE D.41. HS strategies probability distribution across *'Social, Political Accountability and Dissatisfaction'*.

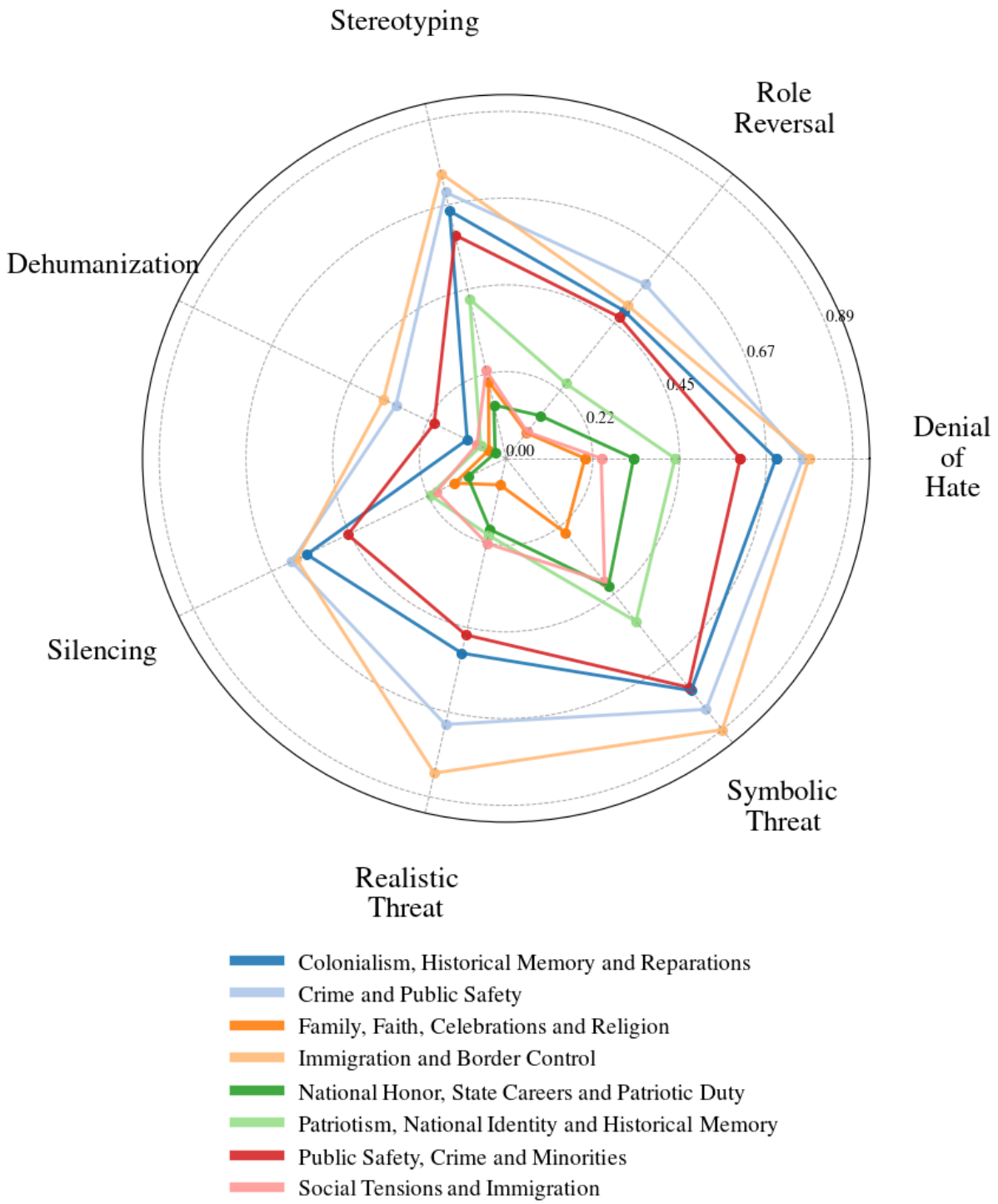


FIGURE D.42. HS strategies probability distribution across *'National Identity and Tradition, Ethnic Issues, Immigration and Security'*.

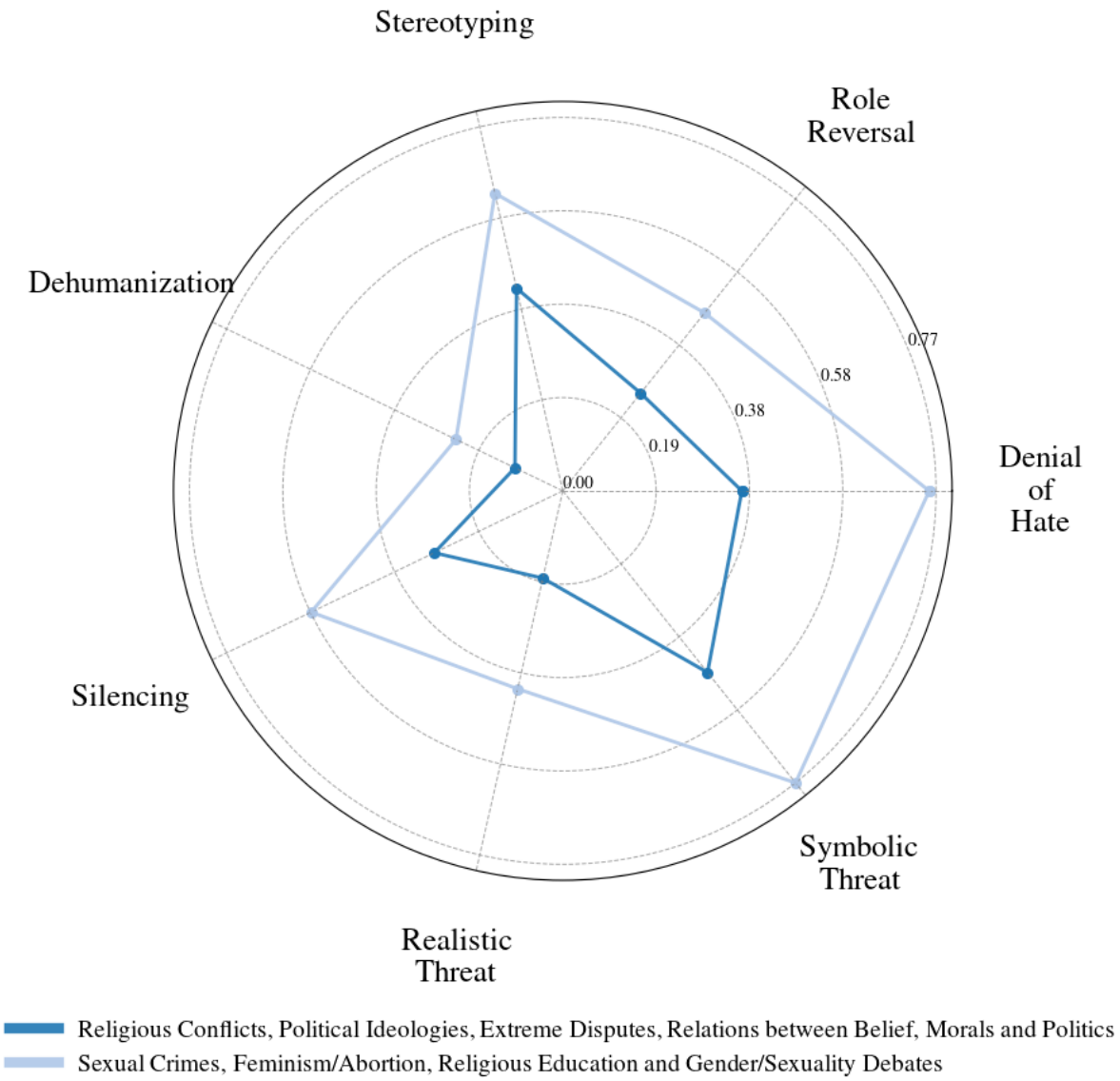


FIGURE D.43. HS strategies probability distribution across *'Ideological Positions, Moral, Sexual and Religious Debates'*.

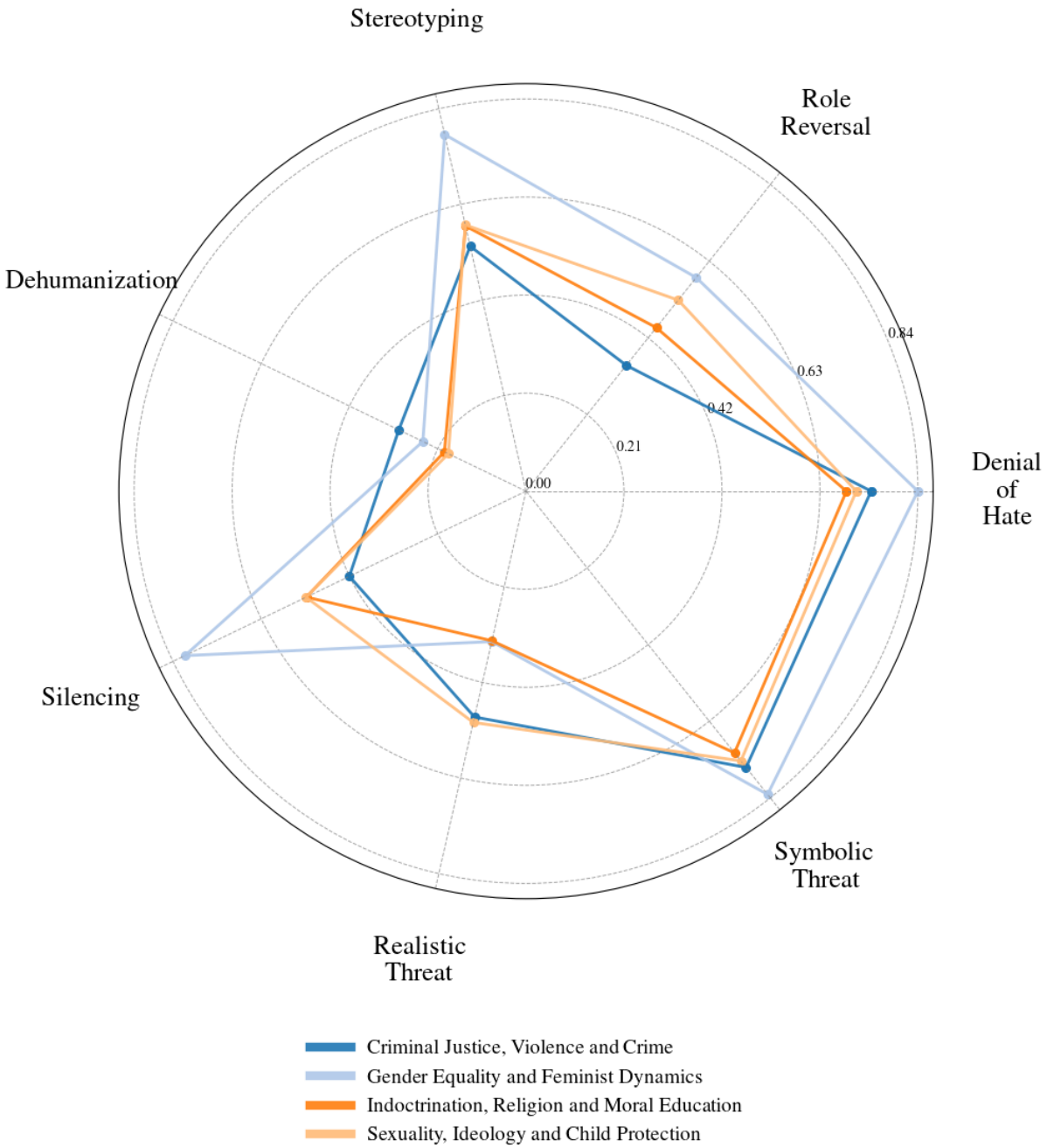


FIGURE D.44. HS strategies probability distribution across *'Sexual Crimes, Feminism/Abortion, Religious Education and Gender/Sexuality Debates'*.

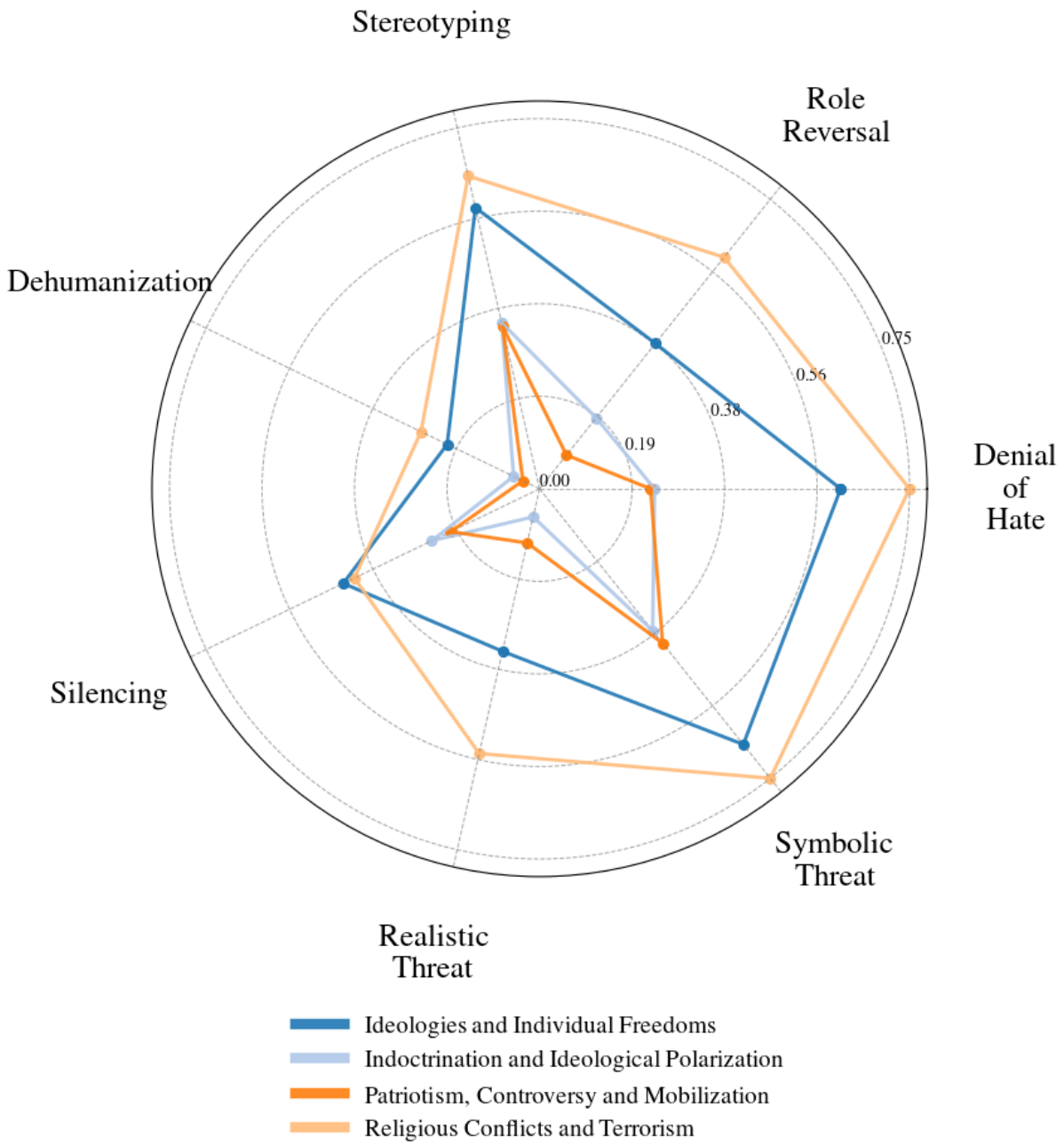


FIGURE D.45. HS strategies probability distribution across *'Religious Conflicts, Political Ideologies, Extreme Disputes, and Relations between Belief, Morals and Politics'*.

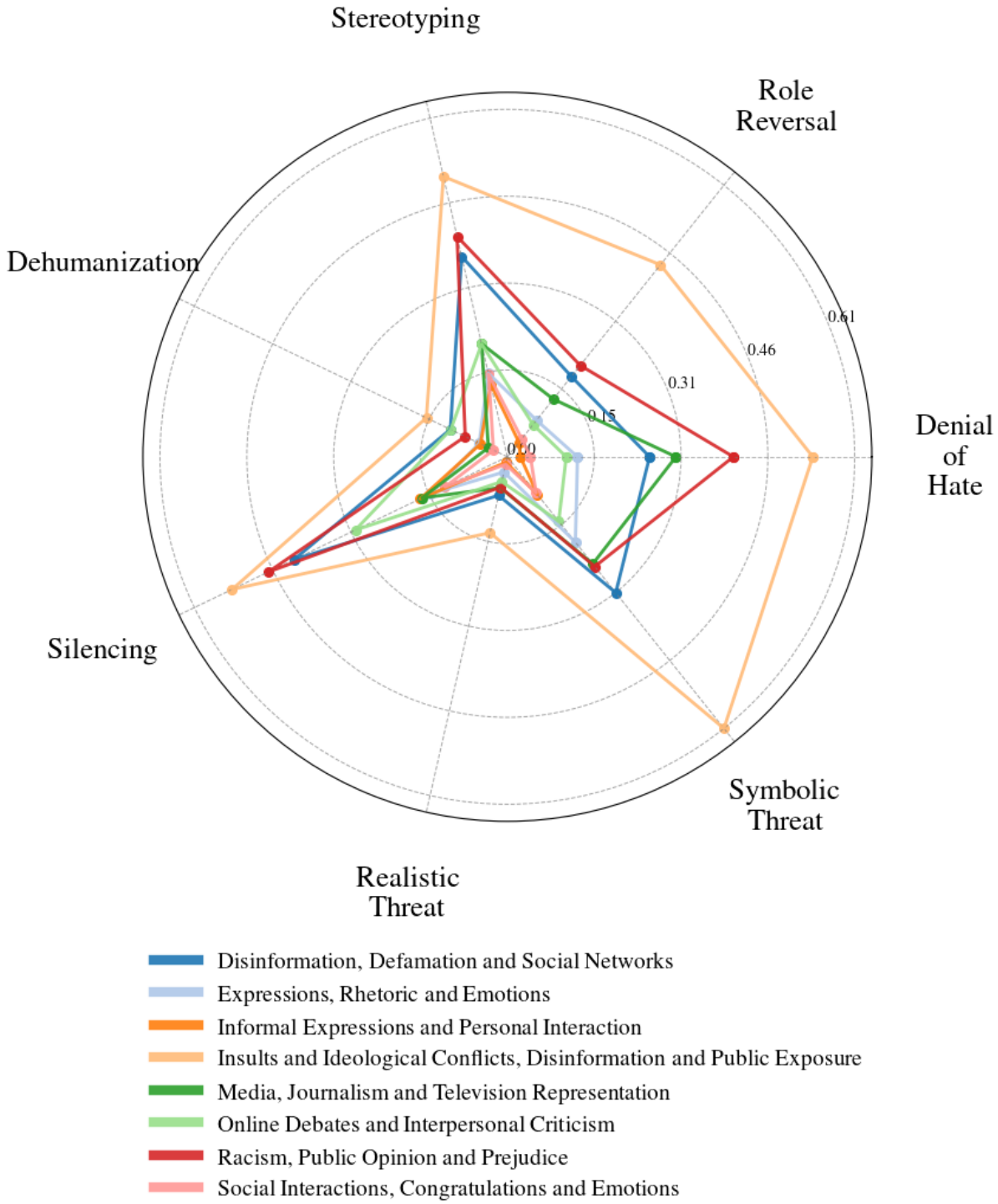


FIGURE D.46. HS strategies probability distribution across *'Discourse of Social and Digital Interaction, Internet Colloquial Tone'*.

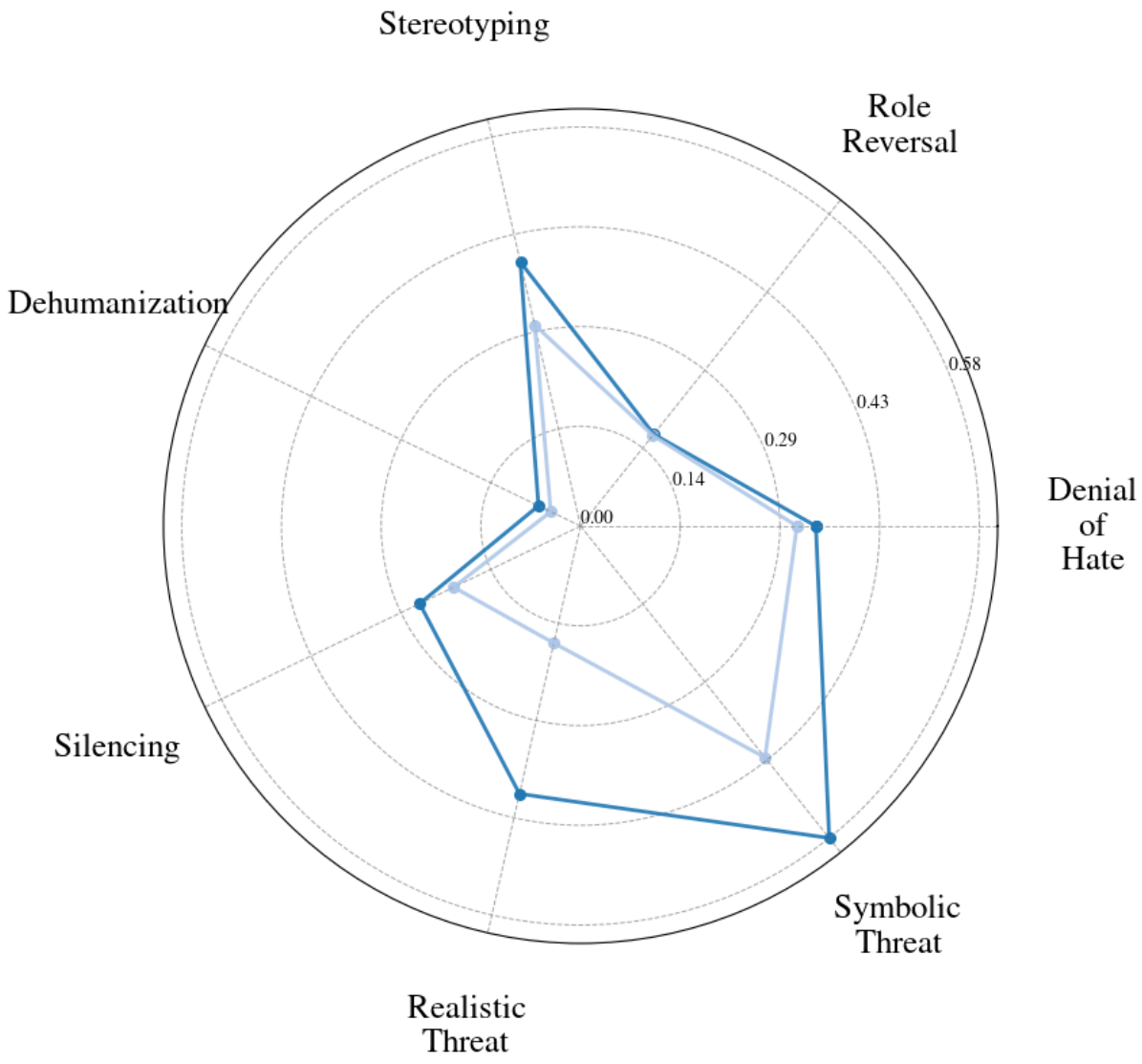


FIGURE D.47. HS strategies probability distribution across *'Party, Electoral and Parliamentary Dynamics'* Dimension.

