

Lund University
Department of Political Science

STVK12
Tutor: Agustin Goenaga

Computational Scaling of Party Cohesion in Social Media

A Case Study on Twitter Campaigns in Brazil

Emma Olivia Baker

Bachelor's Thesis

26th May 2022

Abstract

This thesis explores party cohesion in politicians' social media campaigns in the context of Brazil. It highlights the importance of studying the interaction between digitalisation of political communication and the party system in developing democracies. Brazil makes an interesting case as it is known for its exceptionally unstable party system. This study utilises theories from party discipline and personalisation of politics literature to explain possible cross-party and individual-level variations in party cohesion in social media. It measures party cohesion on social media using the concept of rhetorical cohesion – referring to politicians' similarity in their use of language and speech. Moreover, this thesis uses a quantitative research design utilizing a computational scaling algorithm (Wordfish) and statistical data analysis to investigate rhetorical cohesion in a dataset retrieved from Twitter. The findings of the thesis indicate that there is no clear ideological determinant for party cohesion in social media campaigns in Brazil. However, the Worker's Party (PT) continues to demonstrate relatively high level of party cohesion, indicating that this finding, aligning with previous studies, also applies in the context of social media. Additionally, this thesis finds strong evidence that presidential candidacy is a positive individual-level predictor of rhetorical cohesion with one's own party.

Keywords: party cohesion, Brazil, social media, political campaigns, Wordfish

Word count: 9997

Contents

| | | |
|-----------|--|-----------|
| 1 | Introduction | 1 |
| 2 | Aims and Research Questions | 1 |
| 3 | Previous Literature | 3 |
| 4 | Background and Case Selection | 4 |
| 5 | Theoretical Framework | 5 |
| 5.1 | Party Cohesion and Party Discipline | 6 |
| 5.2 | Personalisation of Politics | 8 |
| 5.3 | Rhetorical Cohesion | 10 |
| 5.4 | Individual-level factors and rhetorical cohesion | 10 |
| 6 | Data | 12 |
| 6.1 | Twitter Data | 12 |
| 6.2 | Sample | 13 |
| 6.3 | Ethical Considerations | 14 |
| 7 | Methodology | 14 |
| 7.1 | Wordfish | 15 |
| 7.2 | Statistical Analysis | 18 |
| 7.3 | Limitations | 19 |
| 8 | Results | 21 |
| 8.1 | Party-level Analysis | 21 |
| 8.2 | Individual-level Analysis | 22 |
| 9 | Discussion and Conclusion | 25 |
| 10 | Bibliography | 29 |

1 Introduction

For more than a decade now, political communication has undergone significant changes with the development of digital communication tools. While the internet is still making its way to reach parts of the globe, political representatives are increasingly using social media to reach out to their electorate. At the same time, during the past few decades, scholars have witnessed increasing attention to individual candidates at the expense of political parties. According to some scholars, the use of digital communication tools, namely social media, has given individual politicians more autonomy over their political party and thus political campaigns have become more individualised. The technological advancements in communication tools demonstrate a need to expand perspectives to see how the cohesiveness of political parties is reflected in online political communication.

Latin America is one of the regions in the world with a significant proportion of their population - and politicians - active on social media. Additionally, the region is generally found to have a rather unstable and weak party system. This thesis investigates party cohesion in social media in the context of Brazil. Like other countries in Latin America, Brazil is known for its weak party system. Lack of party discipline has been presented as a possible explanation for weak and unstable political parties in the country. However, party cohesion and discipline in Brazil has been primarily studied in the context of parliamentary politics.

This thesis aims to uncover how cohesive different political parties in Brazil appear in their representatives' Twitter campaigns as well as which individual-level political conditions could potentially correlate with politicians adopting more individualised rather than party-centred social media campaigns. Its research design uses a mix of computational data analysis and statistical analysis to reach these aims.

This thesis has the following structure. First, it presents its main aims and research questions. Second, it summarises the previous research on individualised campaigning in social media, and explains why Brazil makes an interesting case for this study. Third, it lays down its theoretical framework by linking two theories that have aimed to explain varying levels of party cohesion: party discipline and personalisation of politics. Finally, it explains which data and methodology it uses to answer the research questions, before presenting and analysing the results.

2 Aims and Research Questions

This thesis aims to analyse party cohesion in terms of representatives' social media campaigns in Brazil. Specifically, the goal of this thesis is two-fold. First, it aims to distinguish how coherent social media campaigns are within different political parties

in Brazil. Second, it aims to explore whether certain political characteristics of individual politicians, namely length of ones political career and presidential candidacy, affect the extent to which they deviate from their respective political party. The overall aim is to uncover which cross-party and individual-level characteristics could affect to what extent representatives within political parties adopt more cohesive social media campaigns. It does this through the concept of rhetorical cohesion, which in this case describes the similarity of language in representatives' Twitter campaigns. In other words, this thesis wishes to explore if certain political parties in Brazil appear more or less cohesive in social media, and if specific individual-level political characteristics correlate with more individualised campaigning in social media. In short, it aims to answer the following two research questions:

How do political parties in Brazil differ in terms of rhetorical cohesion of politicians' social media campaigns?

What political characteristics of individual representatives affect rhetorical cohesion within political parties?

Furthermore, this thesis contributes to the existing scholarly literature on the topic of party cohesion in social media campaigns in the context of developing democracies with relatively weak and unstable party systems. It links this research topic to two strains of theoretical literature: party discipline and personalisation of politics.

The aim is to gain knowledge on the use of digital communication technologies in political communication in Brazil as well as how this interacts with the country's party system.

This thesis aims to answer the research questions by analysing the rhetorical cohesion of Brazilian representatives' Twitter timelines using a combination of computational and statistical data analysis methods. In particular, it uses an unsupervised machine-learning algorithm called Wordfish, which is used to scale political texts on a one dimensional scale based on the the vocabulary they use (Huneus, 2021; Aydogan, Tuna and Yildirim, 2021). This allows one to observe how cohesive or scattered politicians within each party appear on this scale. It also enables this thesis to use statistical analysis to uncover which individual-level factors affect individuals positioning on this scale relative to their political party.

3 Previous Literature

Social media has contributed to changes in political communication and campaigning. It has been argued that social media creates a better means for individual candidates to conduct their campaigns independently from their respective party (Metz, Kruikemeier and Lecheler, 2020; Pedersen and Rahat, 2021). Instead of relying on traditional media outlets for reaching out to their electorate, or their party for campaign resources, political candidates are able to interact with their electorate in a relatively cost-effective way. While some scholars have linked social media campaigns with an accelerated shift toward more individualised or personalised campaign styles and strategies (e.g. Zittel, 2009; Karlsen and Enjoras, 2016; Metz, Kruikemeier and Lecheler, 2020), there is less understanding concerning the cross-party and individual-level conditions that could explain why some parties in the same national context appear more cohesive in their social media campaigns.