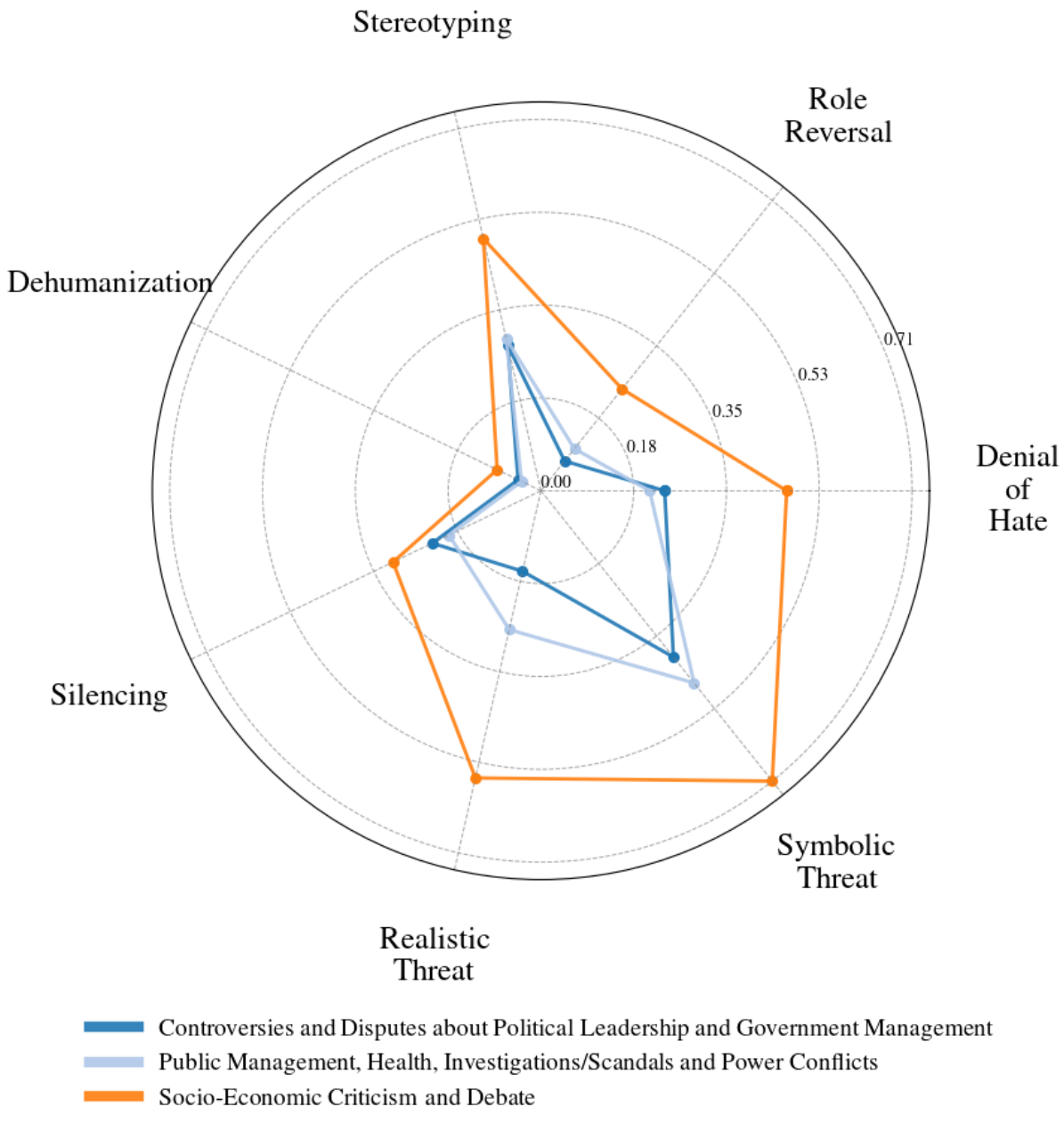


FIGURE D.48. HS strategies probability distribution across *'Governance, Economy and Social Policies, Public Health'*.

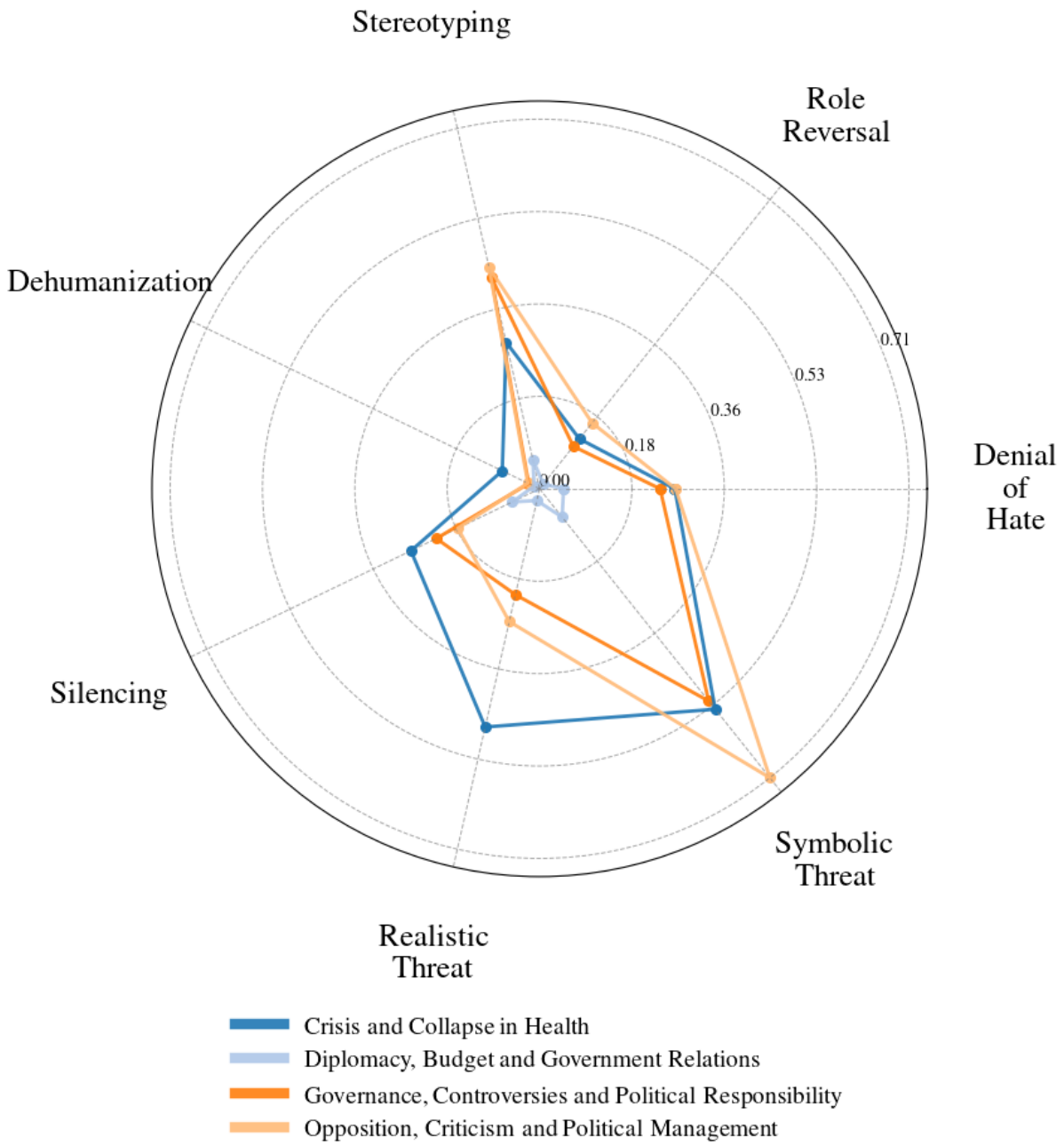


FIGURE D.49. HS strategies probability distribution across *'Public Management, Health, Investigations/Scandals and Power Conflicts'*.

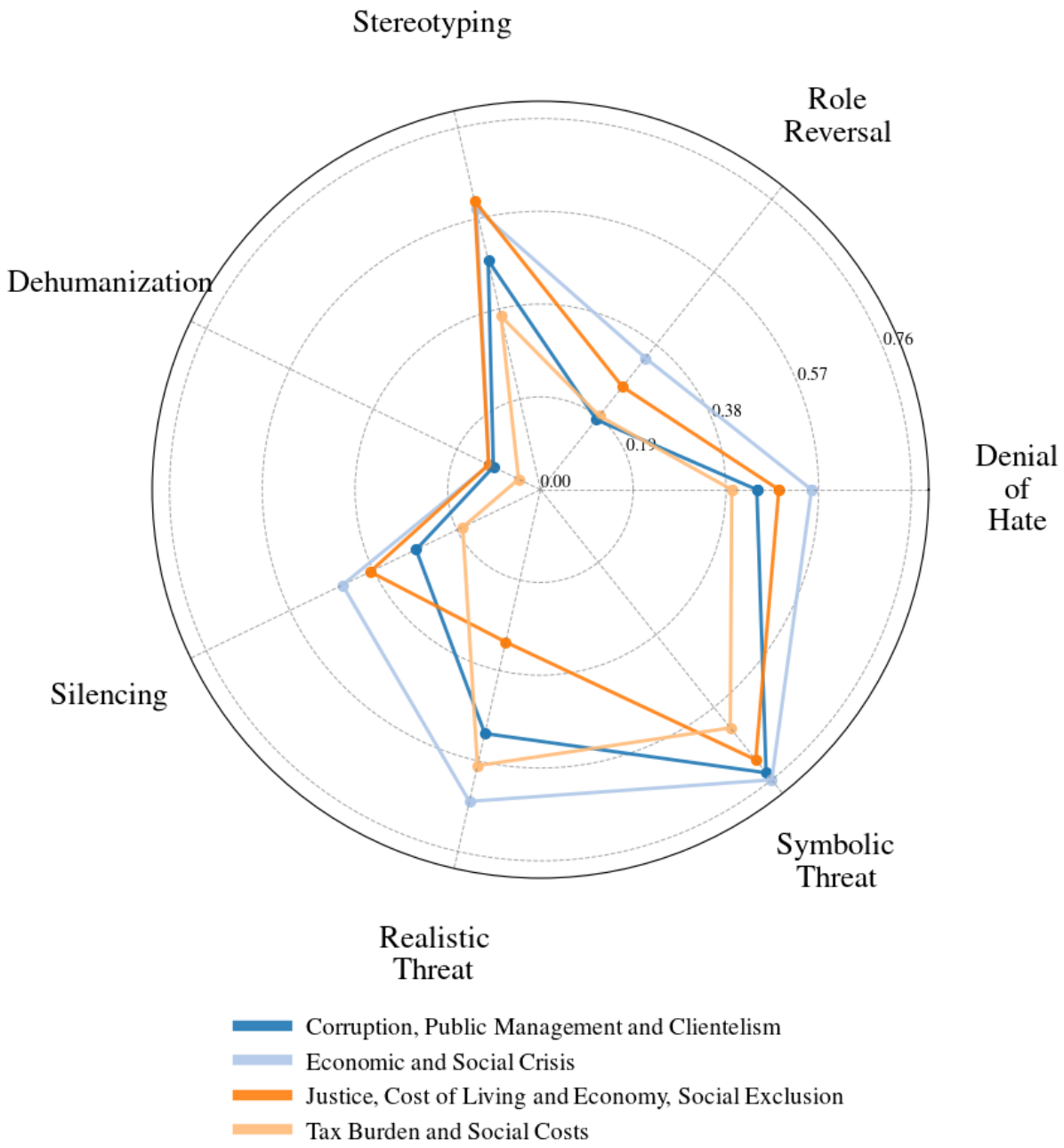


FIGURE D.50. HS strategies probability distribution across *'Socio-Economic Criticism and Debate'*.

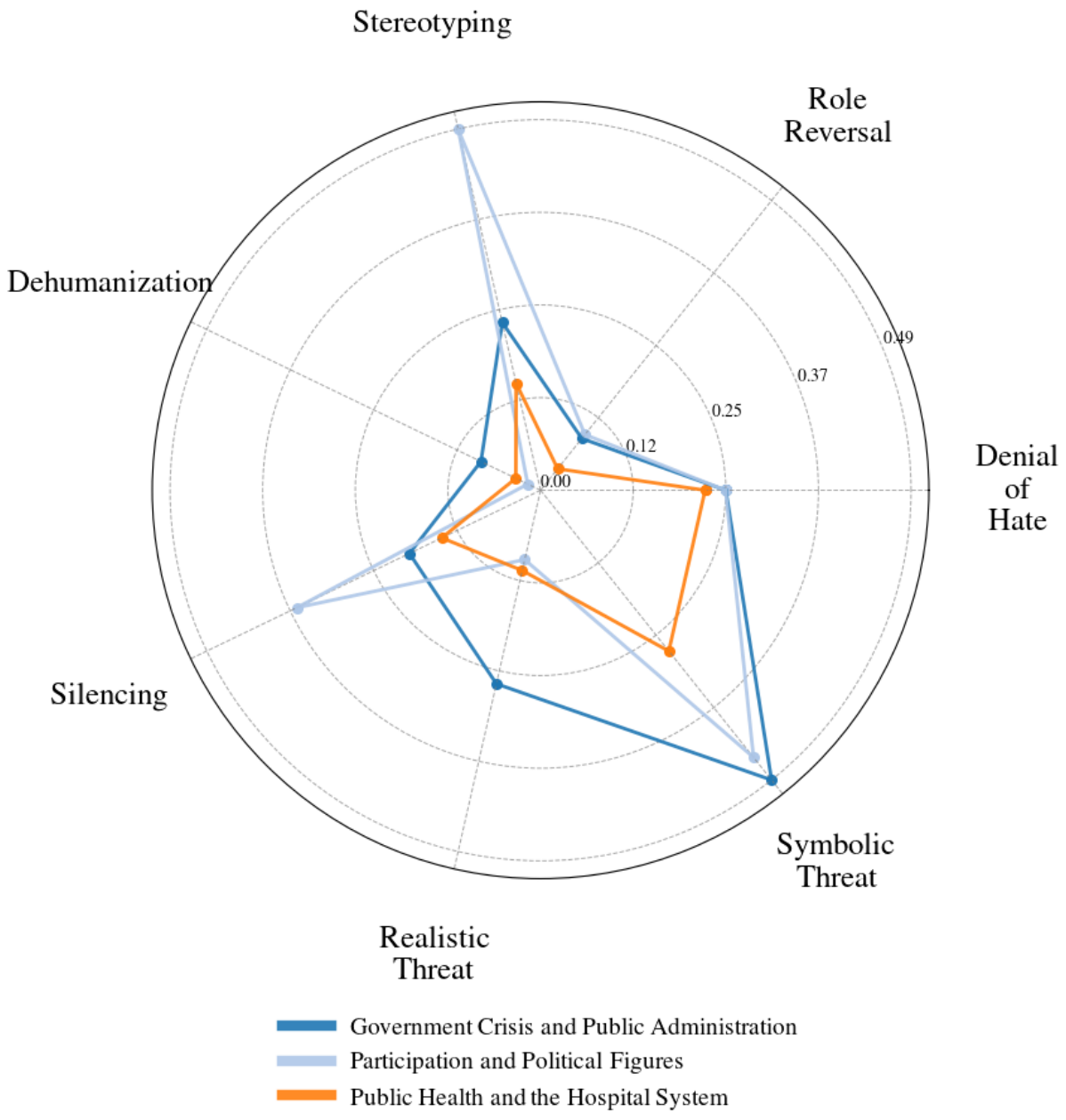


FIGURE D.51. HS strategies probability distribution across *'Controversies and Disputes about Political Leadership and Government Management'*.

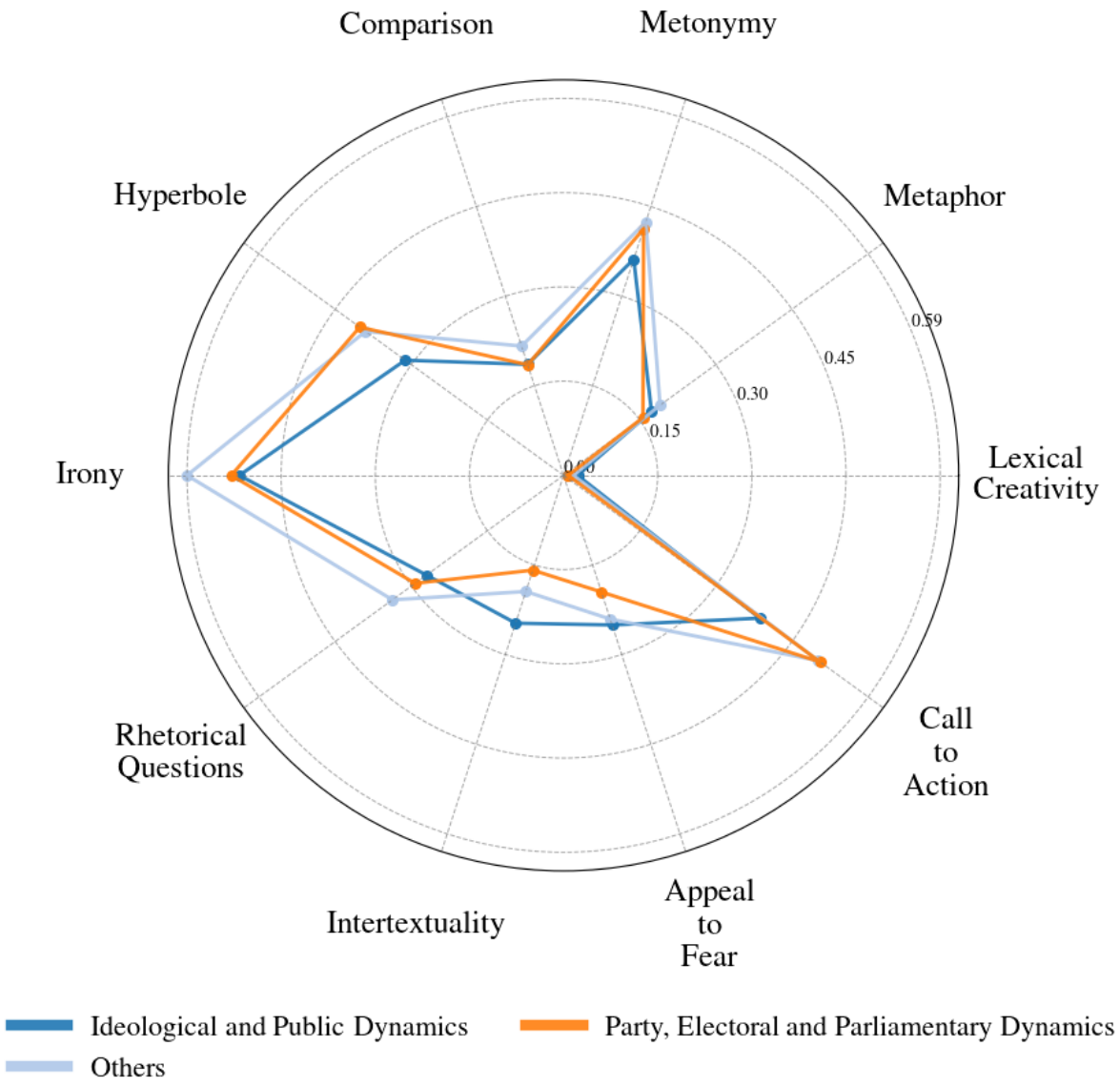


FIGURE D.52. Rhetorical devices probability distribution across *'Political Discourse'* dimension.