Candidate and party campaigning has changed across time, which is at least partly due to changes in the campaigning tools available to them (Karlsen, Rütz and Öhberg, 2020). While in recent years authors have studied the role of social media as a campaign tool, the majority of the pre-existing studies on social media and individualised campaign styles have been conducted in the context of Western democracies (ibid.).

In his paper from 2009, Zittel argued that there are two views when it comes to online campaigning: while the orthodox view sees online communication tools solely as a new means for political parties to conduct centralised campaigns, the revisionist view sees online platforms driving more candidate-centred campaigning. Since Zittel's (2009) publication, multiple studies have discussed this relationship in the context of social media. The existing studies, however, do not provide a sufficiently comprehensive picture of the dynamics of party cohesion and individualised online campaigns styles. Firstly, since the extent of individualisation in representatives campaigns is likely to be influenced by many macro-level political factors such as party systems and electoral systems, the outcomes of case studies are hard to generalise outside of their specific national or sub-national contexts (Adam and Maier, 2010; Lisi and Santana Pereira, 2019; Karlsen, Rütz and Öhberg, 2020). Secondly, there is a lack of research on how different party-level and individual-level factors relate to individualised campaign styles and party cohesion in social media.

In their study based on survey data from 2018, Karlsen and Enjoras found that politicians with individualised (in contrast to more party-centred) campaigns styles were more likely to be active on social media. However, they also found that these individual politicians were less likely to be influential on social media. The

authors point out that their study was conducted on the Norwegian elections, and the Norwegian electoral and party system gives little room for individualised campaigning in general.

Moreover, in their study from 2013, Åström and Karlsson studied the use of online blogs as a tool for online campaigning in the context of elections in Sweden. The authors argue that while modernisation has encouraged individualised campaign communication, the cross-party differences are relatively understudied in the field. The authors find that individualised campaign styles are associated with the individualism-collectivism dimension – parties on the right of the political ideology spectrum that have more individualistic traditions are more likely, compared to parties on the left, to adopt candidate-centred campaign styles. Similarly to Karlsen and Enjoras (2018), Åström and Karlsson (2013) emphasise taking into account the political context of their study. The authors note that Sweden presents a case with a traditionally stable party system, and study the phenomenon explicitly in this context.

Karlsen, Rütz and Öhberg (2020) highlight how while the instrument of campaigning, in this case social media, does affect how likely politicians are to adopt more candidate-centred or party-centred campaigns styles, the effects of the instrument are affected by local political conditions including the electoral, party and political systems.

4 Background and Case Selection

Authors have previously been interested in party cohesion in Brazil and recognised the country for its weak and unstable party system (e.g. Samuels, 1999; Samuels, 2000; Nunes and Melo, 2015). That research, however, has mainly concerned traditional political campaigns and parliamentary politics. It has been widely understood that stable political institutions – including a party system with cohesive political parties – matter for a working democratic system (Owens, 2003; Laver, 1999). Thus, party cohesion matters for political development and studying how it interacts with new technologies used in political communication is an important topic to tackle in the fields of political science and development studies.

As mentioned, the majority of the work on the effect of social media campaigning on party cohesion has been conducted in the context of Western liberal democracies. In most of these contexts, the roots of the party system are noticeably different from that of Brazil. In Western democracies, the party system has its roots in the post-1920s class cleavages. Scholars have linked the decay of these cleavages and thus the disconnect between socio-economic structures and political parties to the

phenomenon of personalisation of politics (Karvonen, 2010: 4; Adam and Maier, 2010). Brazil's political history has followed a different path. In Latin America, and in Brazil in particular, the party system has been less stable throughout the 1900s and stems less so from the institutionalised political organisation of social cleavages (Roberts and Wibbels, 1999). Moreover, Roberts and Wibbels (1999) argue that those parties that were relatively institutionalised for most of the 1900s suffered losses in the strength of their organisational ties after the economic crisis in the 1980s and neoliberal restructuring in the 1990s. Additionally, majority of parties founded during the 1900s, with the Worker's Party (PT) as a notable exception, have kept their internal structures, which are built around legislation from the era of Brazil's military dictatorship (Nunes and Melo, 2015). Studies have found that, in general, Brazil's politics is rather individualised and political parties are internally weak (ibid.; Samuels, 1999; Lucas and Samuels, 2010). Furthermore, parties lack clear ideological differentiations and frequently change their programmatic content (Lucas and Samuels, 2010).

While the role of social media in Brazilian politics has been widely studied before (for example, Arnaudo, 2019; Brito et al., 2019; Santos, 2020 as well as others), few studies have linked it to the party system as a political institution or aimed to translate how online political campaigning styles vary between and within the country's political parties. Due to its high rate of social media use, prevalence of social media campaigns, as well as the discussed unstable party system, Brazil makes a relevant case to further the understanding of the interaction between social media campaigns and party systems.

5 Theoretical Framework

The theoretical framework of this thesis is divided into two sections according to the two research questions. The first section aims to explain cross-party differences in party cohesion through theories of party discipline and personalisation of politics applied in the context of Brazil. After this, it explains how the concept of *rhetorical cohesion* links to party cohesion and helps to study party cohesion in the context of social media campaigns. The second section presents existing theories on individual-level factors that may affect party cohesion. The two main theoretical approaches discussed concern two interconnected strands of research: party discipline and personalisation of politics. They both make comparisons between individualised versus party-centred behaviour patterns of politicians, but while the first focuses on the capacity of the political parties to enforce control over their representatives, the second focuses mainly on external factors giving more autonomy to individual candidates at the expense of political parties.

There are two main reasons for combining the theoretical work on these two strands of research. Firstly, there have been several of studies concerning party discipline in the context of Brazilian politics partly due to their exceptionally weak and unstable party system. However, these studies have been conducted mainly in the context of parliamentary politics and have not considered how their findings would translate to the context of social media. Secondly, unlike the work on party discipline, the pre-existing work on personalisation of politics has connected the phenomenon to the digitisation of political communication, and social media campaigning in particular. However, these studies are mainly situated in the context of Western democracies and therefore the theoretical work cannot be directly translated into the context of the Brazilian political or party system. In the following section, these two theoretical strands are evaluated and applied to the case of social media campaigns in Brazil.

5.1 Party Cohesion and Party Discipline

While party cohesion refers to the extent to which representatives are observed to work together in pursuance of the collective goals of their respective party, party discipline refers to party cohesion enforced by the actions of the leadership of the party (Kam, 2014). For this reason, in contrast to party cohesion, which is an empirical phenomenon, party discipline is a theoretical explanation as to why some political parties appear more cohesive than others (ibid.; Giannetti and Benoit, 2009).

Furthermore, scholars have associated both party cohesion and party discipline with strong parties and party systems. This stems from the representatives of the party being more likely to follow the party line while in power, which leads to more stable parliamentary coalitions (Bowler et al., 1999; Laver, 1999).

Comparative politics literature has provided institutional explanations for varying degrees of party discipline (Giannetti and Benoit, 2009). Giannetti and Benoit (2009) list the main institutional settings that affect party discipline and united voting patterns. Firstly, federal legislative systems have been argued to weaken party unity due to decentralised party governance structures (ibid.). Second, presidential systems have been associated with lower amounts of party discipline as the returns from voting as strong blocks are lower compared to parliamentary systems (ibid.). Thirdly, Giannetti and Benoit (2009) mention how electoral system with more possibilities for so-called “personal vote” give parties less power to control their candidates’ re-election and therefore these types of electoral systems correlate with lower levels of party discipline. This would indicate that open-list proportional representation systems would show less party discipline as opposed to closed-list systems, for example.

According to these institutional explanations, Brazil should present an obvious case of low party discipline and cohesion: it is a federation with a presidential

executive-legislative system and an open-list proportional electoral system (Lucas and Samuels, 2010; Figueiredo and Limongi, 2000). Samuels (1999), however, raises Brazil as an interesting and somewhat exceptional case. Although the Brazilian system is essentially described as a candidate-centric system with a high possibility for personal vote, the voters can also choose to cast a label vote, which gives their vote to the party instead of the candidate (*ibid.*). Regardless, it seems that Brazil presents a case with low levels of party cohesion and weak political parties in general (Samuels, 1999; Lucas and Samuels, 2010). The findings by Figueiredo and Limongi (2000) seem to contradict these findings to some extent. The authors argue that while the electoral system does encourage personal takes on election campaigns, within the legislative Chamber the level of party discipline is high due to the extent of presidential agenda-setting and party-centred legislative work. The authors did agree that their contradicting findings only hold inside the parliament. Therefore, previous studies on party discipline in Brazil point to party cohesion also being low in the context of political campaigns.

In Brazil, low levels of party cohesion are likely to have repercussions for the stability and competitive capacity of the party system in general. As mentioned, Bowler and others (1999) discuss how multiparty systems benefit from party discipline and cohesion due to more stable structures in legislative work and party competition. Both the ability to build and maintain coalitions as well as the capacity of the opposition to function as a block improve when parties are able to show coherence in voting patterns.

Although institutional theories explain cross-national differences in levels of party discipline and party cohesion, there is less understanding on variance between parties operating in the same political system (Samuels, 1999; Giannetti and Benoit, 2009). Some theories have been developed to distinguish the characteristics of parties more likely to adopt party discipline. Firstly, Bowler and others (1999) argue that smaller parties with little likelihood to achieve majority in the decision-making procedures are less likely to adopt strict party discipline due to less significant returns from voting as a block, compared to bigger parties. Secondly, party discipline has been found to vary depending on the ideological position of the party (Owens, 2003). This has also been the case in Brazil, where Mainwaring and Liñán (1997) found party discipline to be low in all parties but those on the left.

Samuels (1999) reached similar conclusions, especially in terms of the Worker's Party (PT), which according to the author has internal structures in place that encourage its members to align with the party line. More recent studies have achieved similar conclusions. In his thesis, Jucá (2022) argues that Brazil continues to present a puzzle for party discipline concerning its contradicting institutional incentives. Jucá

(2022) found recognisable levels of party discipline in the Worker’s Party (PT), but also in the clientelistic Party of the Republic, now the Liberal Party (PL). Jucá’s (2022) thesis argues that previous theories on comparative politics have ignored the influence of clientelism on party discipline and that the effects of clientelism on party discipline and party systems in general should be studied further – especially in Latin America where clientelism has long traditions in the political system.

Overall, while there is no consensus on the extent of party discipline in the Brazilian party system, the institutional explanations would point to low party discipline and low party cohesion in Brazil. Additionally, while cross-party factors are less studied compared to cross-national institutional factors, scholars have found some indications of the aspects that could explain variations in party discipline and cohesion within the same political system. These above-mentioned studies have not, however, explored how party discipline would generally be translated into the context of social media campaigns.

5.2 Personalisation of Politics

Like party discipline, personalisation of politics aims to explain the differences in the extent to which individual representatives adopt behaviour and preferences aligning with or departing from their respective political party. The literature has linked personalisation of politics to changes in the socio-technical media environment (Karvonen, 2010: 4) including the rise of social media as a tool for political communication (Krukemeier, Gattermann and Vliegthart, 2018; Metz, Kruikemeier and Lecheler, 2020; Pedersen and Rahat, 2021). The link between social media and personalisation of politics is based on the notion that social media allows political candidates to campaign independently from the resources provided by their political party (ibid.).

From a broad perspective, personalisation of politics refers to a shift or a change “[. . .] altering the balance between political individuals and collectives” (Pedersen and Rahat, 2021: 216). The phenomenon, however, is defined and approached in different ways by different authors. In their comprehensive literature review on personalisation of politics, Adam and Maier (2010) note that the personalisation of politics hypothesis is in fact based on two separate propositions. The first proposition suggests that the role and autonomy of individual representatives has increased at the expense of political parties, while the second proposition suggests a discursive shift toward non-political traits of representatives over political traits. This thesis discusses personalisation of politics with an exclusive focus on the first proposition.

Personalisation of politics is argued to stem from a disconnect between the citizens and the political system (Adam and Maier, 2010). Specifically, it is linked to the socio-economic changes that have led to the decline of the societal cleavages that

the current party system is largely based on in the majority of Western democracies (Karvonen, 2010: 4). Consequently, instead of association with political parties, citizens increasingly focus on specific issues and individual candidates, which in turn leads to more individualised campaign styles (ibid.; Gauja, 2015). Since the personalisation of politics hypothesis builds on the roots of the party system typically found in Western liberal democracies, its direct application to contexts such as Latin America and Brazil should be done with precaution. For this reason, this thesis focuses on how personalisation of politics scholarship's explanations on cross-party and individual-level differences in party cohesion overlap with studies concerning party discipline and what takeaways the scholarship can offer in terms of party cohesion in social media.

Both Karvonen (2010: 101-102) and Adam and Maier (2010) question the general validity of the personalisation of politics thesis and argue that institutional conditions should be accounted for when researching personalisation of politics. In their comparative study on campaign individualisation, Lisi and Santana Pereira (2019) study the effect of institutional factors on the extent to which candidates adopt a personalised focus in their political campaigns. The authors found that institutional conditions, most importantly the electoral system, campaign funding, access to media and the degree of internet dissemination, affect the level of personalisation. Karvonen (2010: 101-102) also found that intensified personalisation has mainly occurred in the context of candidate-centred electoral systems.