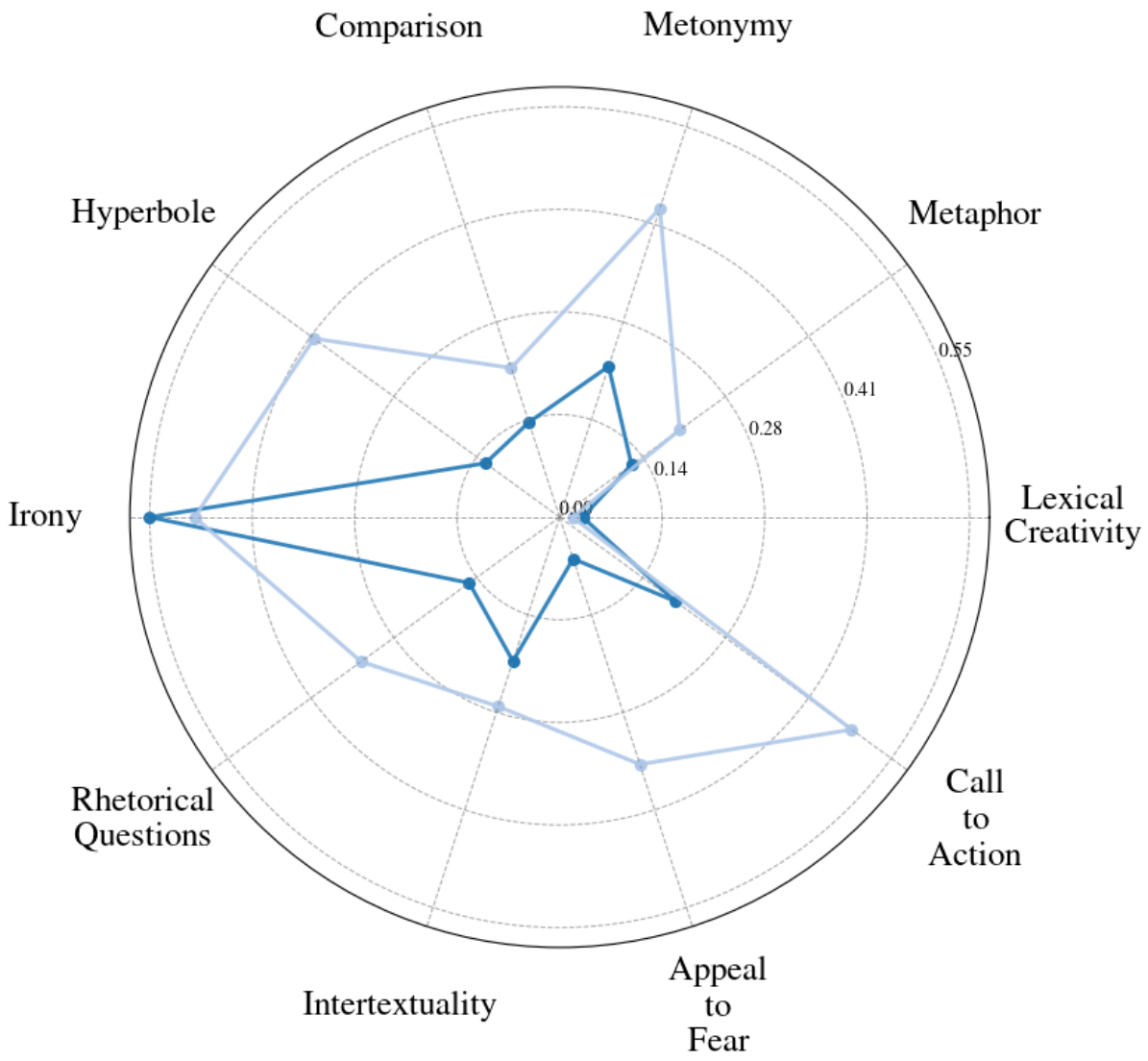


FIGURE D.53. Rhetorical devices probability distribution across *'Ideological and Public Dynamics'* Dimension.

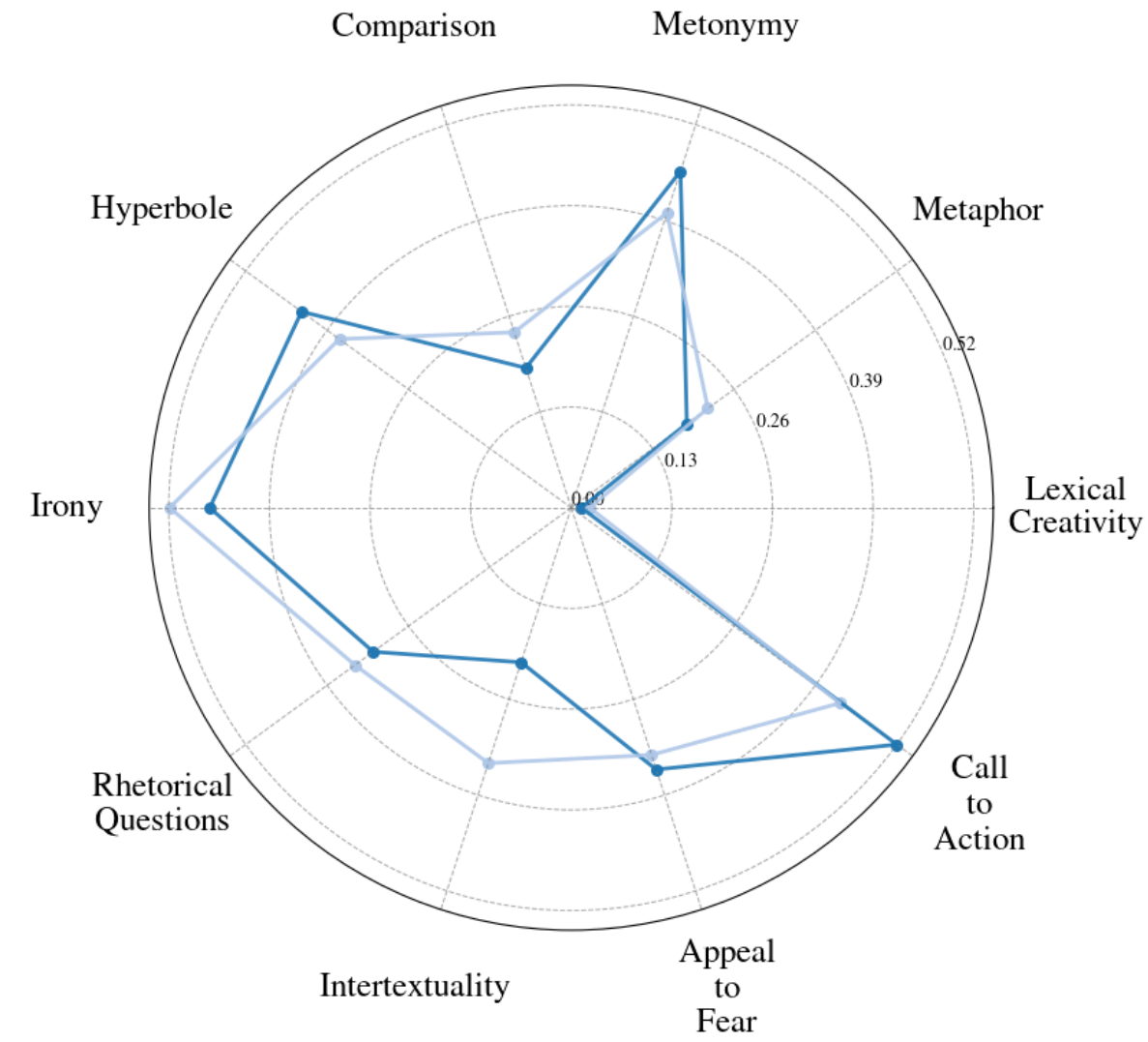


FIGURE D.54. Rhetorical devices probability distribution across *'Public Order, Clash of Values, Religious and Identitarian Issues'*.

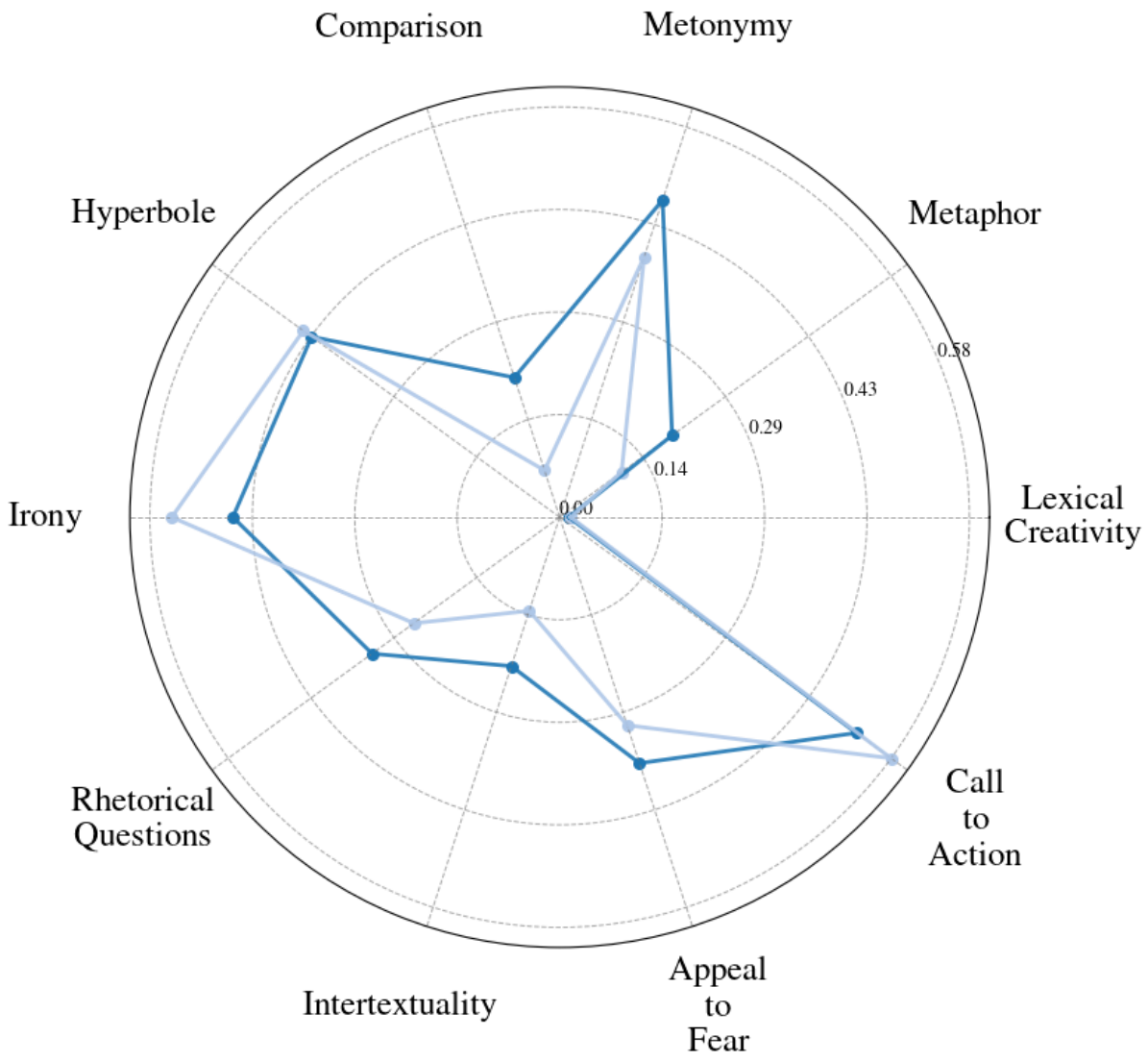


FIGURE D.55. Rhetorical devices probability distribution across *'Conservative Approach, Public Order, Nationalism/Patriotism and Traditional Values'*.

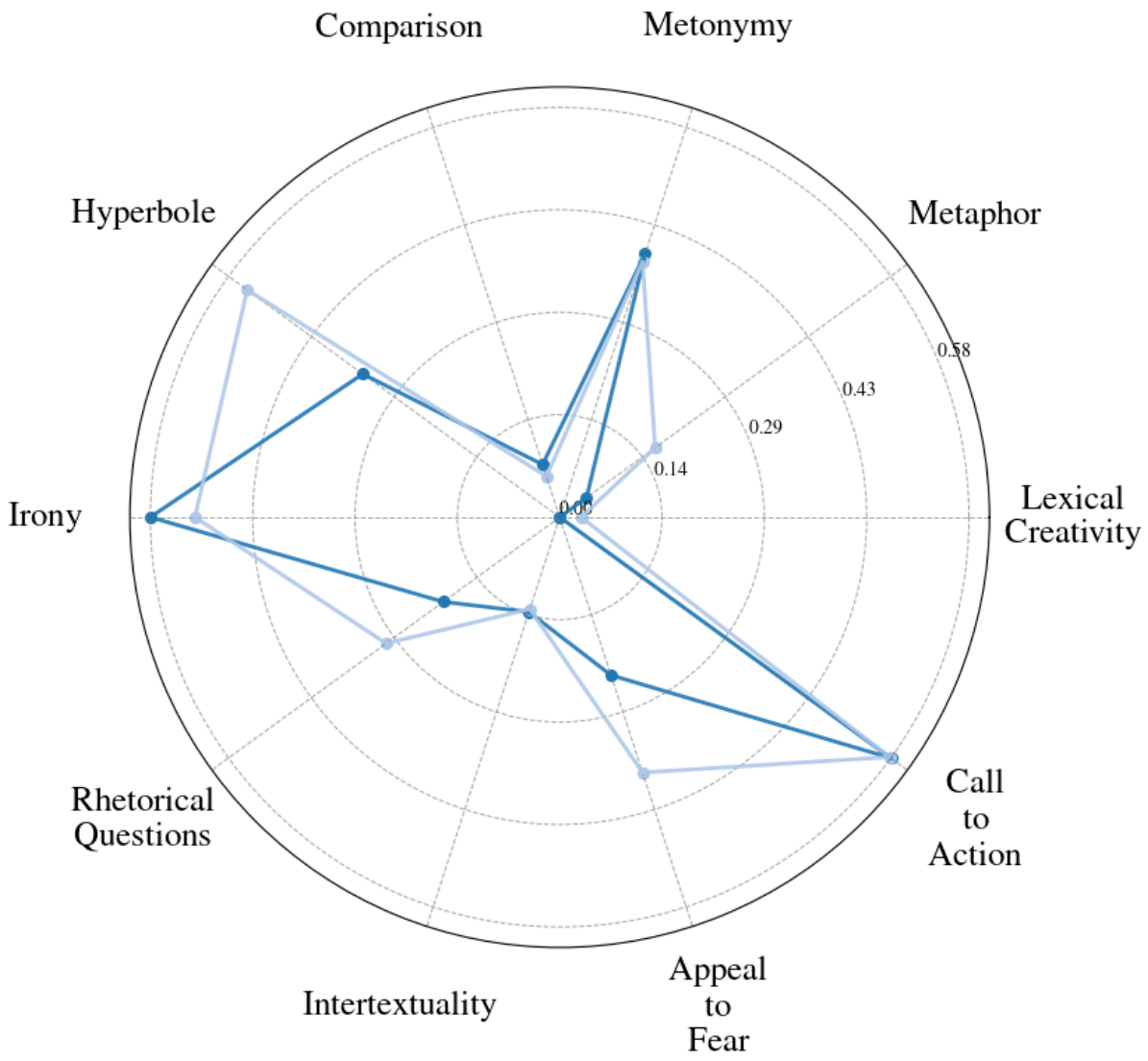


FIGURE D.56. Rhetorical devices probability distribution across *'Social, Political Accountability and Dissatisfaction'*.

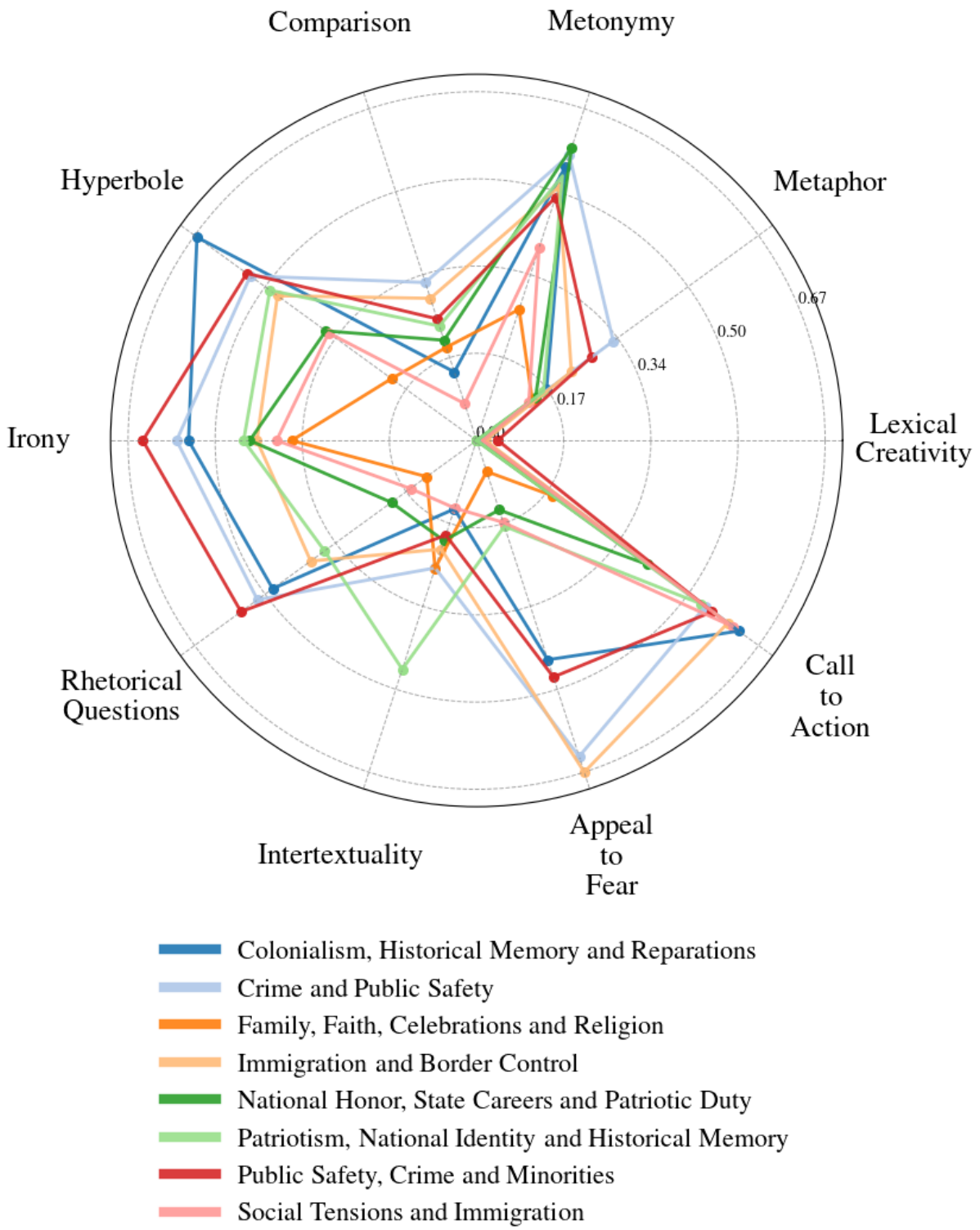


FIGURE D.57. Rhetorical devices probability distribution across *'National Identity and Tradition, Ethnic Issues, Immigration and Security'*.

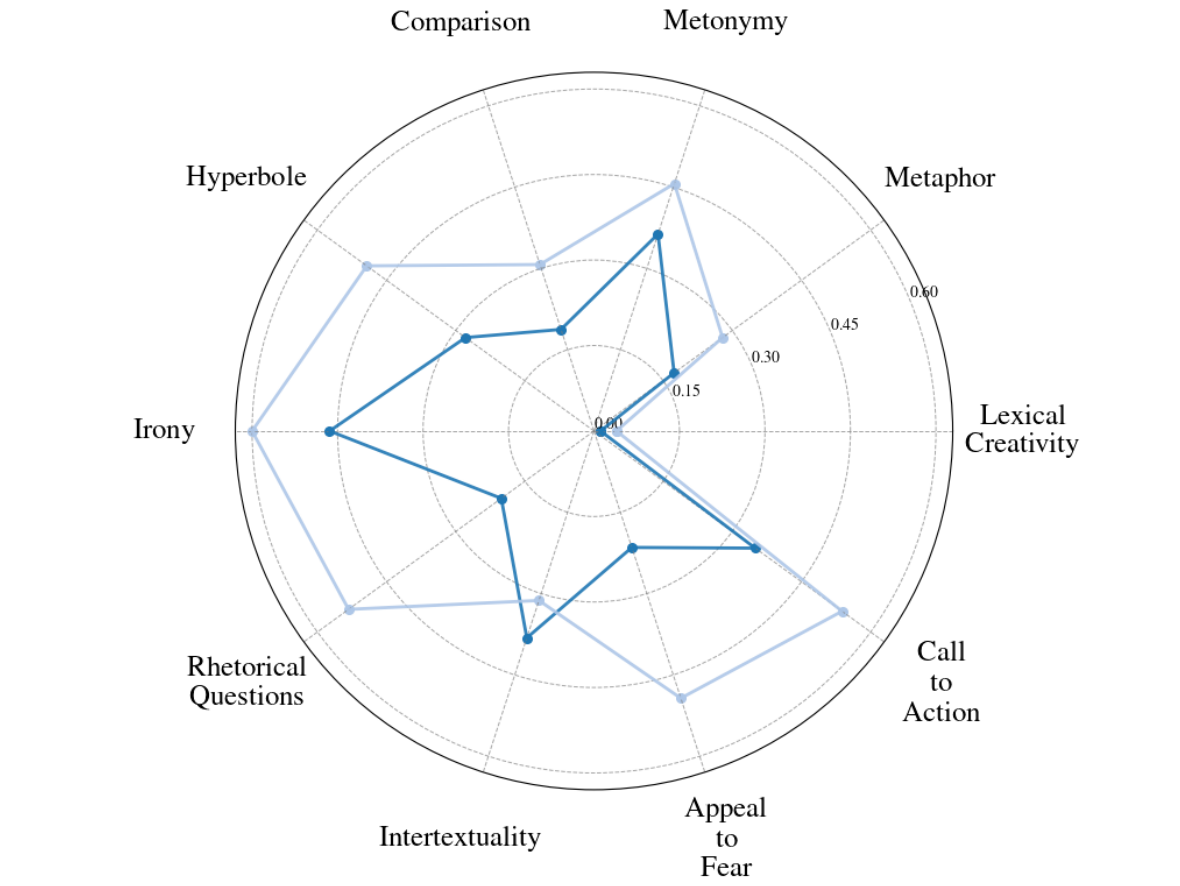


FIGURE D.58. Rhetorical devices probability distribution across *'Ideological Positions, Moral, Sexual and Religious Debates'*.

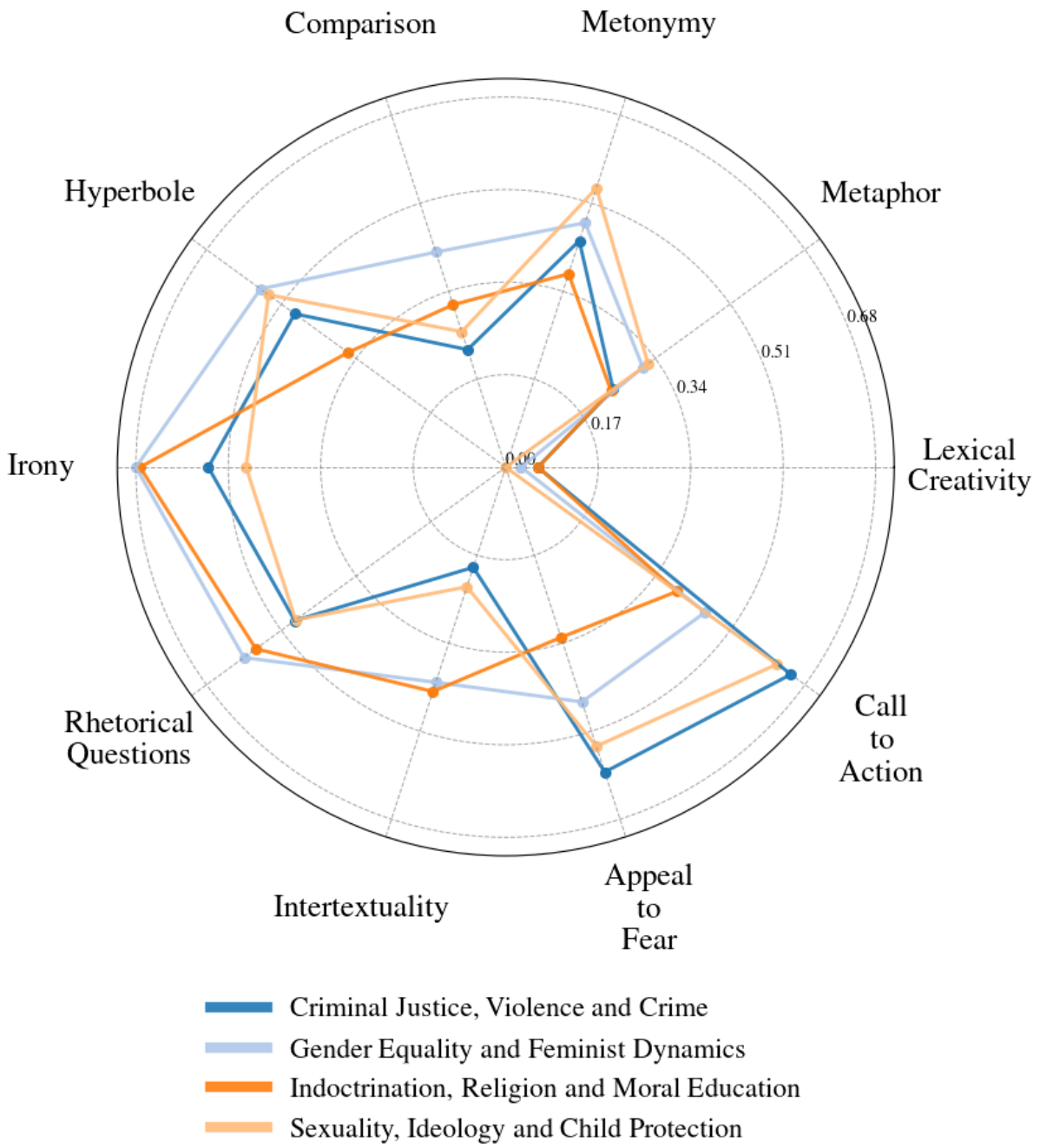


FIGURE D.59. Rhetorical devices probability distribution across *'Sexual Crimes, Feminism/Abortion, Religious Education and Gender/Sexuality Debates'*.

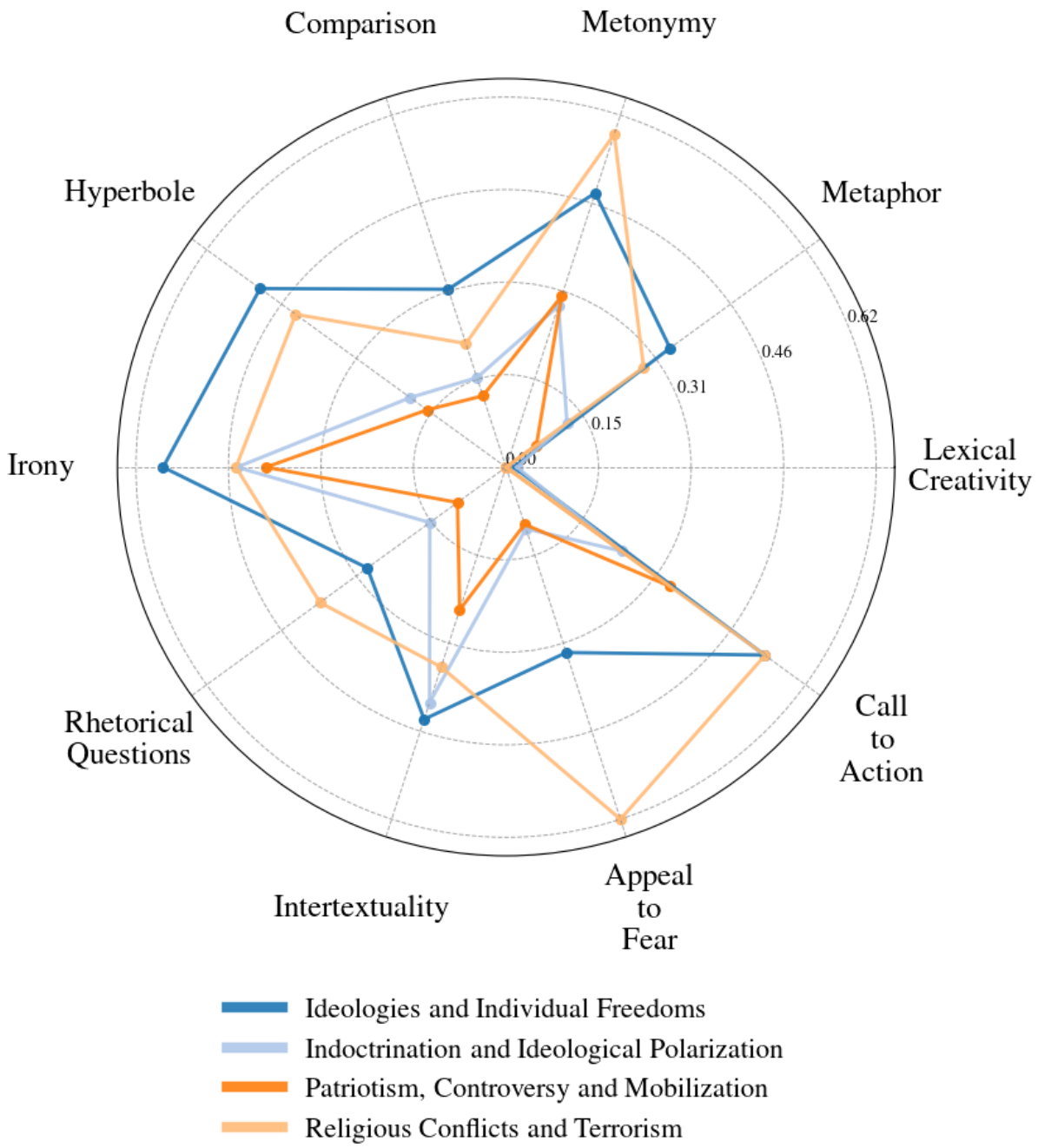


FIGURE D.60. Rhetorical devices probability distribution across *'Religious Conflicts, Political Ideologies, Extreme Disputes, and Relations between Belief, Morals and Politics'*.

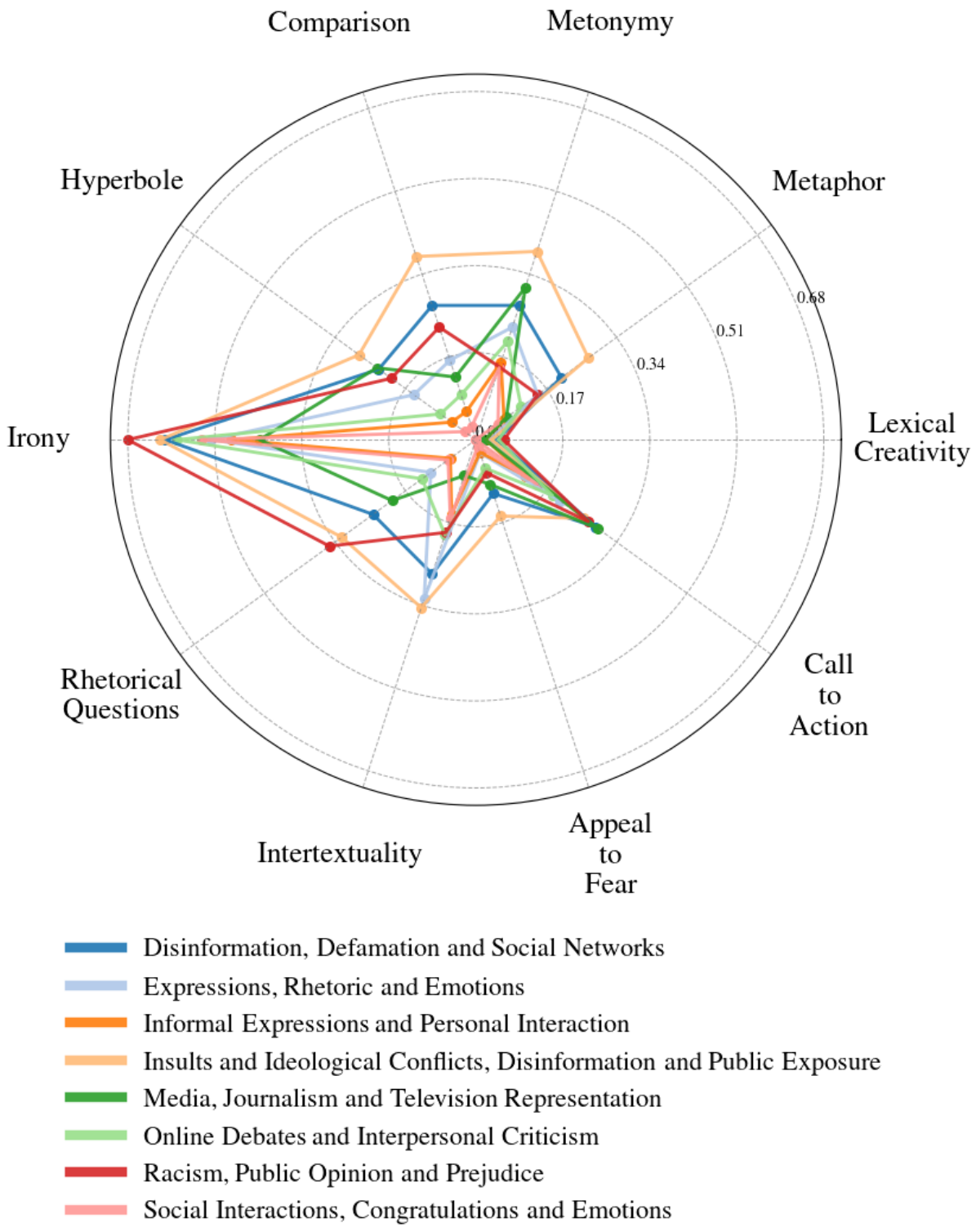


FIGURE D.61. Rhetorical devices probability distribution across *'Discourse of Social and Digital Interaction, Internet Colloquial Tone'*.

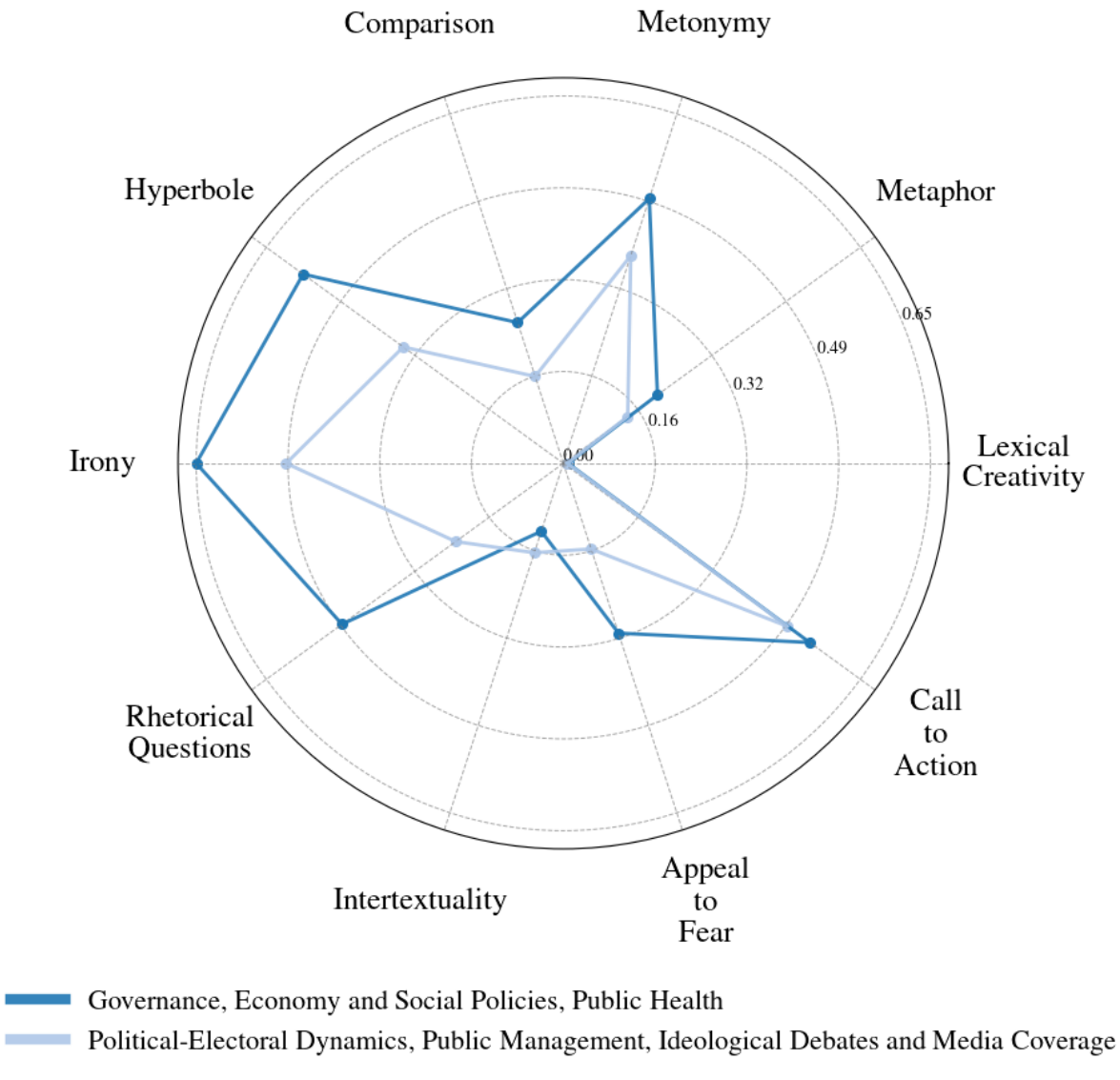


FIGURE D.62. Rhetorical devices probability distribution across *'Party, Electoral and Parliamentary Dynamics'* Dimension.

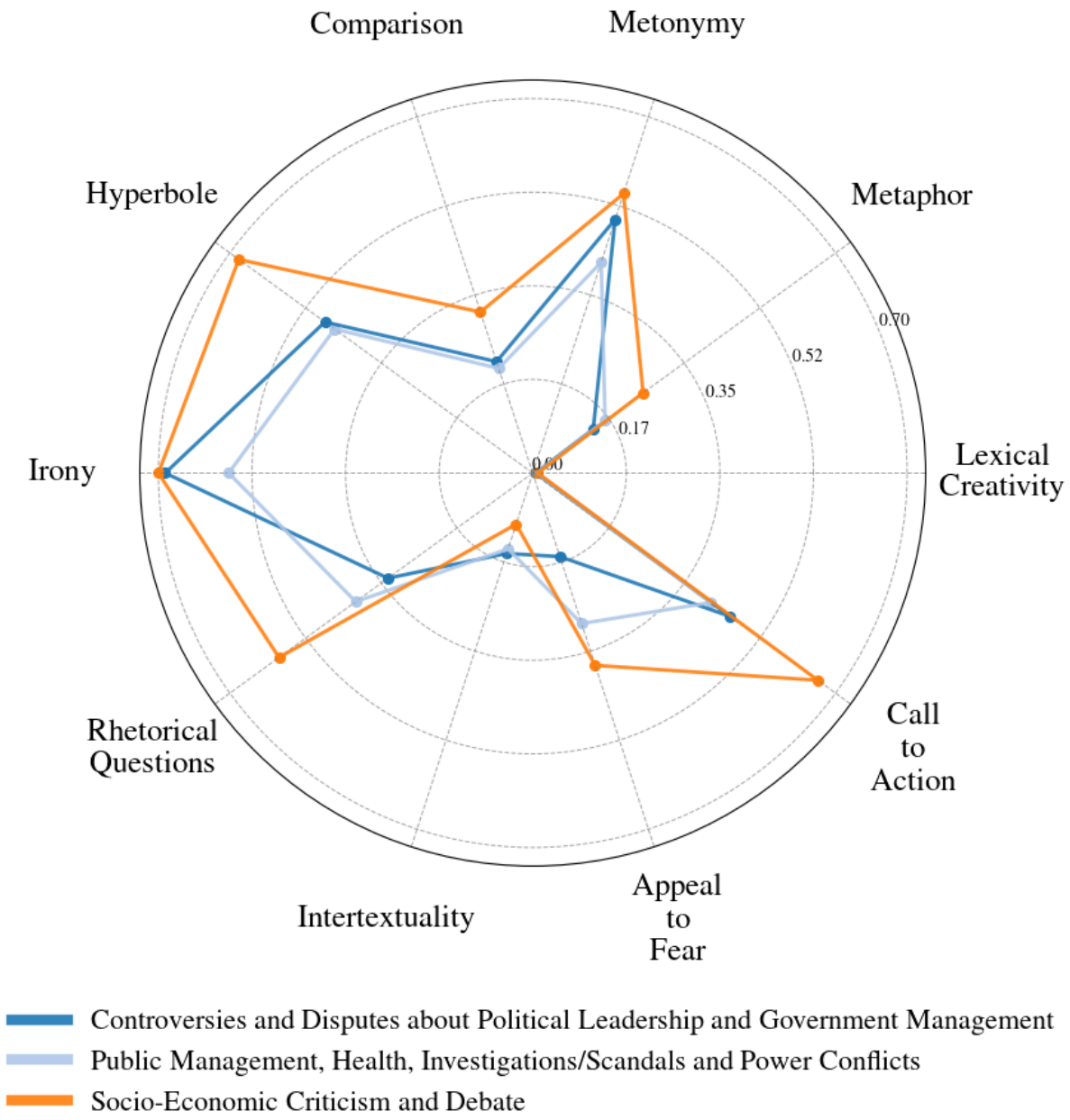


FIGURE D.63. Rhetorical devices probability distribution across *'Governance, Economy and Social Policies, Public Health'*.