Some authors have also aimed to explain cross-party differences in personalised campaigns. In their previously mentioned study from 2013, Åström and Karlsson looked at cross-party differences in personalised campaign blogging. They argue that individualised versus more party-centred campaigns are associated with the individualism-collectivism dimension: traditionally leftist parties possess more collective and right wing parties more individualistic values. Their study, based on data from Sweden, shows how parties on the right appear more candidate-centric in their campaign blogging.

In her paper from 2015, Gauja approaches the personalisation of politics with an assumption that political parties are reactive to the contemporary developments in political culture. Thus, Gauja (2015) examines the ways in which different political parties could have adopted strategies to adapt to the personalisation of politics. She mentions online tools and platforms as a way for parties to achieve more personalised communication with their voters. Gauja (2015) goes further to suggest that large mainstream parties are likely to shift into so-called 'catch-all' parties without a clear ideological focus, whereas more niche issue focused parties are more sensitive toward their mean voter.

Personalisation of politics and party discipline are essentially two theoretical explanations of the same phenomenon: party cohesion. These theoretical explanations are not, however, mutually exclusive. In fact, their findings concerning institutional and party-level explanations seem to be overlapping. Rather, these theories differ in their explanation as to why there are differences in party cohesion. Party discipline explains levels of cohesion through parties' internal structures of enforcing cohesive behaviour, whereas personalisation of politics explains variations through external socio-technical changes.

5.3 Rhetorical Cohesion

This thesis uses the concept of rhetorical cohesion to measure the extent of observed similarities in the use of language in the political campaigns of representatives. The concept is closely linked to the concept of party cohesion. While the definition of party cohesion varies between studies to some extent (Giannetti and Benoit, 2009), in the context of this thesis it refers to the degree of similarity in behaviour, policy preferences and values held by the members of the same political party. Rhetorical cohesion, on the other hand, is defined as the extent to which party cohesion is observed through similarities in the use of language. In this case, the concept of rhetorical cohesion allows for the investigation of party cohesion in the context of social media campaigns and the textual data social media offers.

With the above theories and scholarship on party discipline and personalisation of politics in mind, these hypotheses on cross-party differences in rhetorical cohesion follow:

H1: The parties on the left of the ideological spectrum show higher levels of rhetorical cohesion in social media campaigns.

H2: The Worker's Party in Brazil appears rhetorically cohesive in terms of its representatives' social media campaigns.

5.4 Individual-level factors and rhetorical cohesion

The findings of Lisi and Santana Pereira (2019) in their study on campaign personalisation indicate some individual-level factors that could predict party cohesion. Firstly, the authors find that individual candidates with longer political careers are more likely to adopt individualised campaign styles. Secondly, they found that right-wing politicians favour individualised campaign styles promoting their personal qualities over party values. Finally, they found men are more likely than women to depart

further from the party line.

Samuels (1999) also associates individual candidates' campaign resources with the likelihood of the realisation of party discipline. He points out that in the PT party in Brazil the recruitment and promotion of candidates with fewer independent resources was common. Samuels (1999) argues that candidates' dependency on their party's campaign resources has contributed to the party's ability to institutionalise party discipline and thus remain more cohesive in comparison to other political parties in Brazil. Furthermore, Close and Gherghina (2019) present the degree to which individual preferences depart from the preferences of one's party as an important predictor for party cohesion. However, such data can be hard to retrieve without comprehensive surveys on politicians' preferences. Overall, the individual level factors, compared to institutional and cross-party factors, are relatively understudied in research concerning party cohesion (Adam and Maier, 2010).

Of these above-mentioned theories, this thesis focuses on testing the effect of the length of one's political career on party cohesion. Thus, the hypothesis follows:

H3: A longer political career is associated with more individualised rhetoric in social media campaigns.

Furthermore, this thesis is interested in the extent to which parties' presidential candidates contribute to rhetorical cohesion. Some authors find that in presidential systems, such as Brazil, political parties with presidential candidates become "presidentialised". This means that their other representatives adopt integrated strategies with their presidential candidate due to the "coattail effect" - the association with the presidential candidate will increase congressional candidates' likelihood to get elected (e.g. Borges, Albala and Burtnik, 2017; Ribeiro and Fabre, 2020). This would suggest that rather than presidential candidates aligning their campaign strategy with the party legislators, the legislators would appear to follow the strategy of the presidential candidate. Samuels (2000) argues that an aligned strategy between the president and the Congressional candidates is in fact mutually beneficial as the campaign alignment makes it easier for the executive to receive political support from the legislators after the election.

At the same time, Samuel (2000) argues that in the case of Brazil, with its weak national party labels and organisations, congressional candidates have little incentive to line up behind the presidential candidates. Instead, Samuels (2000) says that gubernatorial backing is more important for federal legislators and candidates in Brazil. In their more recent study based on individual-level survey data, Borges and Lloyd (2016) found that presidential coattails were in fact detectable in the context

of Brazilian elections. However, with country’s weak and unstable party system in mind, this thesis suggests the following hypothesis concerning presidential candidates and rhetorical cohesion on social media:

H4: Presidential candidates do not appear more rhetorically cohesive with their party compared to the representatives in the Chamber of Deputies.

6 Data

6.1 Twitter Data

The primary data source is tweets written by Brazilian politicians extracted from the social media platform Twitter. The initial dataset includes the accounts of 136 Brazilian representatives and 12 presidential pre-candidates for the upcoming election in October 2022. The tweets were collected using Twitter’s streaming API. The Python library Tweepy was used to make requests to the Twitter API. In total, the dataset includes 11 589 tweets, of which the Chamber of Deputies accounts for 10 177 and the presidential candidates 1412. The tweets were collected in 2-4 day intervals due to access restrictions in the Twitter API. As the Twitter API allows access to tweets from the last seven days, the tweets were collected regularly during a period of five weeks.

On average, the Deputies tweeted 74 times over the collection period, whereas the presidential candidates tweeted 118 times on average. There were quite significant variations in the Twitter activity across the politicians in the dataset. While some politicians did not tweet at all, the most active representative tweeted 1127 times during the collection period. All political parties that currently have seats in the Chamber of Deputies were included in the initial dataset.

Due to a lack of Twitter activity for some of the politicians in the sample, the sample was reduced to 124 politicians prior to running the scaling algorithm. After running the algorithm, the parties with less than five observations were removed from the dataset to enable proper statistical analysis of the data. The final amount of politicians in the set was thus 103, of which 5 were presidential candidates. This meant that the final dataset consisted of 9738 tweets across 10 parties of which 5 have presidential candidates active on Twitter.

All the parties with more than 20 seats in the Chamber of Deputies are represented in the final dataset. Additionally, one party (NOVO) with less than 20 representatives in the Chamber is included in the final dataset. Interestingly, seven of the eight sitting politicians of NOVO had their Twitter accounts available on the

website of the Chamber of Deputies. Additionally, their presidential candidate, who is not a member of the Congress, is included in the dataset. Table 1 presents the final numbers regarding the representation of the Brazilian political parties in the dataset. The number of seats in the Chamber are retrieved from the Congress website

Table 1: Sample

| Party | Nr of seats | Nr in dataset | Tweets |
|--------------|-------------|---------------|--------|
| PL | 77 | 20 | 2734 |
| PT | 57 | 17 | 3642 |
| PP | 55 | 8 | 121 |
| UNIÃO | 54 | 12 | 870 |
| PSD | 47 | 10 | 367 |
| REPUBLICANOS | 34 | 11 | 368 |
| MDB | 37 | 6 | 336 |
| PSB | 23 | 6 | 281 |
| PSDB | 22 | 5 | 350 |
| NOVO | 8 | 8 | 669 |

6.2 Sample

The sample of 98 (137 initially) representatives from the Chamber of Deputies, which in total consists of 513 Deputies, represents what can be called a convenience sample. The representatives included in the sample had their Twitter usernames available in the official website for the Brazilian Congress. This poses the potentiality of biases in the dataset, as it is possible that certain characteristics of representatives would make them more likely to publish their Twitter handles publicly on the Congress website. Hypothesising this, however, is out of the scope of this thesis and rather poses a possible question for future research.

However, the sample seems to somewhat resemble the characteristics of the Chamber of Deputies in terms of its social characteristics. The average age of the representatives in the dataset is 49.67, which is quite close to the average age (51.58) of the Chamber (Inter-Parliamentary Union, 2022). Unfortunately, the gender division in the dataset is rather male-dominated with only 20 of the representatives included being female. This, however, is a significantly higher proportion of women ($\approx 19\%$), compared to that of the Chamber of Deputies itself ($\approx 15\%$) (Inter-Parliamentary

Union, 2022).

Additionally, the 5 (12 initially) accounts for presidential pre-candidates were manually searched and added to the database.

6.3 Ethical Considerations

Social media and Big Data have made an unforeseeable quantity of social data accessible for research in different fields of sciences over the past decade or so (Woodfield and Iphofen, 2017). At the same time, the ethical considerations of using social media data for research has been a topic under ongoing debate (ibid.). The debate has mainly concerned informed consent and anonymity (Williams et al., 2017). While some scholars see social media data as public information, others, such as Williams and others (2017), argue that this view does not appropriately take users views and perspectives into consideration.

In the context of this study, the profiles of the users studied are not considered to represent private individuals. Firstly, because their political career establishes them as public representatives and, secondly, because it can be assumed that they made an informed decision to allow their Twitter handle to be available on the Congress website. Therefore, while limiting the dataset to those representatives with Twitter handles available on the site poses some analytical limitations, it also helps to minimise the ethical risks that could arise from the use of the representatives' social media data. While some of the presidential candidates did not have their information listed on the Congress website, their accounts were confirmed to represent them as political candidates rather than private individuals.

7 Methodology

This thesis uses a quantitative research design to answer the research questions. Specifically, it uses a combination of computational data analysis with a machine-learning NLP (Natural Language Processing) algorithm and statistical analysis with multiple linear regression. The programming language R was used to conduct both stages of the data analysis. This two-stage methodology is explained here. The first sub-section describes how the NLP algorithm Wordfish is used to scale the representatives' Twitter timelines on a one-dimensional scale, which is consequently used to observe levels of rhetorical cohesion within the different political parties. The second section explains how multiple variable regression is used to model individual-level political factors' effects on rhetorical cohesion.

7.1 Wordfish

Wordfish is an unsupervised NLP algorithm that scales political texts based on word frequencies in a given dataset (Slapin and Proksch, 2008). As the algorithm is unsupervised, it does not require any training with training data and thus, it is relatively time-efficient and can be used in any language (Huneus, 2021). Wordfish, along with other computational scaling methods, is commonly used in political science research to scale political ideology (Wilkerson and Casas, 2017; Aydogan, Tuna and Yildirim, 2019). However, as mentioned Wordfish does not require any training data and therefore creates the scale it uses based on the textual input data it is given. Wordfish takes a word-document matrix as an input and scales the political texts based on the similarity of their word choices (Aydogan, Tuna and Yildirim, 2019). Specifically, it places the text documents into a unidimensional scale based on how similarly they use certain key words (Huneus, 2021). The algorithm model assumes that the frequency of word use follows a Poisson distribution (ibid.). The Poisson distribution is a probabilistic distribution based on a single parameter λ (ibid.). For each individual word j and a politician i , it is assumed that there will be such a distribution (λ_{ij}). The distribution is the following:

$$\text{Wordcount}_{ij} \sim \text{Poisson}(\lambda_{ij})$$

where

$$\lambda_{ij} = \exp(\alpha_i + \psi_j + \beta_j \times \omega_i)$$

The model estimates the parameter λ_{ij} based on two parameters for both document i and word j . The model assigns two parameters for each document i : position ω (which reflects the underlining position of the document, or in this case the politician) and fixed α (which essentially captures the document length). And two parameters for each word j : fixed ψ (which captures the word frequency) and word weight β (which captures the importance of the word in differentiating the document position).

The program computes its estimates the following way (Slapin and Proksch, 2008): after setting some initial values for the parameters of either the document or words, the values for the other are estimated in the next stage. After this, the algorithm alternates between the two sets of parameters until it finds that the gap between consecutive estimates is very small. Once this is achieved, the algorithm reports the results.

Wordfish takes on two underlying assumptions (Huneus, 2021). First, it as-

sumes that words always carry the same meaning. Although it is clear that this is not the case, the algorithm performs relatively well regardless of not being able to distinguish between different meanings or contexts of specific words (Huang, 2017). Second, it assumes that texts are ordered by a latent dimension X , which reflects the ideological differentiation of documents. Huneeus (2021) notes that the validity of this assumption reflects the extent to which the method is robust when compared to other methods, and how well the documents included in the analysis are representative of political ideology. In the context of this thesis however, the dimension is not expected to reflect political ideology per se, but rather rhetorical differences between the politicians included in the dataset. To some extent, this presents an advantage, as there is a validity risk in assuming that a certain ideological scale mirrors the word choices of politicians.

In fact, Aydoğan, Tuna and Yildirim (2019) mentioned how the unsupervised nature of Wordfish resembles roll call vote analysis techniques, which makes it a particularly interesting method to analyse party cohesion in terms of social media campaigning.

Before running the algorithm, the data was pre-processed. First, the politicians who did not tweet during the period of data collection were removed from the set. After this, the data was cleaned, which included removing all special characters, username-mentions and url's. Additionally the words were normalised by making all letters lowercase and removing accents from Portuguese letters. After this, the existing data was converted into a corpus format and then to Document-feature Matrix format (DFM).