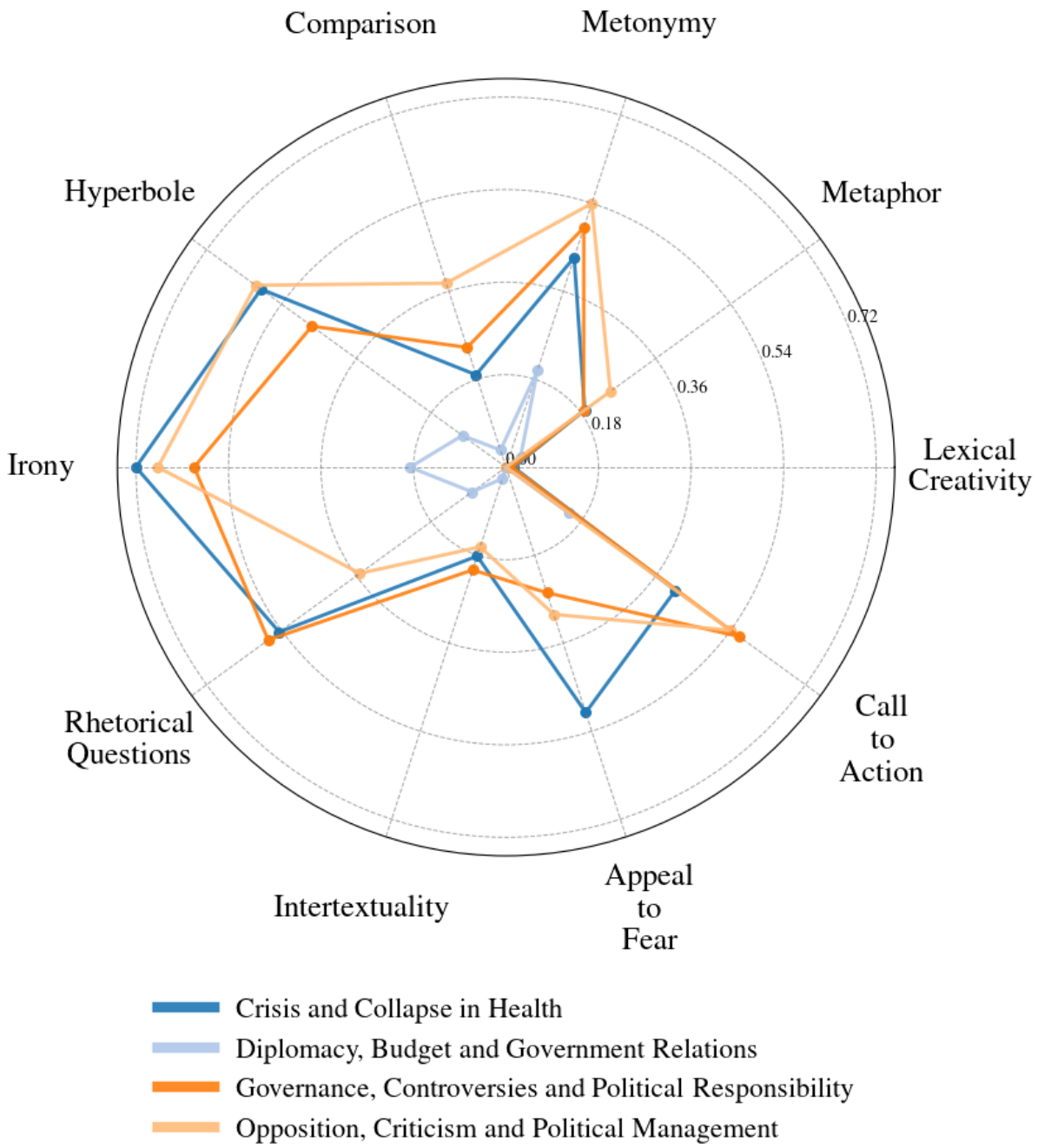


FIGURE D.64. Rhetorical devices probability distribution across *'Public Management, Health, Investigations/Scandals and Power Conflicts'*.

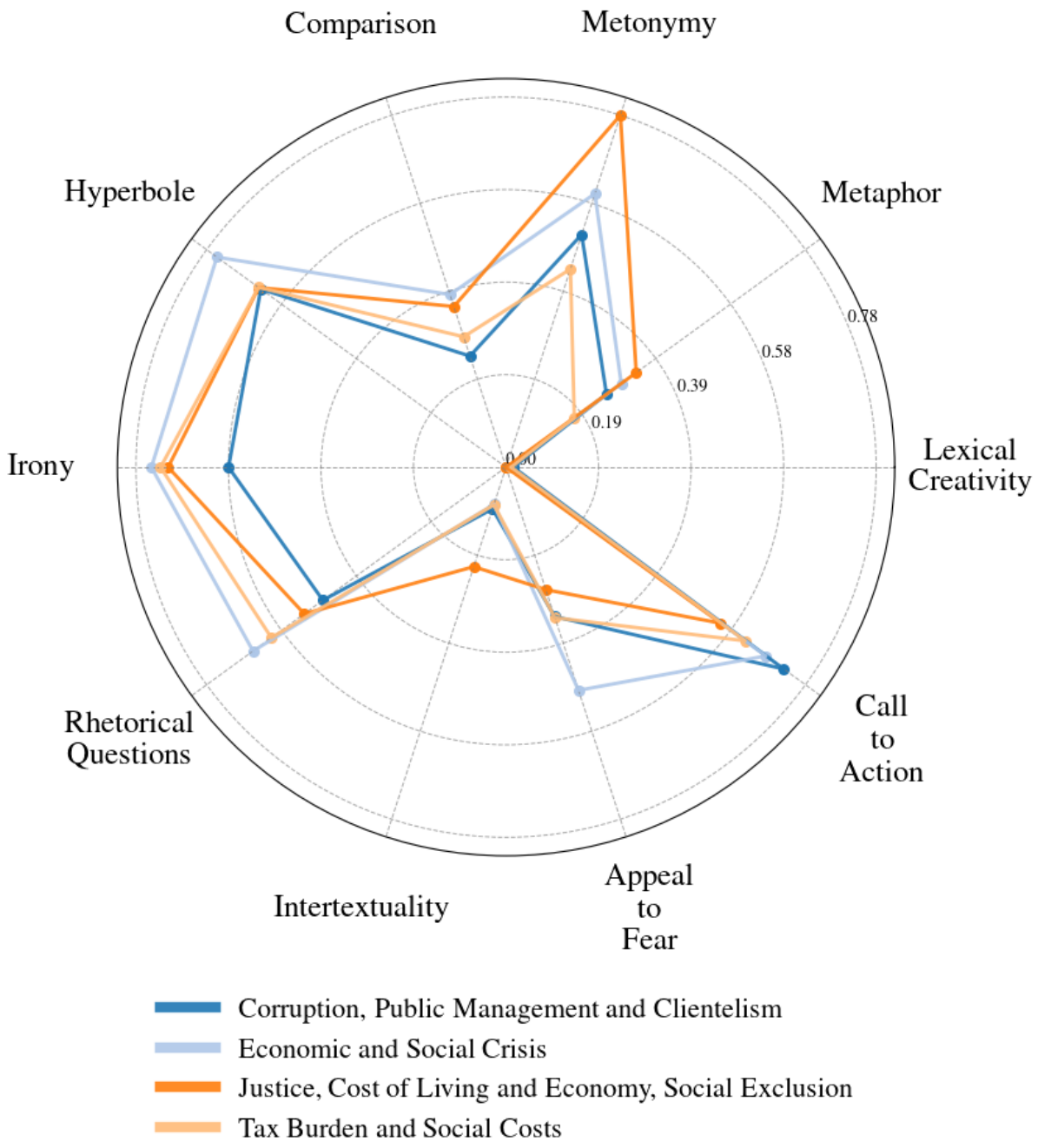


FIGURE D.65. Rhetorical devices probability distribution across *'Socio-Economic Criticism and Debate'*.

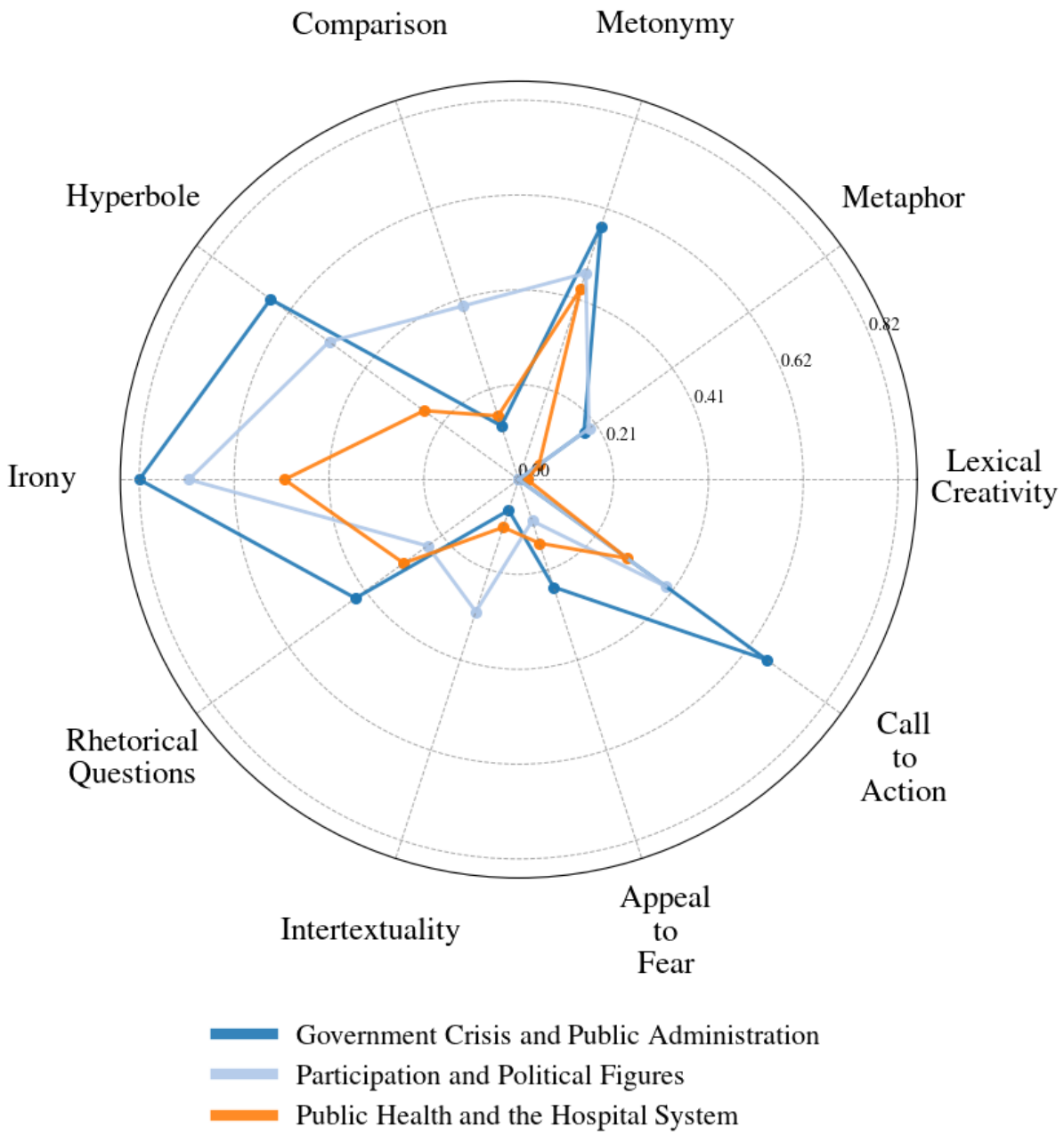


FIGURE D.66. Rhetorical devices probability distribution across *'Controversies and Disputes about Political Leadership and Government Management'*.

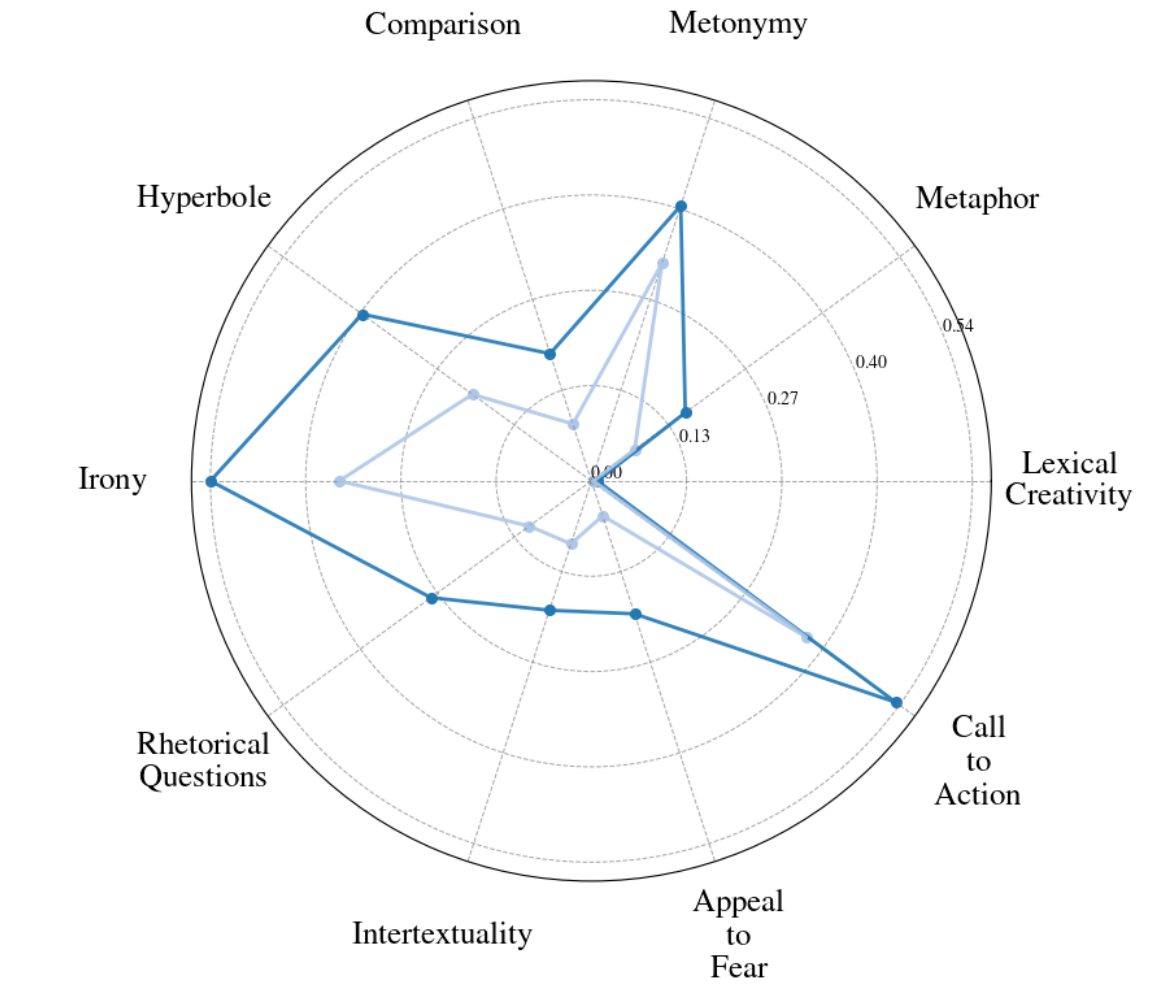


FIGURE D.67. Rhetorical devices probability distribution across *'Political-Electoral Dynamics, Public Management, Ideological Debates and Media Coverage'*.

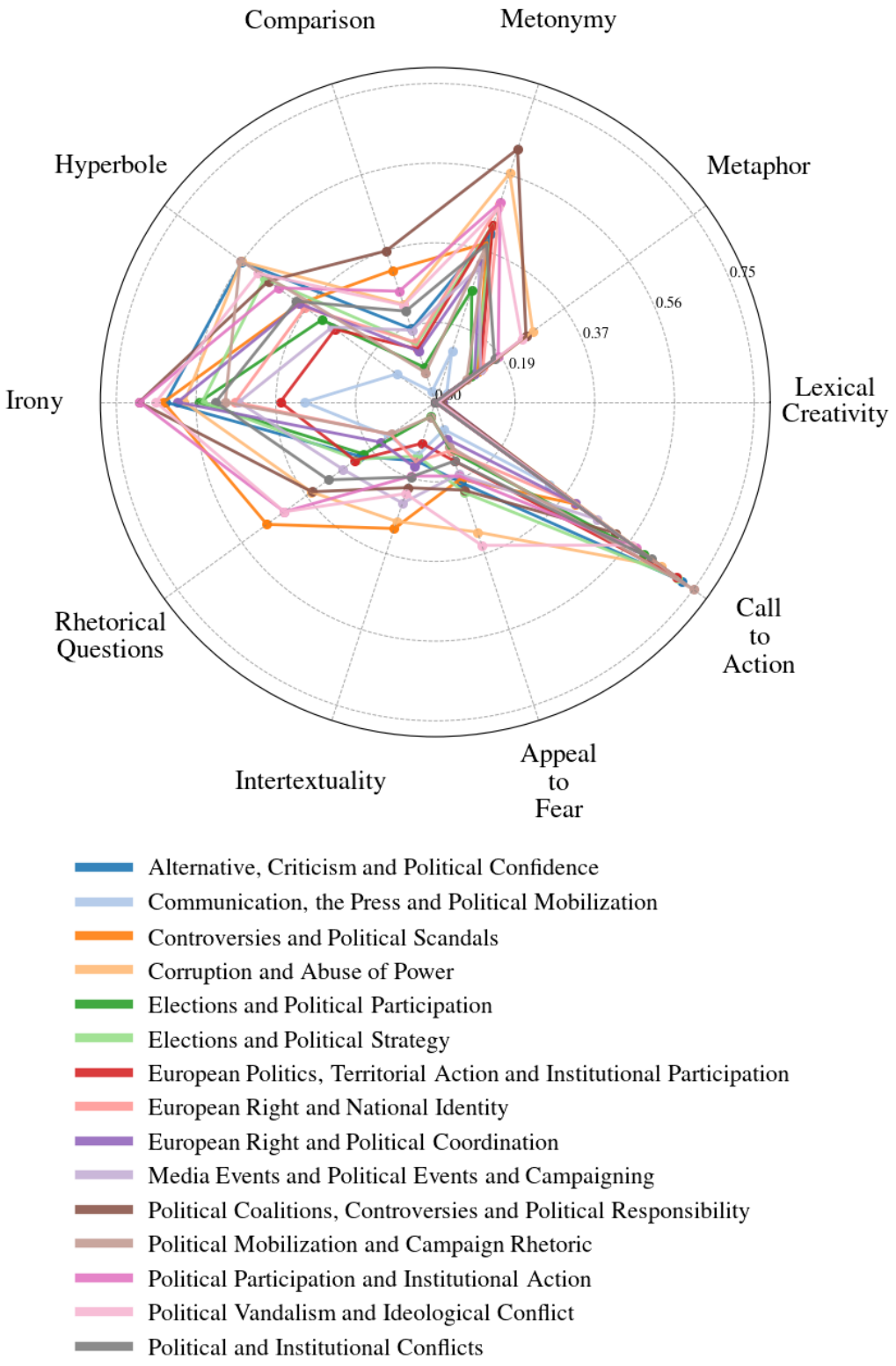


FIGURE D.68. Rhetorical devices probability distribution across *'Political Dispute, Ideological Positions, Communication and Media'*.

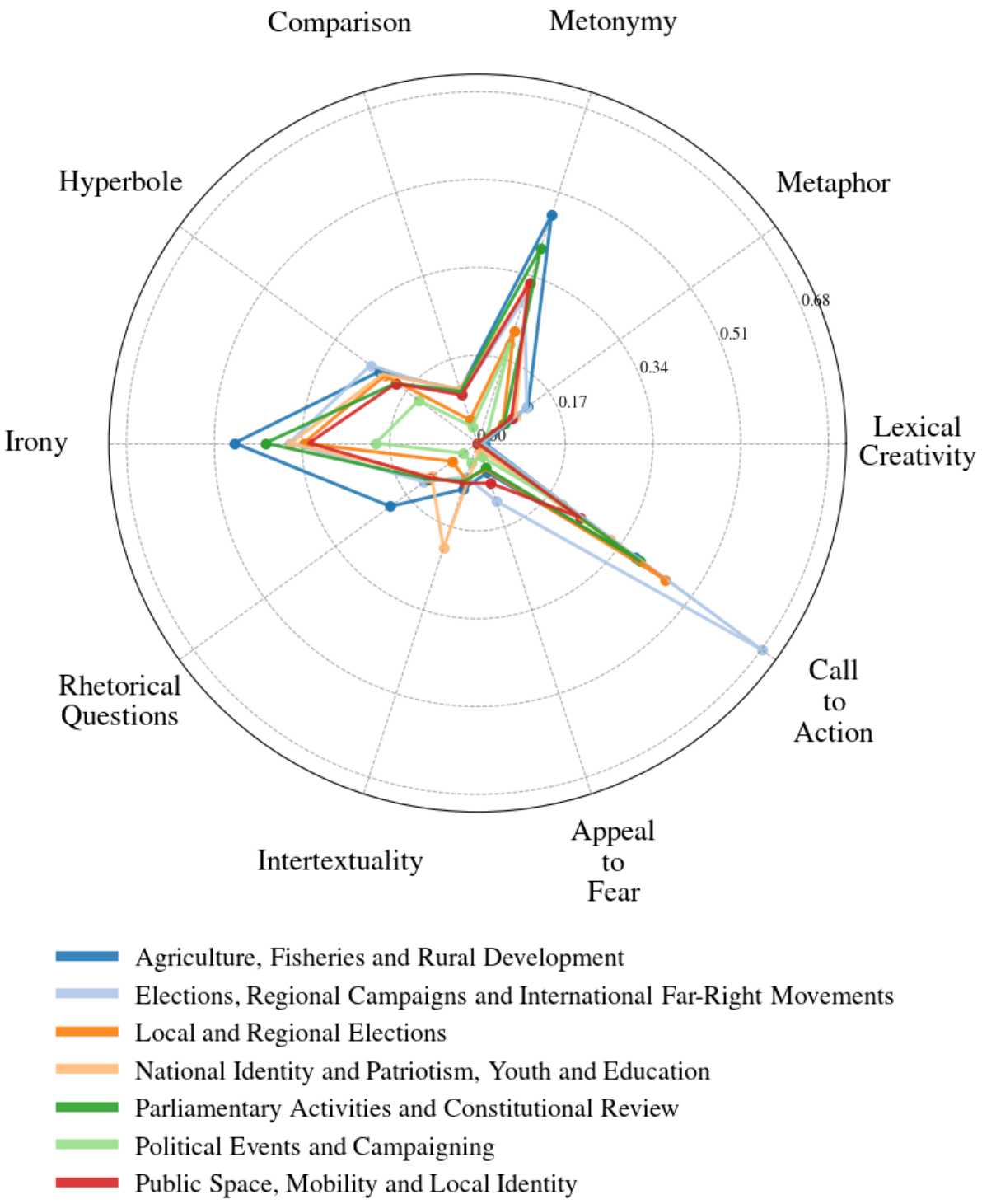


FIGURE D.69. Rhetorical devices probability distribution across *'Political-Parliamentary and Electoral Cycle, Mobilization, Interaction and Event Organization'*.