Finally, numbers, symbols, punctuation and stopwords were also removed from the data. Huneeus (2021) also recommends the removal of very uncommon and common words by removing words with a frequency less than or equal to the 5th percentile and those with a frequency greater than or equal to the 95th percentile. Huneeus (2021) argues that these words at the extremes of the frequency distribution can bias the algorithm. This is not, however, common practise when using Wordfish for data analysis and, in this case, the words at the extreme ends of the distribution were considered likely to be important in distinguishing some politician's rhetoric from their party line. Furthermore, this study cares less about the reliability of the algorithm's ideological scale and rather focuses on scaling rhetorical differences. After the data was pre-processed, the data was passed to the algorithm and the resulting theta values (equivalent of the ω - the document position parameter) of each document were saved to the dataset.

Figure 1 visualises the parameters assigned to the words in the dataset. Examining the weights and frequencies of specific words helps to understand how the

scale represent, but rather it is intrigued by how differences in speech and vocabulary differentiate politicians rhetorically.

As mentioned, this thesis investigates rhetorical cohesion of political parties as defined by the party cohesion in terms of similarities in the use of language in social media campaigns. The distance between individual politicians’ Wordfish theta scores and the mean score of their respective party is used as a measurement of rhetorical cohesion. In this first part of the data analysis, which analyses differences in rhetorical cohesion at the party-level, the distribution of politicians’ scores within each political party are compared using descriptive statistics and data visualisation techniques in order to answer the first research question.

7.2 Statistical Analysis

To investigate the relationship between individual-level political factors and rhetorical cohesion, this thesis conducts statistical analysis in the form of multiple linear regression (MLR). In short, to answer the second research question, an MLR model is used to evaluate the correlation between the dependent variable (rhetorical cohesion), and independent variables (individual-level political factors). Rhetorical cohesion of individual candidates with respect to their political party is measured by calculating the distance between individual politicians’ Wordfish theta scores and the mean score of the political party they represent. A bigger value for the dependent variable thus indicates a lower level of rhetorical cohesion with one’s party and a more individualised campaign style, whereas a lower value for the dependent variable indicates that the politician’s rhetoric is closer to the party line. The different variables and their types are listed in Table 2.

Table 2: Variables

| Type | Name | Label | Classification |
|-------------|------------------|---------------------------------|----------------|
| Dependent | party_distance | Theta distance from party mean | Continuous |
| Independent | years_in_office | Years in political office | Discrete |
| Independent | president | If running for presidency | Binomial |
| Control | party | Memberin which political party | Categorical |
| Control | gender | Dummy variable for gender | Binomial |
| Control | number_of_tweets | Number of tweets in the dataset | Discrete |

The model has two independent variables, which measure political individual-

level factors that could influence the independent variable: the length of political career and running for presidency. The length of one’s political career is measured as years spent in federal political office, whereas whether a politician is running for presidency is coded as a dummy variable (1 indicates presidential candidate and 0 not).

In addition, there are three control variables: political party, gender and the number of tweets in the dataset. To control the possible effect of a political party on the dependent variable, political party is added as a control variable. This categorical variable is converted into a factor variable. The control variable for gender is coded into the dataset as a dummy variable so that 1 represents female gender and 0 male gender. In their study, Karlsen and Enjoras (2016) found a positive relationship between Twitter activity and individualised campaign styles. For this reason, there is a possibility that those politicians that demonstrated higher Twitter activity could show bigger divergence from the party average in terms of their Wordfish scores. For this reason, a control variable for the number of tweets the politician tweeted during the collection period is added to the model to measure Twitter activity.

The MLR model follows:

$$y = \alpha + \beta_1x_1 + \beta_2x_2 + \beta_3x_3 + \beta_4x_4 + \beta_5x_5$$

where y represents the dependent variable (distance between theta and party mean) and α represents the intercept. The coefficients of the two independent variables are β_1 and β_2 , whereas x_1 represents the independent variable (length of political career) and x_2 is the second independent variable (presidential candidacy). Control variables are represented as x_3 x_4 , x_5 with their coefficients β_3 , β_4 and β_5 respectively.

First, the model is tested without including any control variables:

$$y = \alpha + \beta_1x_1 + \beta_2x_2$$

After this, the control variables are included to the equation one by one to see what effects they may have on the coefficients of the independent variables, their significance, and the significance of the model.

7.3 Limitations

There are a few limitations with the methodology. Firstly, the sample size (103) could affect the extent to which the statistical analysis could produce statistically significant results. However, it is worth noting that although the sample size seems rather small, multiple linear regression does not account for the relatively small population size

(525) with respect to the sample size.

Secondly, as mentioned, Wordfish bases its model on frequency of certain words that it considers important for the position of political texts. However, it assumes that each individual word carries the same meaning in each different context. This causes some limitations to the extent to which Wordfish can be used to measure rhetorical cohesion. The algorithm has been found to perform relatively well in scaling political ideology regardless (Huang, 2017). Furthermore, as this study only focuses on the scale in terms of differences in language, it can be assumed that using different words, regardless of the context, points to some level of difference in the politicians' rhetoric.

Although the frequency and differentiation of used words gives a relatively reliable estimate of the differences in the use of language between politicians, different rhetoric styles and strategies can be differentiated in other ways than merely in the form of vocabulary choice. Here, however, rhetorical cohesion is only measured to the extent of the results produced by the Wordfish algorithm, and therefore the scope is limited to the frequency of using specific words. It is acknowledged that other methodologies, such as varieties of qualitative methods as well as more complex NLP techniques, would be able to adopt more nuances in their analysis of rhetorical differences and measures of rhetorical cohesion. The trade-offs between quantitative and qualitative methodologies are well known and although the research design in this thesis allows a relatively transparent analysis on a larger set of social media data, its findings would be strengthened if complemented with qualitative methods in future research.

Finally, as the main language of Brazil, and therefore the language of the data, is Portuguese, it is important to acknowledge that the author of this thesis is not fluent in Portuguese. However, as Wordfish is an unsupervised algorithm it does not require any input or training data that would have to be analysed or categorised manually. Furthermore, the output of the algorithm provides the theta values for the politicians based on the β parameters the algorithm has calculated for the used words. To ensure the validity of the results, the words with the highest impact on the document positions were validated. As it is only the values the algorithm assigned for single words that affect the positioning of documents rather than larger components of text, it is relatively easy for the author to get help from a reliable dictionary to translate these words.

8 Results

8.1 Party-level Analysis

The descriptive statistics of the Wordfish thetas for each political party are presented in Table 3. When interpreting the results, it is important to note that the complete scale, from the politician with the lowest to the politician with the highest theta value, ranges between -2.248 and 3.506. With this in mind, it appears that all parties have a relatively high range of theta values given the rather large standard deviations.

Furthermore, it can be observed that none of the parties average on the extreme ends of the scale produced by Wordfish. In the case of MDB, we can see quite a large difference between the mean (0.43) and the median (-0.2) indicating that the distribution is skewed. This most likely has to do with one outlier ending up at the extreme positive end of the scale (visible in Figure 2) and the party having a relatively low number of observations (6). This most likely explains the relatively high standard error of the party.