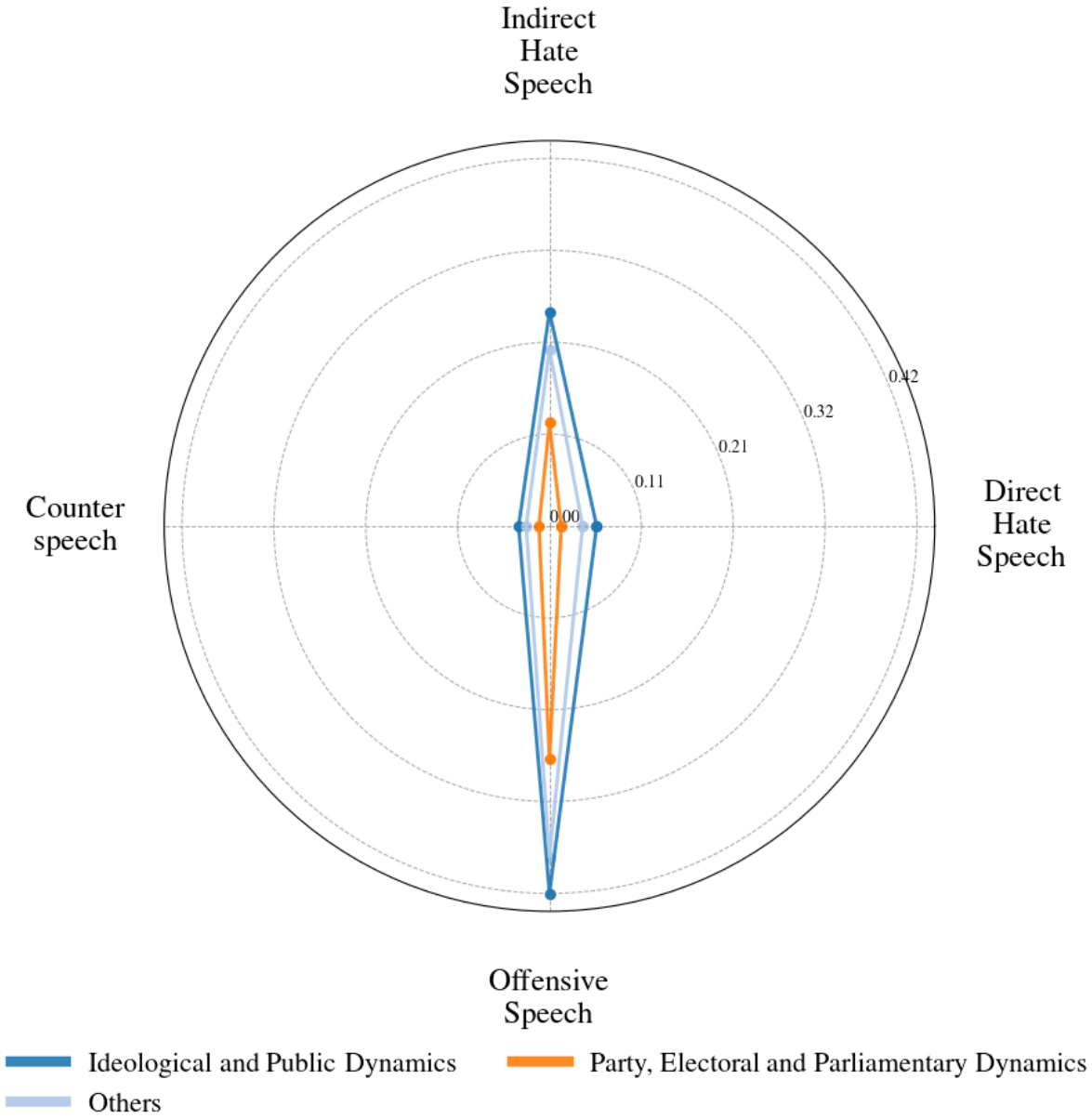


FIGURE D.70. Speech types probability distribution across *'Political Discourse'* dimension.

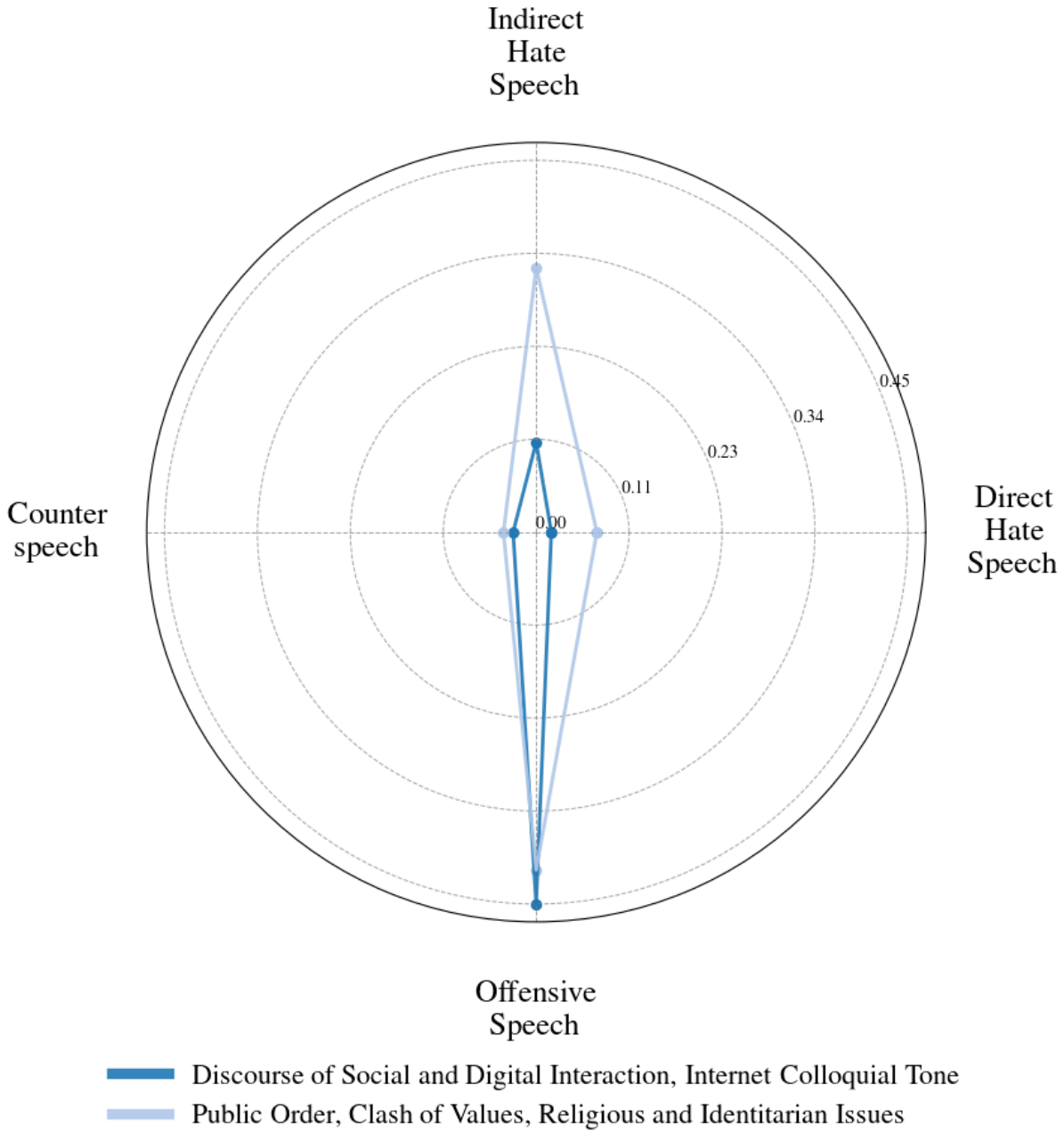


FIGURE D.71. Speech types probability distribution across *'Ideological and Public Dynamics'* Dimension.

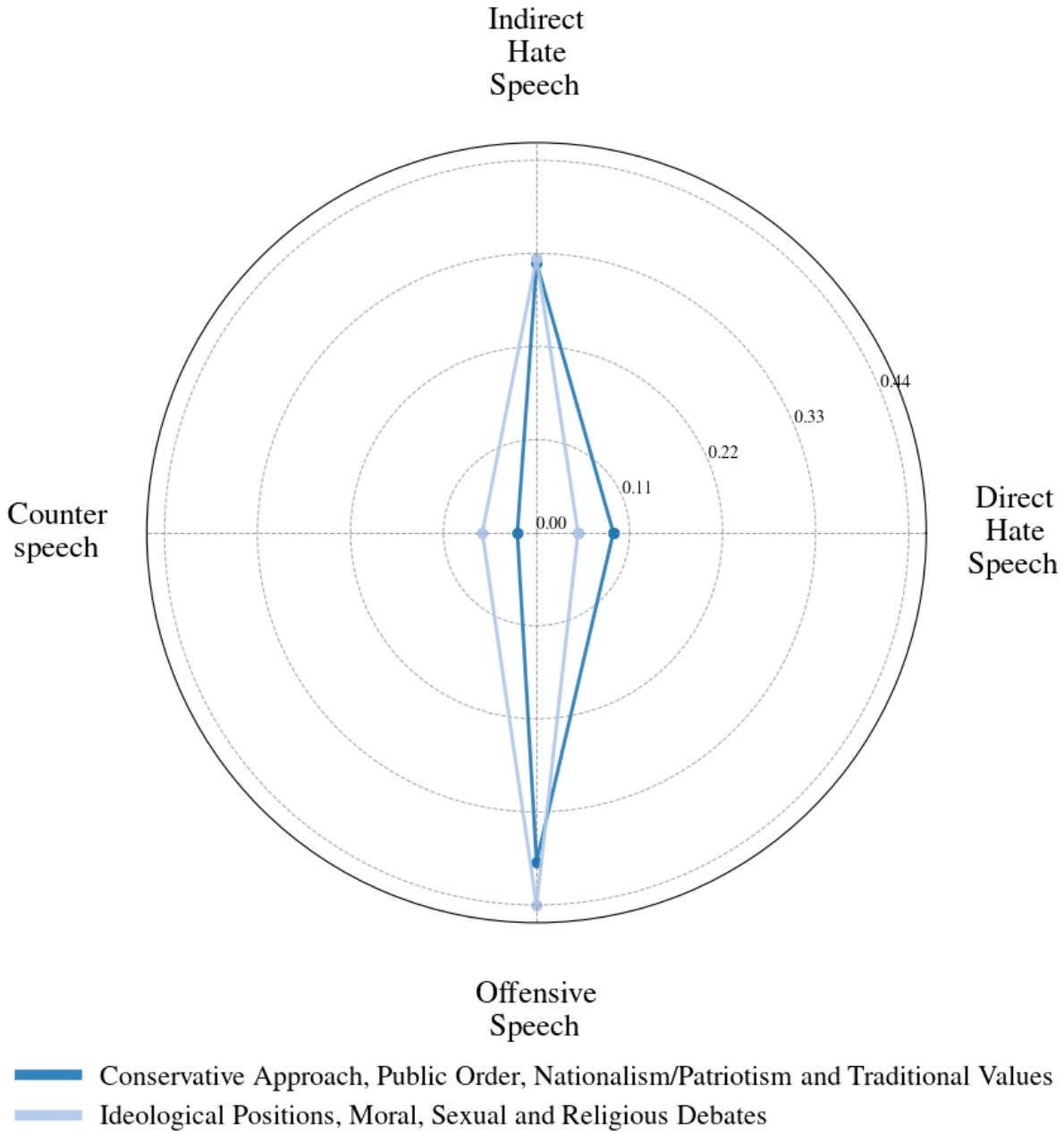


FIGURE D.72. Speech types probability distribution across *'Public Order, Clash of Values, Religious and Identitarian Issues'*.

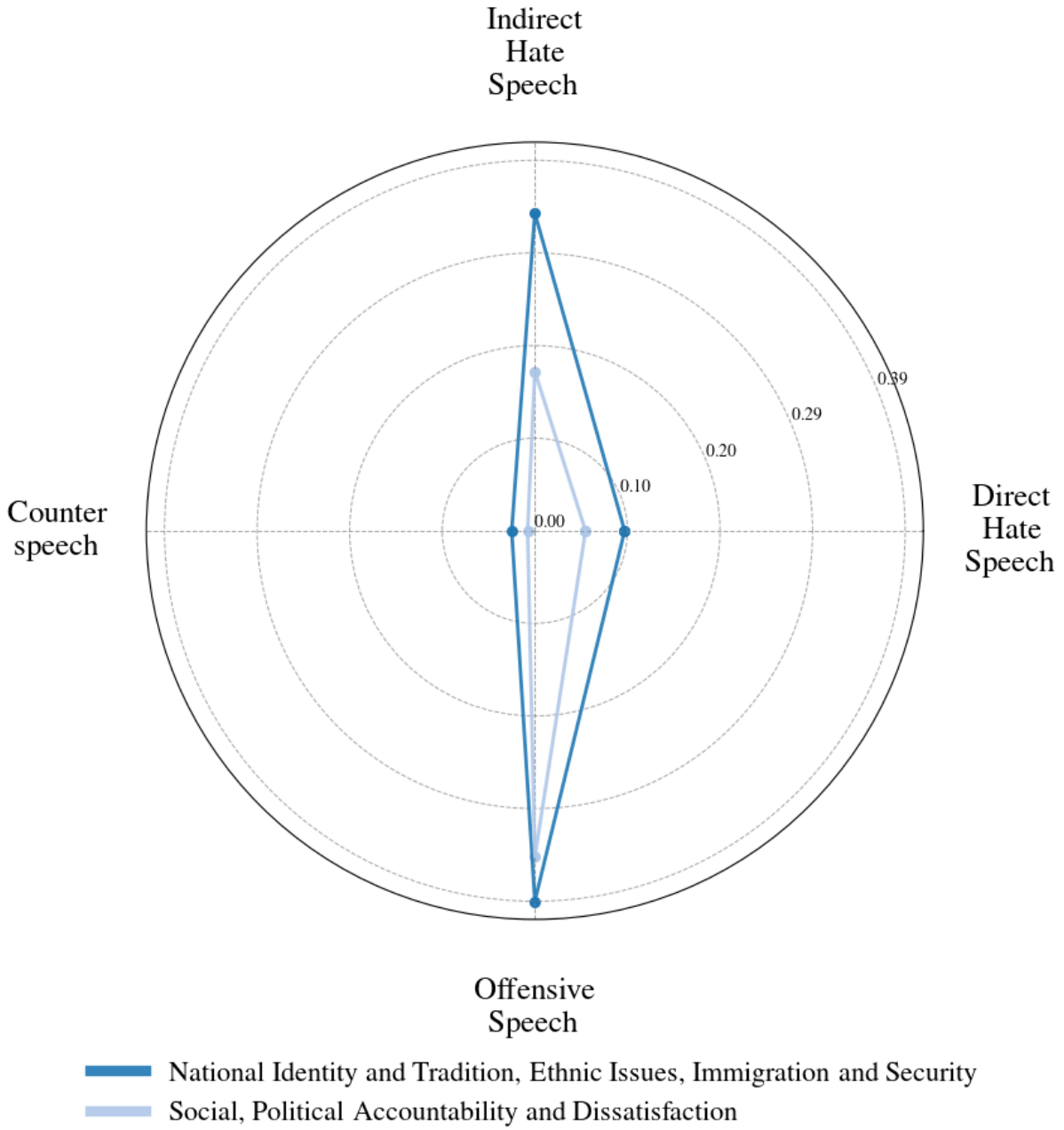


FIGURE D.73. Speech types probability distribution across *'Conservative Approach, Public Order, Nationalism/Patriotism and Traditional Values'*.

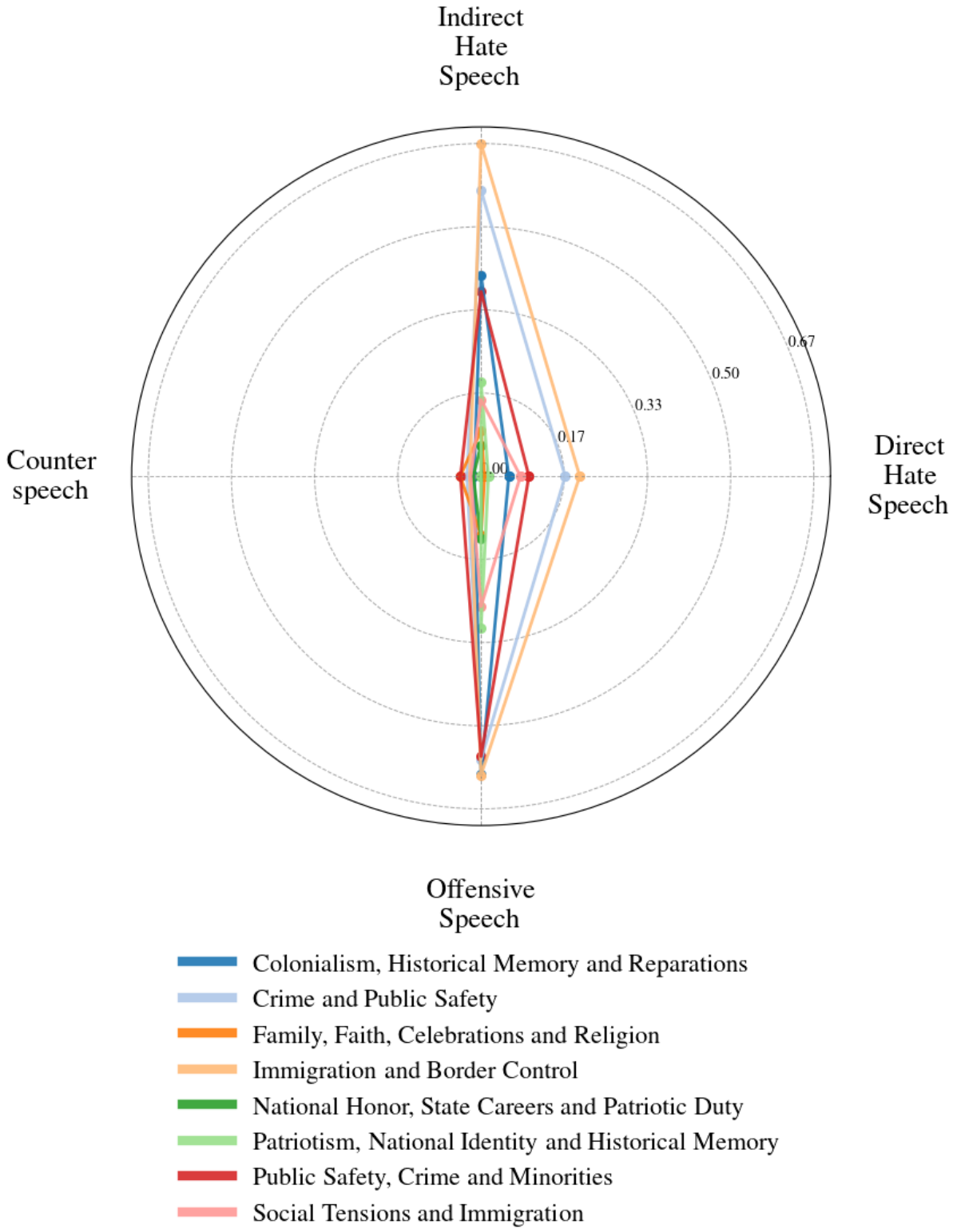


FIGURE D.74. Speech types probability distribution across *'National Identity and Tradition, Ethnic Issues, Immigration and Security'*.

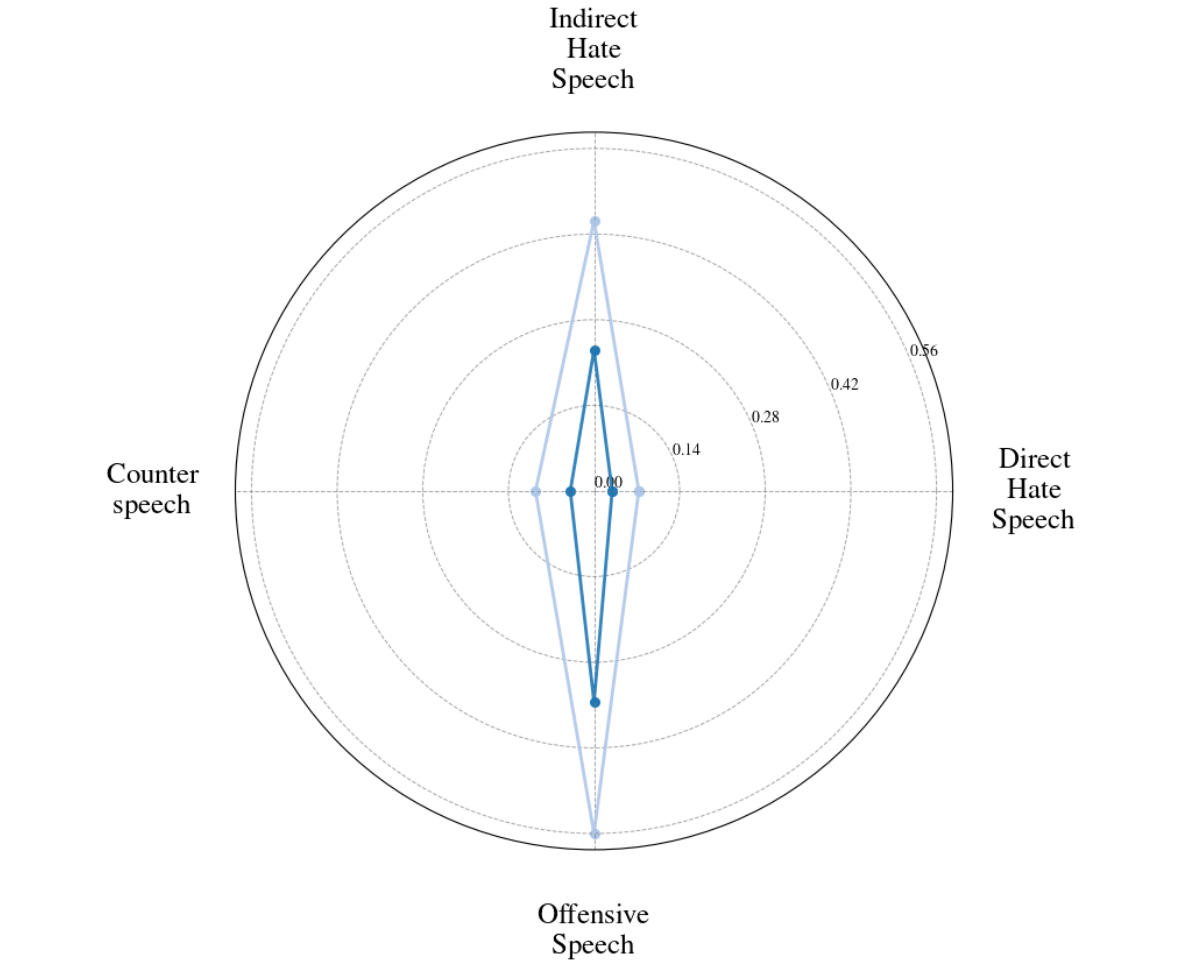


FIGURE D.75. Speech types probability distribution across *'Ideological Positions, Moral, Sexual and Religious Debates'*.