Table 3: Wordfish scores within Political Parties

| Party | n | mean | sd | median | se |
|--------------|----|-------|------|--------|------|
| MDB | 6 | 0.43 | 1.66 | -0.2 | 0.68 |
| NOVO | 8 | 0.35 | 0.4 | 0.31 | 0.14 |
| PL | 20 | 0.45 | 0.96 | 0.65 | 0.22 |
| PP | 8 | -0.29 | 0.77 | -0.32 | 0.27 |
| PSB | 6 | -0.32 | 0.78 | -0.5 | 0.32 |
| PSD | 10 | -0.25 | 1.09 | -0.58 | 0.34 |
| PSDB | 5 | 0.59 | 1.21 | 0.26 | 0.54 |
| PT | 17 | 0.04 | 1.0 | 0.12 | 0.24 |
| REPUBLICANOS | 11 | -0.35 | 1.14 | -0.41 | 0.34 |
| UNIÃO | 12 | -0.16 | 1.09 | -0.14 | 0.31 |

As the dataset includes quite a few outliers in different parties, the box plot in Figure 2 provides a bit clearer description of the distribution of Wordfish thetas within the different political parties. The figure shows how parties such as MDB, PL, PSD, PSDB, Republicanos and União all appear to be scattered over a relatively wide space on the y-axis, while only NOVO, PP, PSB and PT appear somewhat cohesive, excluding the few outliers. Of the largest parties – defined by number of seats in the

Chamber – PT and PP seem to stand out with their cohesiveness.

As noted above already, the line denoting the median score of MDB shows that the documents on the negative end of the theta axis appear much more clustered than those on the positive end of it. Similar observations appear in PSD and in PSDB, to some extent. Additionally, we can see that MDB has one very far outlier in their distribution, in the far positive end of the division. After investigating what this persons’ Twitter timeline looks like to find out what would explain this big of a divergence from the rest of the party, it appeared that they almost solely discussed topics related to the COVID-19 in their profile.

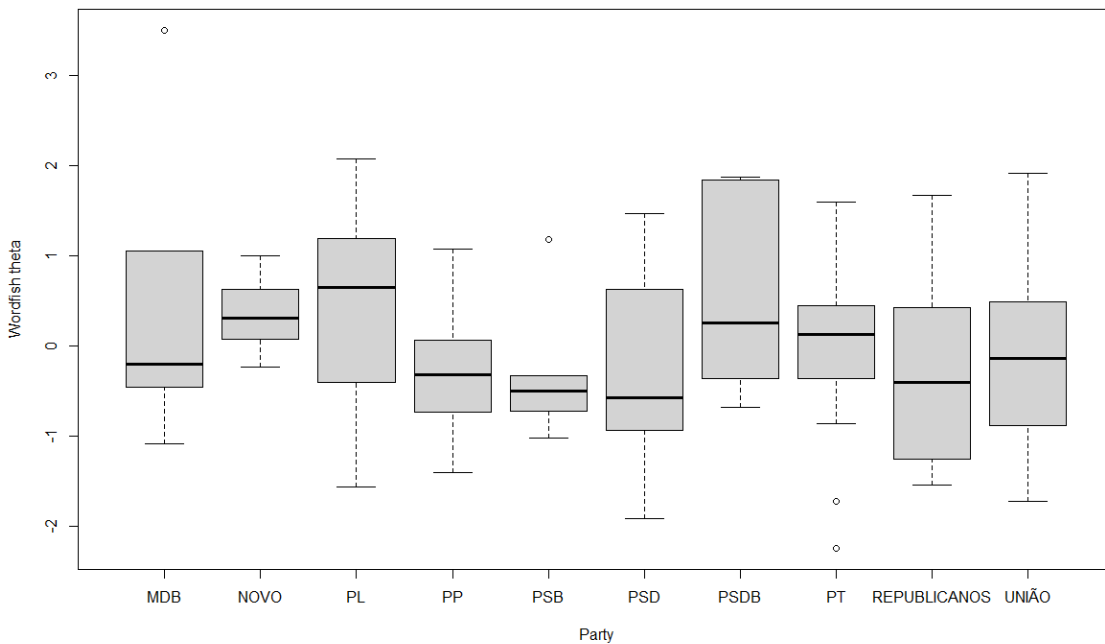


Figure 2: Box plot visualising the rhetorical cohesion of Twitter campaigns within different political parties

Of all the parties in the sample, NOVO and PSB appear the most cohesive in terms of their representatives’ social media campaigns, which is also notable in terms of the standard deviation of their theta scores (Table 3). It is important to note however, that for some parties, the amount of observations reflects only a proportion of representatives included in the sample. This is the lowest for PP with only eight of their 55 sitting members represented in the dataset.

8.2 Individual-level Analysis

Table 4 shows the output of the model that measures the relationship between the political individual-level independent variables and rhetorical cohesion. Specifically, the independent variables included in the equation represent the length of the politi-

cian's political career (measured in years in political office) and whether the politician is a pre-candidate for presidency. The dependent variable is measured as the distance between the politicians' Wordfish theta and the mean theta of their respective political party. As specified in the note below the table, values marked with 2 or more asterisks (**) have p-values below 0.05, and constitute statistically significant results for the purposes of this thesis.

The regression output of the first model (1), which does not account for control variables, is statistically significant with p-value <0.05 . The R-squared value of 0.06 would indicate that this model would explain 6% of the variation in the dependent variable. Additionally, the coefficient for the independent variable for presidential candidacy is statistically significant with a negative coefficient of -0.553. This would imply that running for presidency would have a negative affect on the distance of one's Wordfish theta from the party mean and therefore a positive effect on rhetorical cohesion.

The other independent variable, measuring length of political career, has a significance level (0.054) just slightly over the significance level 0.05. However, as the sample size in this case is not very large, adding control variables reveals whether the effect of the variable on the dependent variable remains stable (i.e. whether the value of the coefficient remains the similar).

The regression output with political party added as a control variable is outlined in the second model (2) present in the table. The coefficients and the standard errors for each political party are listed in the table. The output shows that while the model itself is no longer statistically significant (most probably due to multiple non-significant predictors), whether a politician is running for president still seems to be a negative predictor of their theta score's distance from the party mean i.e. it has a positive correlation with rhetorical cohesion with their respective political party. Additionally, the effect of the second independent variable, measuring length of political career, remains stable. However, the p-value has increased notably above the significance level (0.092).

While some of the parties do not appear to be significant predictors of the dependent variable, the parties that appeared the most cohesive when observing the distribution of Wordfish scores at a party level (Figure 2) have p-values lower than 0.05. This observation is logical: representatives of a party with more cohesive or clustered theta scores are more likely to score closer to the mean of their party. Whereas for the other parties, the representatives' distances from the mean vary more and thus such a prediction would be less accurate.