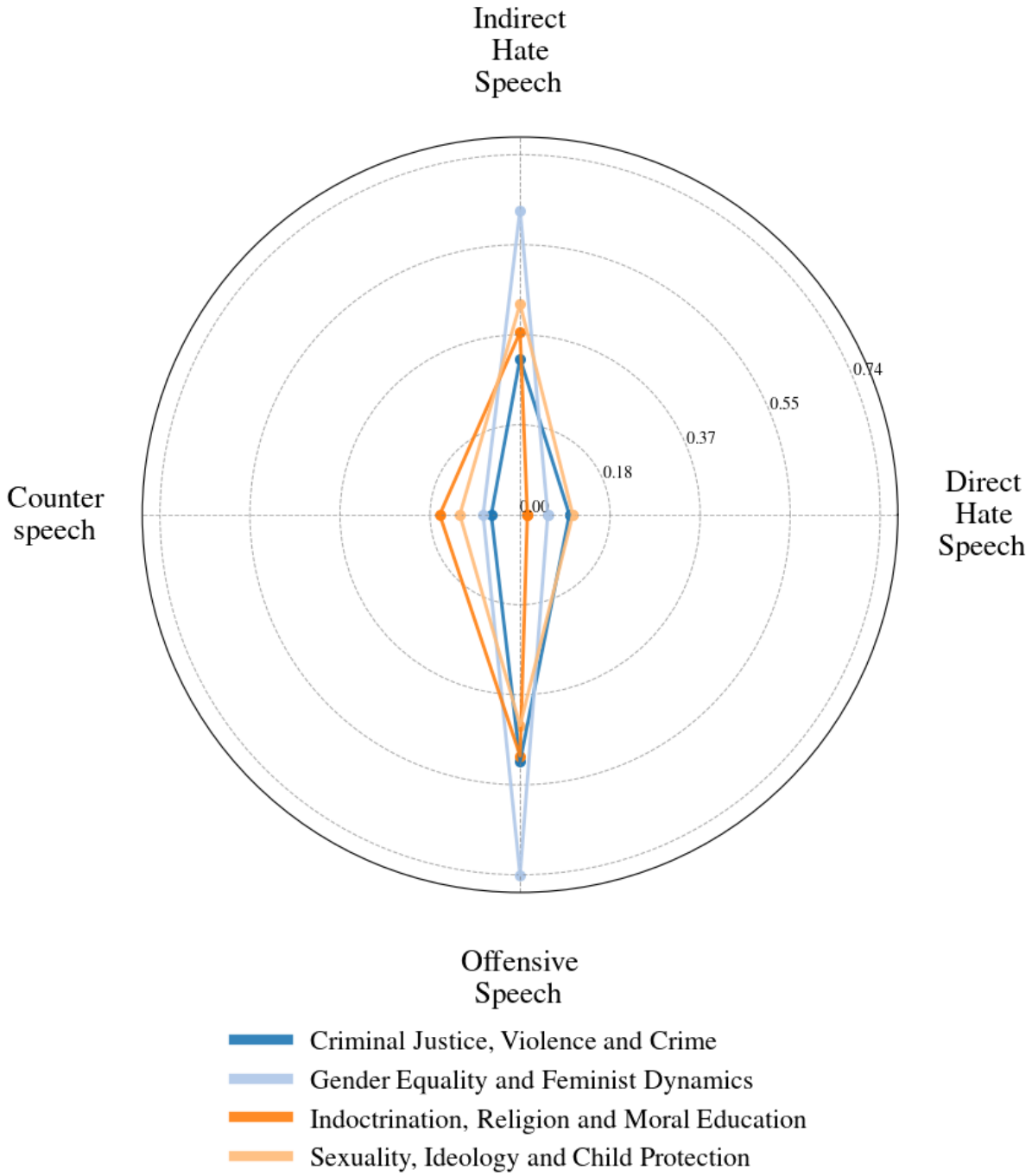


FIGURE D.76. Speech types probability distribution across *'Sexual Crimes, Feminism/Abortion, Religious Education and Gender/Sexuality Debates'*.

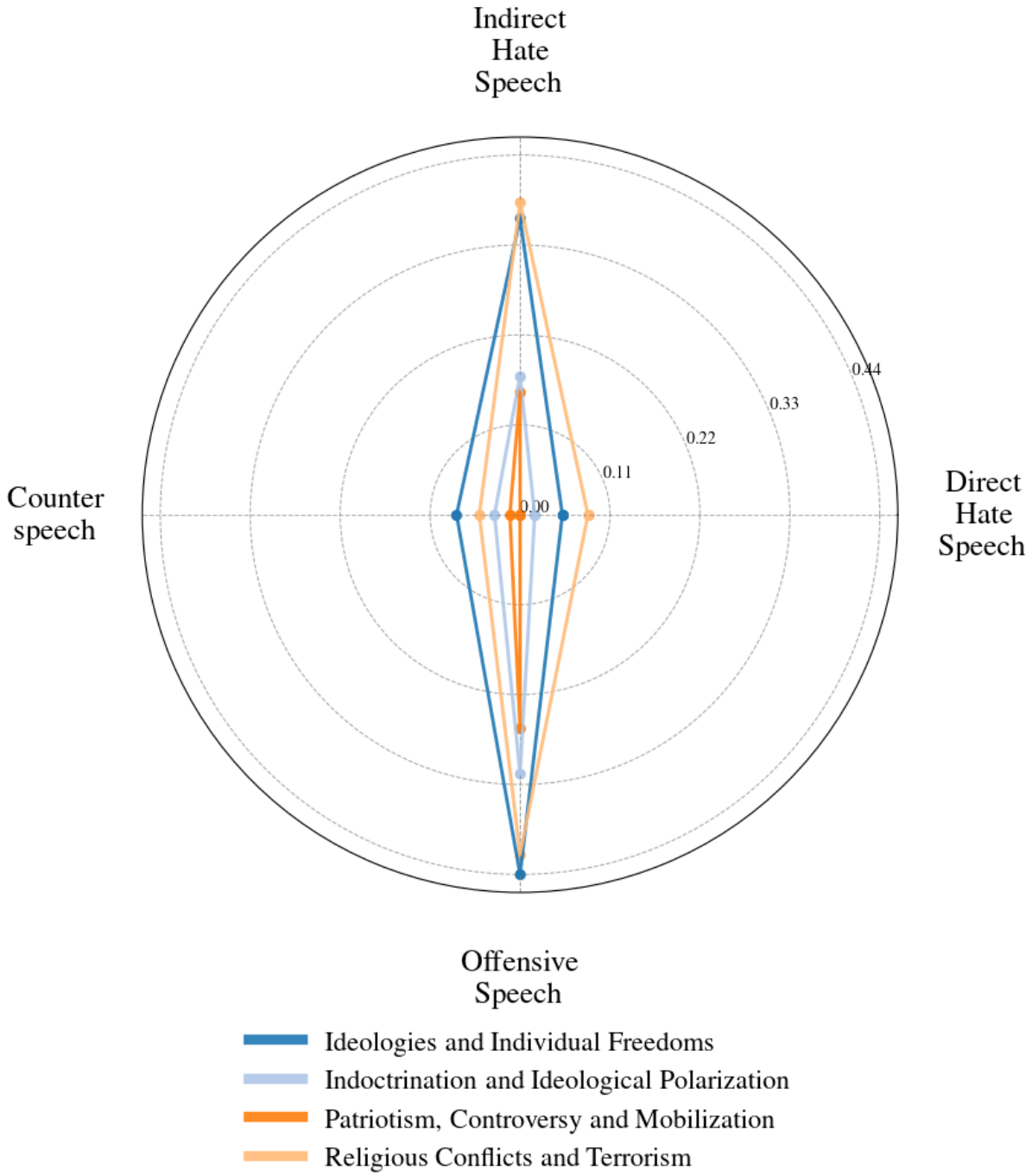


FIGURE D.77. Speech types probability distribution across *'Religious Conflicts, Political Ideologies, Extreme Disputes, and Relations between Belief, Morals and Politics'*.

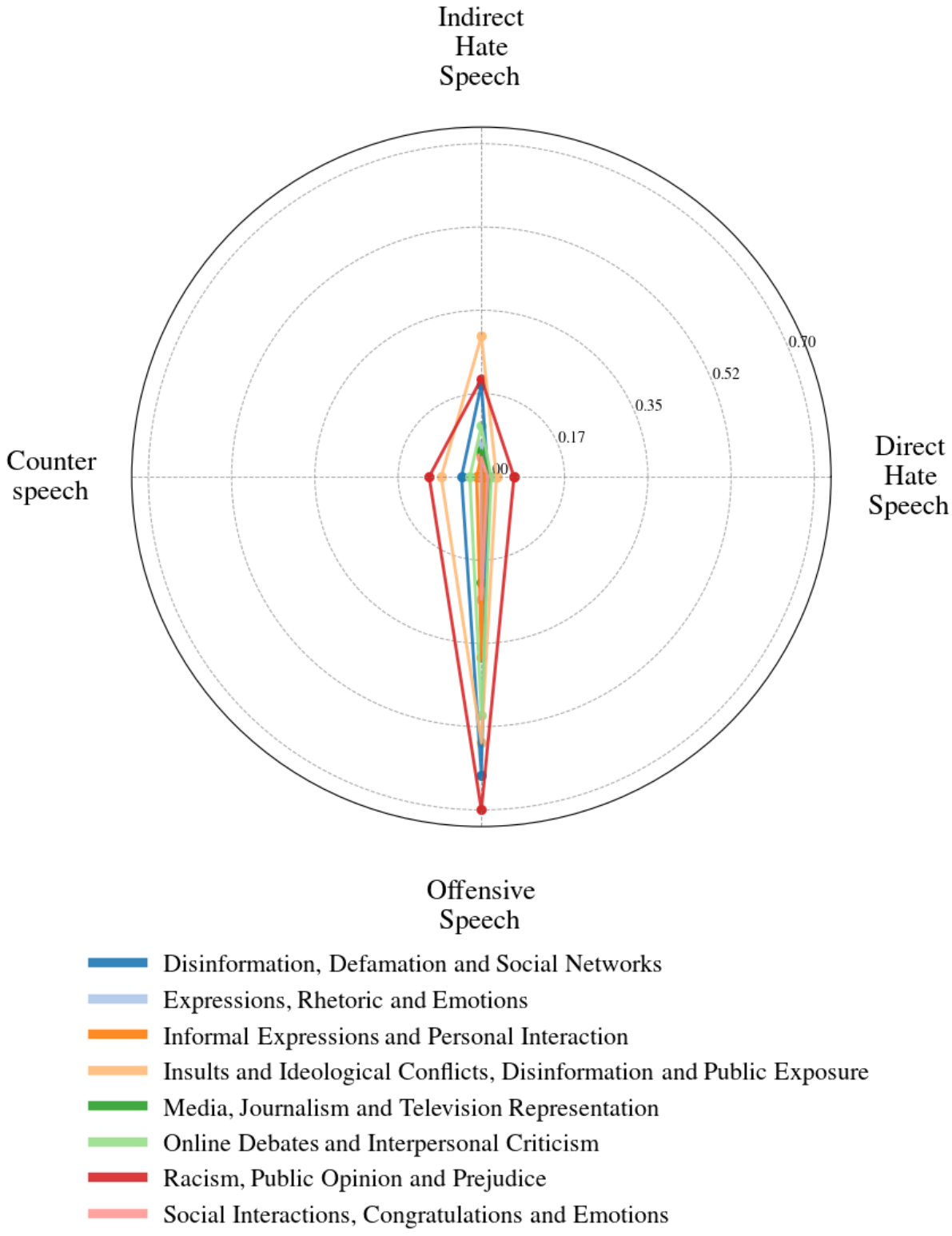


FIGURE D.78. Speech types probability distribution across *'Discourse of Social and Digital Interaction, Internet Colloquial Tone'*.

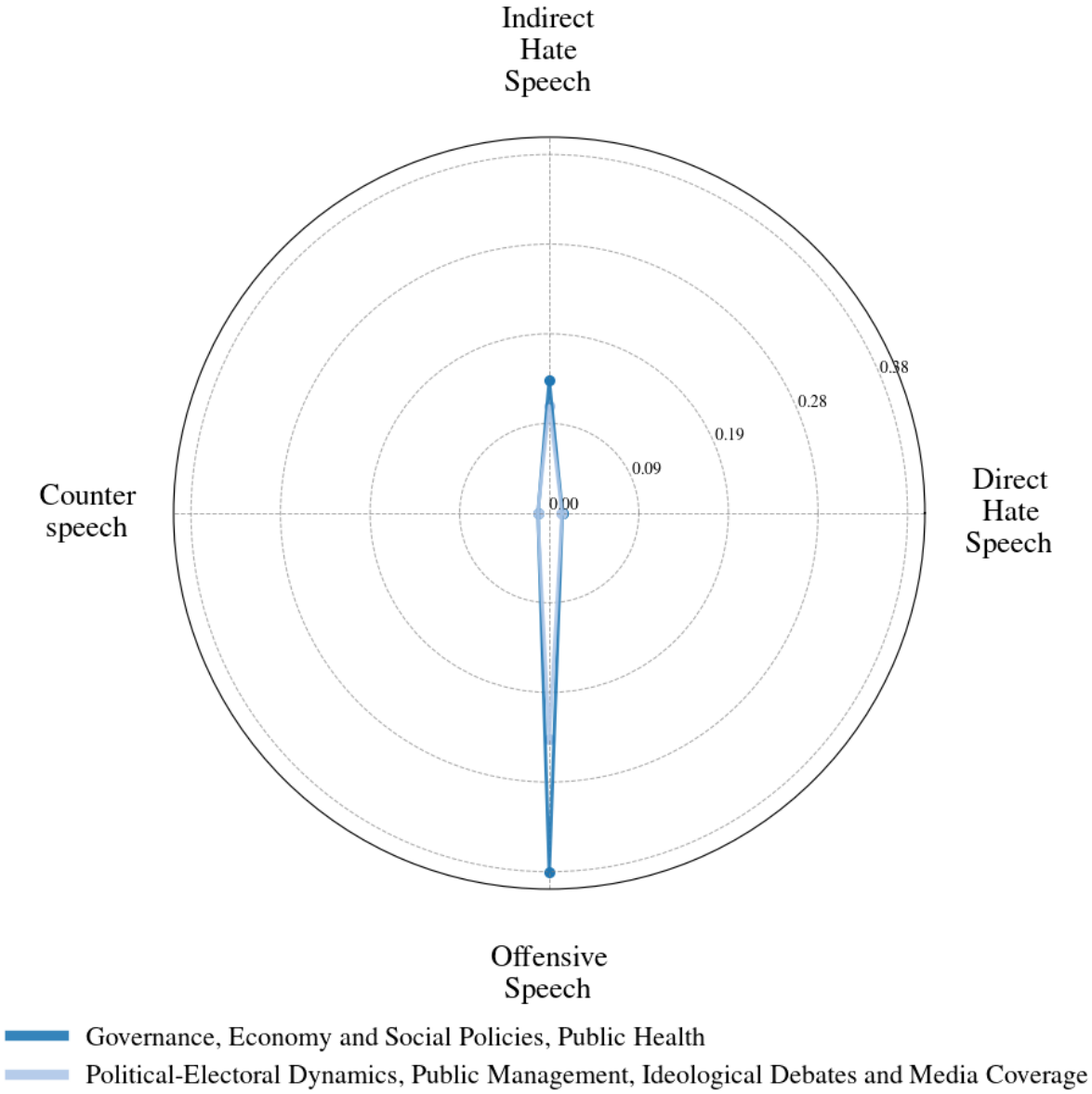


FIGURE D.79. Speech types probability distribution across *'Party, Electoral and Parliamentary Dynamics'* Dimension.

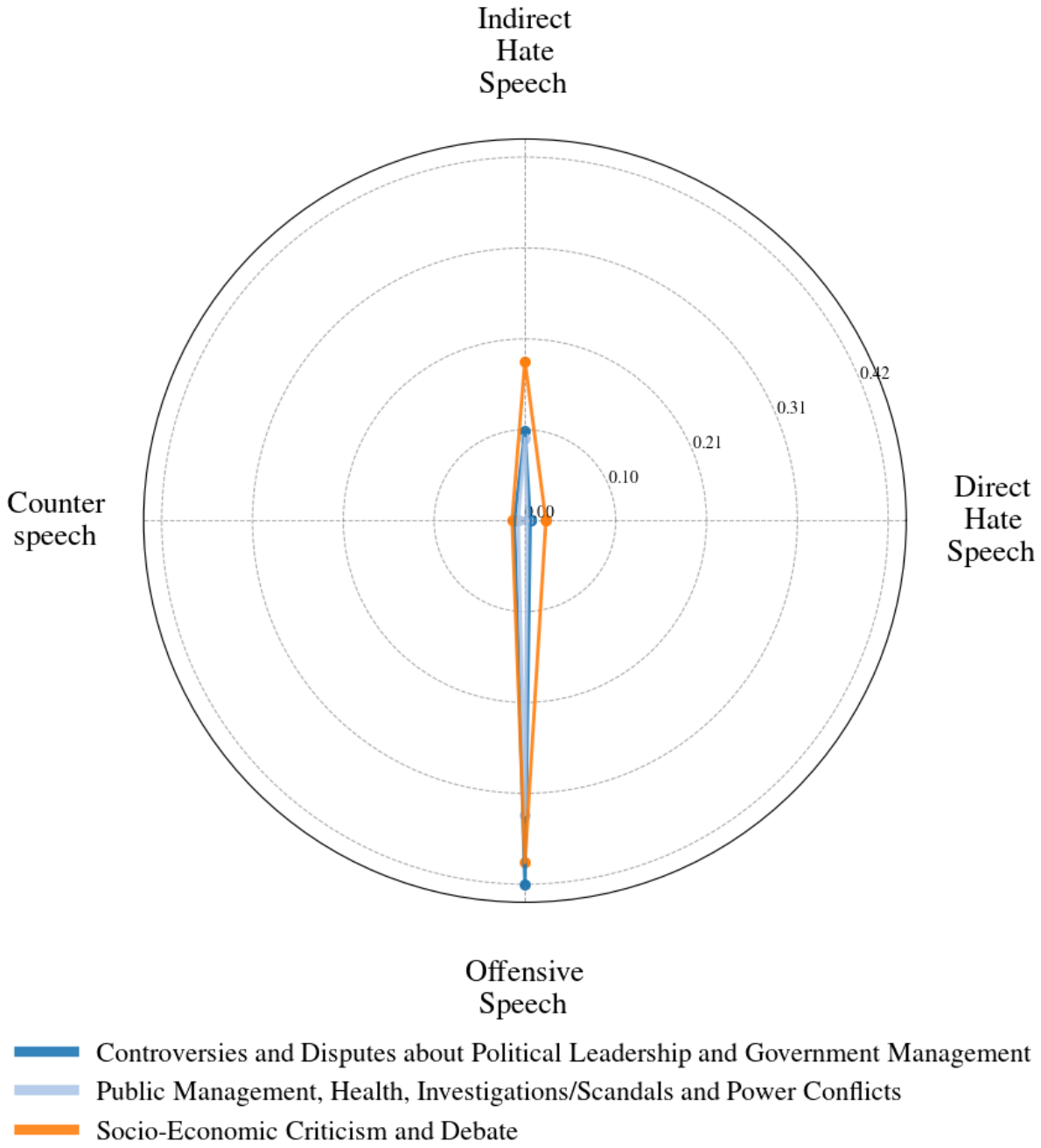


FIGURE D.80. Speech types probability distribution across *'Governance, Economy and Social Policies, Public Health'*.

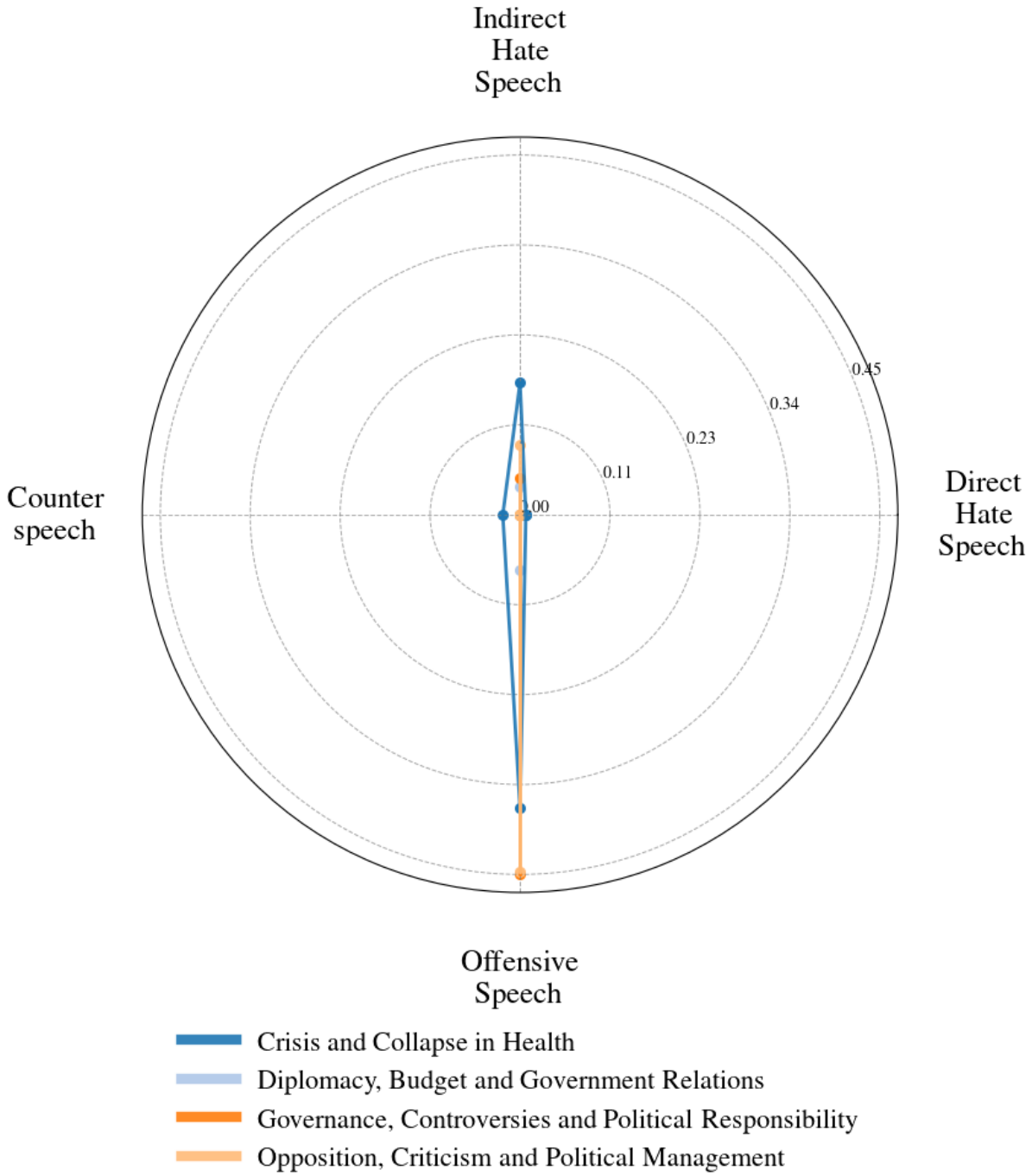


FIGURE D.81. Speech types probability distribution across *'Public Management, Health, Investigations/Scandals and Power Conflicts'*.

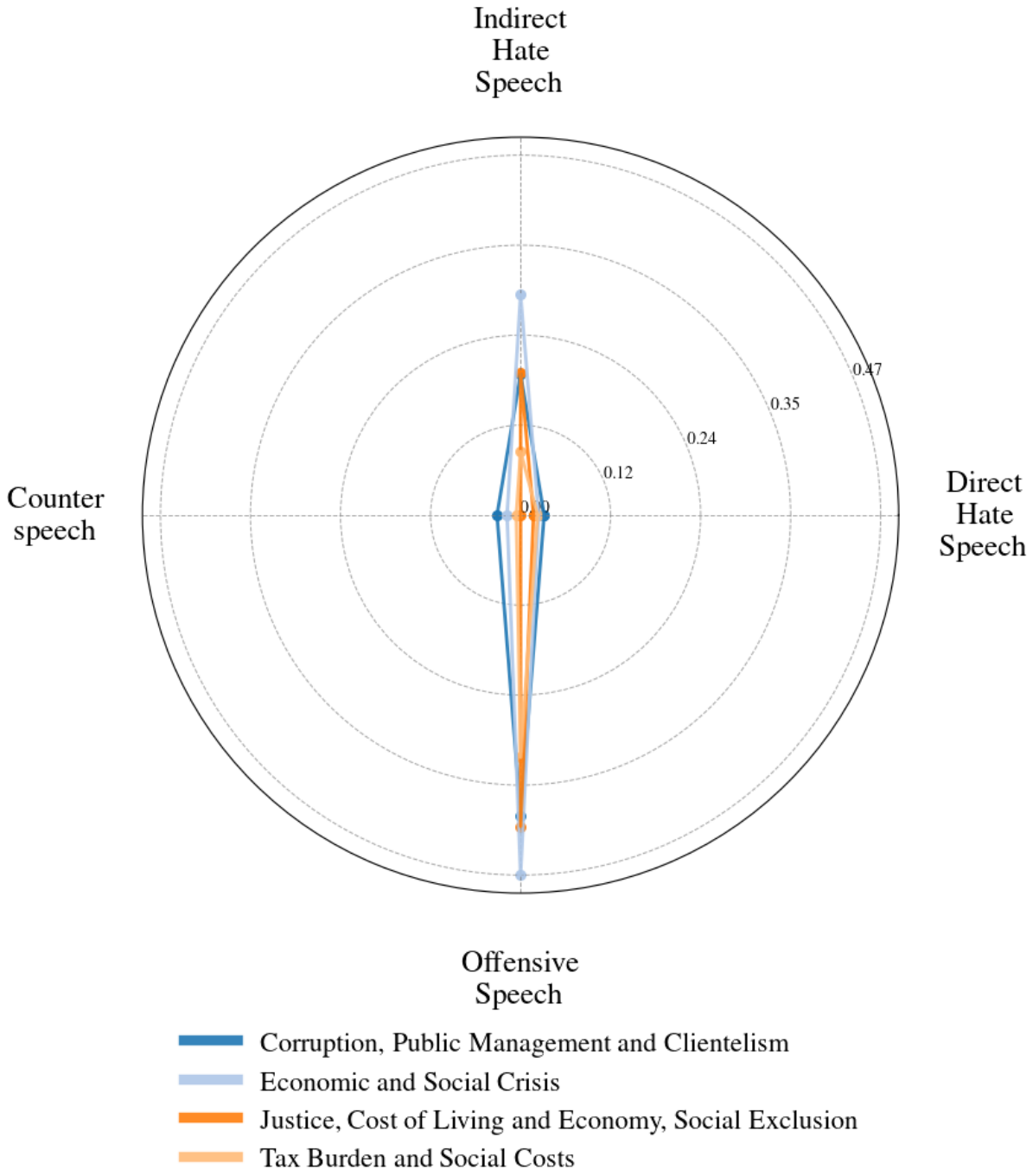


FIGURE D.82. Speech types probability distribution across *'Socio-Economic Criticism and Debate'*.

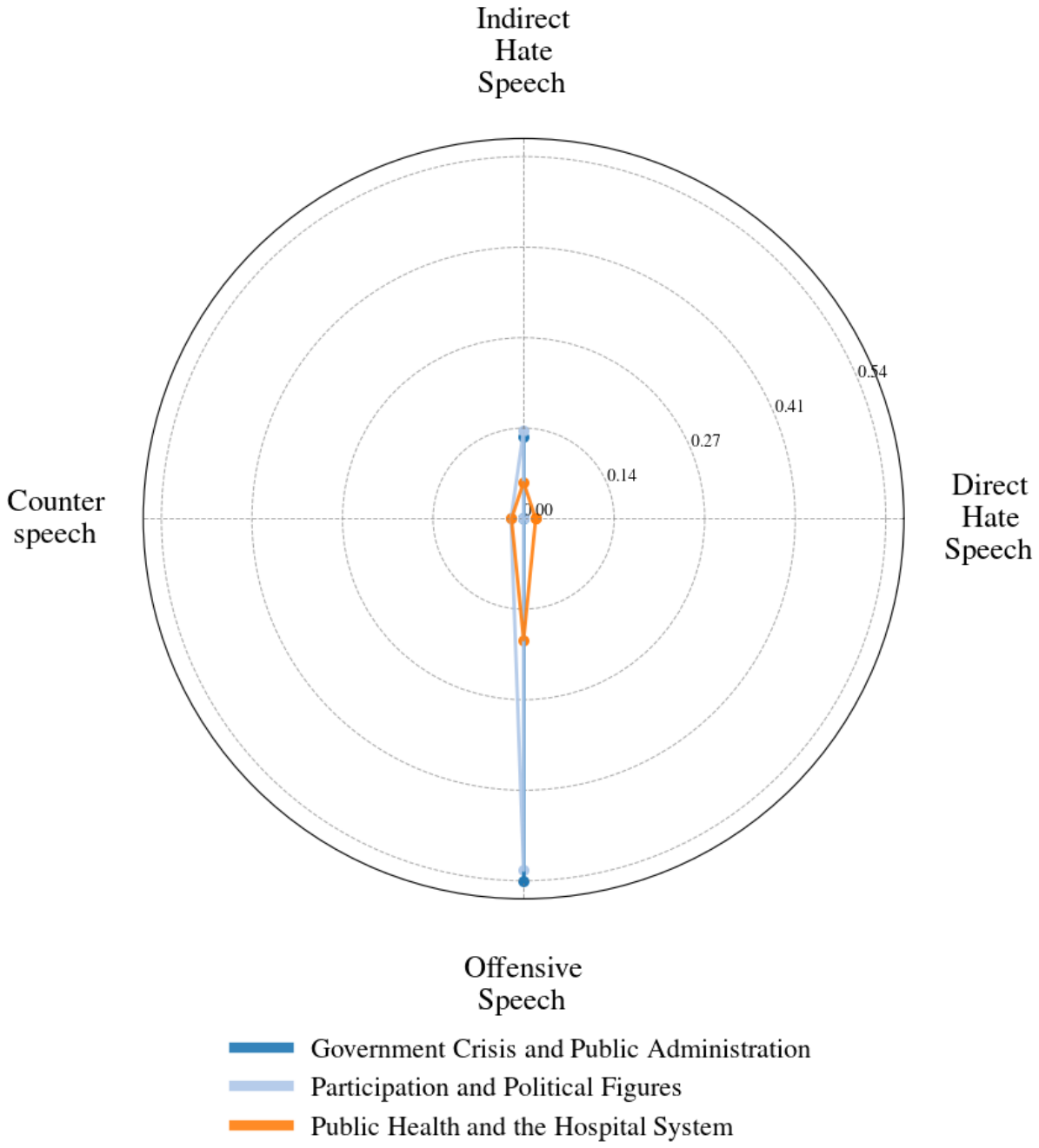


FIGURE D.83. Speech types probability distribution across *'Controversies and Disputes about Political Leadership and Government Management'*.

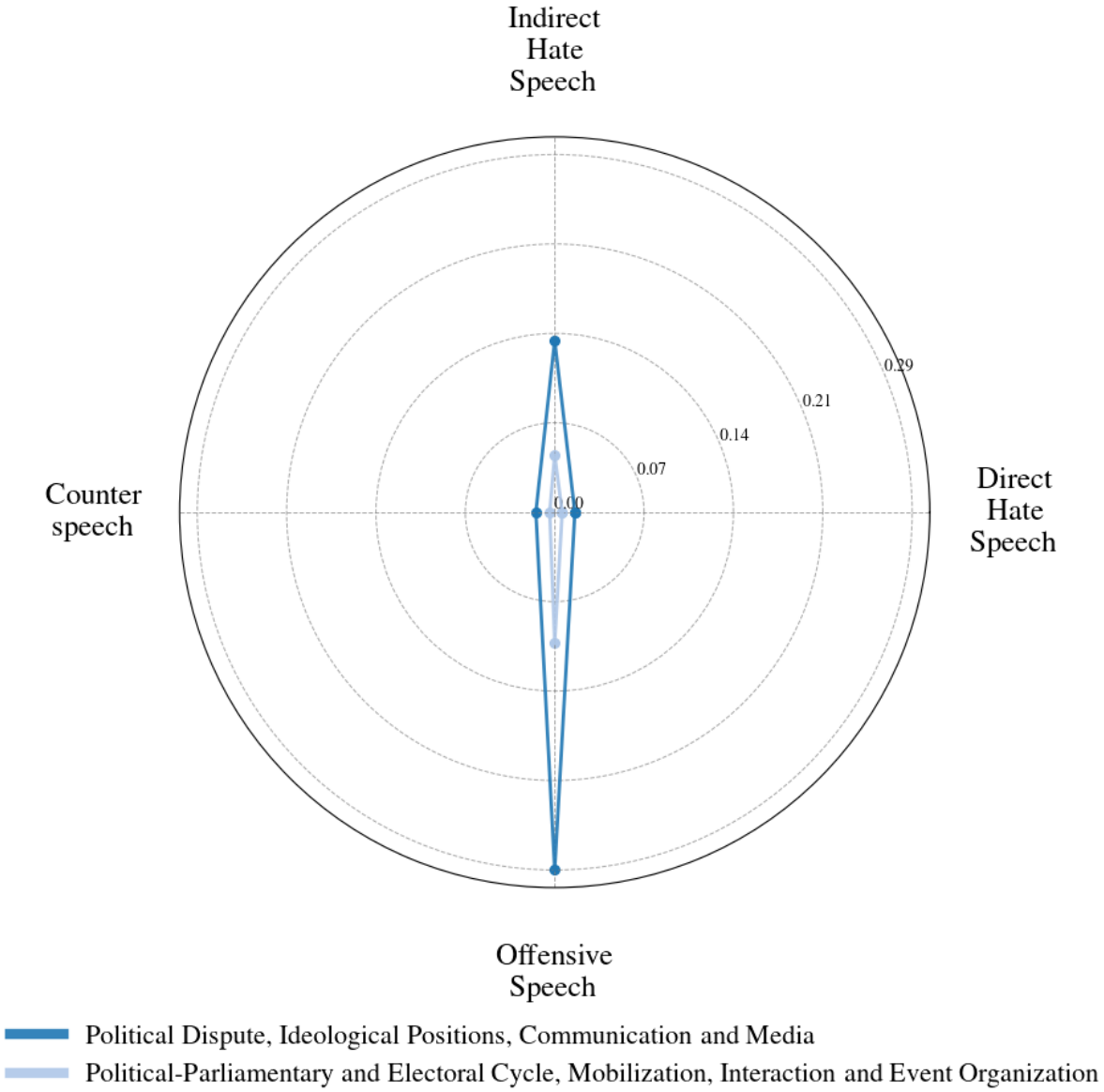


FIGURE D.84. Speech types probability distribution across *'Political-Electoral Dynamics, Public Management, Ideological Debates and Media Coverage'*.

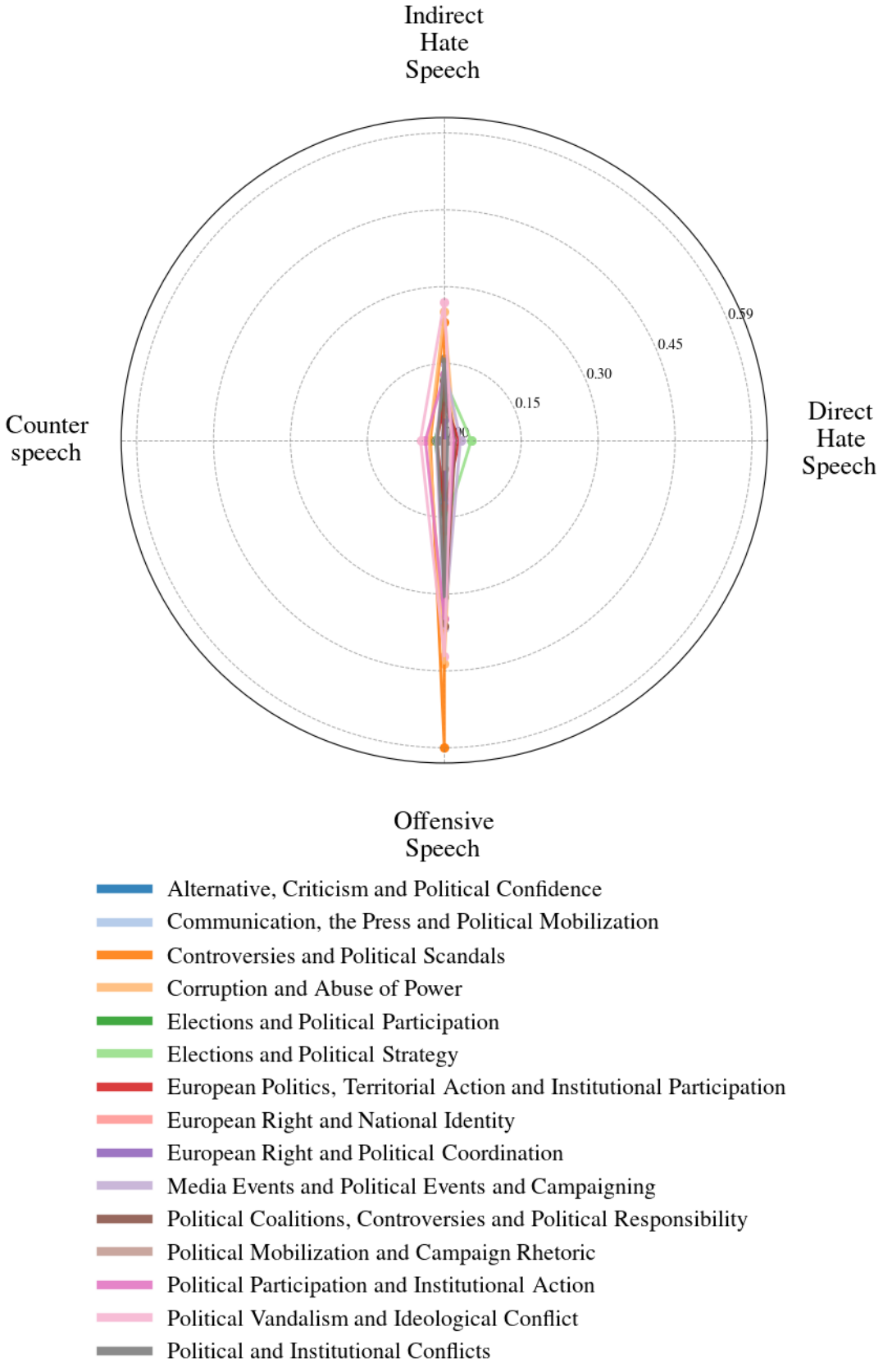


FIGURE D.85. Speech types probability distribution across *'Political Dispute, Ideological Positions, Communication and Media'*.

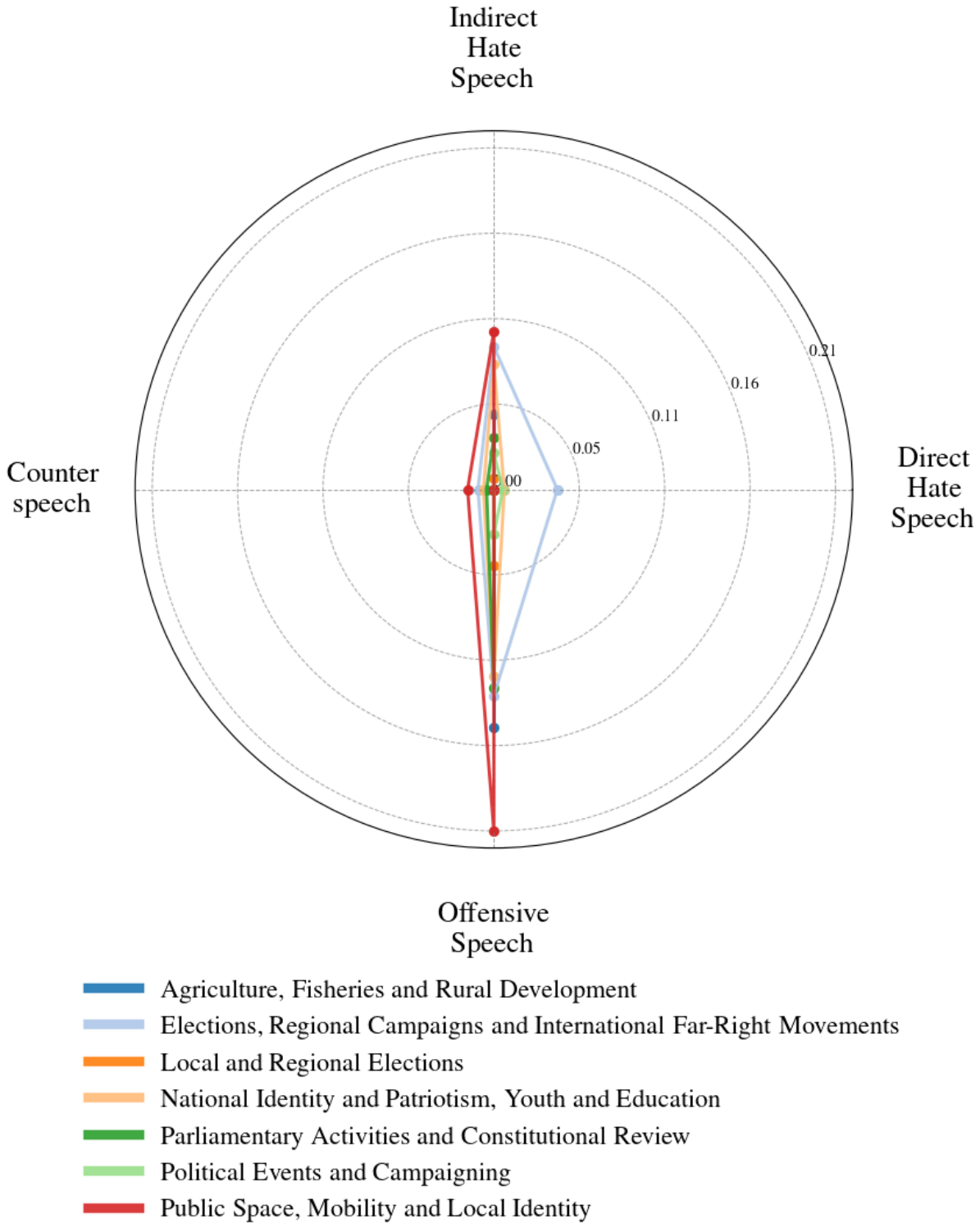


FIGURE D.86. Speech types probability distribution across *'Political-Parliamentary and Electoral Cycle, Mobilization, Interaction and Event Organization'*.