After this, a control variable for gender is added to the model (3). It is observed that gender is not a significant predictor of rhetorical cohesion with a p-value 0.536.

Table 4: Regression Results

| | <i>Dependent variable:</i> | | | |
|-------------------------|--------------------------------|----------------------|----------------------|---------------------|
| | Theta distance from party mean | | | |
| | (1) | (2) | (3) | (4) |
| years_in_office | 0.021* (0.011) | 0.021* (0.012) | 0.020 (0.012) | 0.019 (0.012) |
| president | -0.553** (0.278) | -0.591** (0.281) | -0.581** (0.283) | -0.595** (0.284) |
| partyNOVO | | -0.818** (0.320) | -0.828** (0.321) | -0.840** (0.323) |
| partyPL | | -0.438 (0.272) | -0.436 (0.273) | -0.465* (0.277) |
| partyPP | | -0.743** (0.317) | -0.737** (0.319) | -0.725** (0.320) |
| partyPSB | | -0.734** (0.341) | -0.754** (0.343) | -0.758** (0.344) |
| partyPSD | | -0.325 (0.304) | -0.335 (0.306) | -0.332 (0.307) |
| partyPSDB | | -0.267 (0.354) | -0.245 (0.357) | -0.249 (0.358) |
| partyPT | | -0.642** (0.279) | -0.632** (0.280) | -0.684** (0.290) |
| partyREPUBLICANOS | | -0.388 (0.297) | -0.395 (0.299) | -0.392 (0.299) |
| partyUNIÃO | | -0.359 (0.298) | -0.363 (0.299) | -0.372 (0.300) |
| gender_dummy | | | -0.097 (0.156) | -0.115 (0.159) |
| number_of_tweets | | | | 0.0003 (0.0005) |
| Constant | 0.670*** (0.086) | 1.158*** (0.257) | 1.181*** (0.261) | 1.170*** (0.262) |
| Observations | 103 | 103 | 103 | 103 |
| R ² | 0.060 | 0.180 | 0.183 | 0.188 |
| Adjusted R ² | 0.041 | 0.080 | 0.074 | 0.069 |
| Residual Std. Error | 0.591 (df = 100) | 0.579 (df = 91) | 0.581 (df = 90) | 0.583 (df = 89) |
| F Statistic | 3.192** (df = 2; 100) | 1.811* (df = 11; 91) | 1.681* (df = 12; 90) | 1.584 (df = 13; 89) |

Note:

*p<0.1; **p<0.05; ***p<0.01

The independent variable distinguishing presidential candidates remains significant ($p=0.043$) in this model, and its effect remains stable. Furthermore, the same control variables for political parties remain significant after controlling for the effect of gender on the dependent variable.

In the final model (4), the number of tweets per politician is included as a control variable. It can be observed that this control variable does not appear as a statistically significant predictor of rhetorical cohesion - it has a significance level of $0.471 > 0.05$ and a relatively low but positive coefficient. Again, the coefficients for both independent variables remain relatively stable as well as the significance level for presidential candidacy.

Furthermore, the controls for political parties show very stable coefficients in all the models in which they are included. This means that their effect does not seem to be strongly influenced by the effect of gender or Twitter activity. Interestingly, this observation seems to hold for all parties in the sample.

Additionally, we can see that the R-squared value of the final model is 0.188, indicating that the model explains $\approx 19\%$ of the variation in the dependent variable. The focus here, however, is not to focus on the predictive accuracy of the model as a whole but rather the reliability of the independent variables' effect on the dependent variable.

In sum, it can be concluded with confidence that presidential candidacy has a positive relationship with rhetorical cohesion with ones respective political party. In other words, presidential candidates are likely to rhetorically align with the rest of their party in their social media campaigning. Moreover, the other independent variable measuring length of political career lost significance when controlling the effects of gender, party and Twitter activity. This indicates that years in political office is not a reliable predictor of rhetorical cohesion. However, due to the small sample size and the relatively stable coefficient value of this independent variable in the different models, it is possible that the variable could still be negatively correlated (because the value of the coefficient is positive) with rhetorical cohesion. Regardless, this requires further research.

9 Discussion and Conclusion

Linking the results to the theoretical discussion delivers mixed conclusions. The results found here link relatively well with the conclusions of previous studies on the party cohesion and party discipline in Brazil. At the same time, the results on the cross-party and individual level factors affecting party cohesion in general are somewhat contrasting to the relationships found in the pre-existing research.

Regarding the cross-party differences, it is clear that PT keeps showing levels of cohesion unprecedented in the other large parties such as PL, UNIÃO and PSD, or among older parties such as PSDB and MDB (Nunes and Melo, 2015; Leme and Healy, 2017). Surprisingly, the right-wing Progressistas (PP) also appear rather cohesive in social media. This particular finding, however, needs to be taken with precaution due to the relatively small sample size of party representatives from PP.

Overall, it seems that the parties that appeared most rhetorically cohesive in their social media campaigns are located at both left and right ends of the political spectrum (for ideological identification of political parties in Brazil, see for example: Passarelli, 2015). This leads to the rejection of Hypothesis 1. Furthermore, this finding challenges the pre-existing theories on party discipline and personalisation of politics that argued party cohesion to be higher in parties with left-wing ideologies. On the other hand, this study does confirm the findings on the cohesiveness of the Worker's Party (PT) from previous research, and brings them into the context of social media campaigns, which indicates that the second hypothesis can be confirmed.

Moreover, the low levels of cohesion manifested by PL seem to point out that the findings of Jucá (2022), who connected the party's clientelistic practises to party discipline, do not translate to social media campaigns on Twitter. Furthermore, of the included parties, NOVO, with the lowest number of sitting Deputies in the sample, appeared to be the most cohesive party on Twitter. However, with the low representation of "smaller" parties in the dataset, more research is needed to make conclusions about the relationship between rhetorical cohesion and party size in Brazil.

Concerning the individual-level factors' relationship with rhetorical cohesion of parties in Brazil, the findings of this thesis both support and challenge theories from the party discipline and personalisation of politics literature. Concerning the third hypothesis, the results on whether a longer political career has a negative relationship with party cohesion should be considered tentative and do require more research. However, it seems that if the low statistical significance in the regression analysis stems from the small sample size, the third hypothesis could be confirmed and length of political career would seem to have a negative relationship with rhetorical cohesion on Twitter.

Furthermore, this thesis found strong support that presidential candidates are likely to appear more rhetorically cohesive with their respective political party than the (other) members of the Chamber of Deputies. This leads to the rejection of the fourth hypothesis. Moreover, this finding contributes to the discussion on the role of presidential candidates within political parties in Brazil. Establishing this relationship, however, does not tell whether the candidate is likely to exercise control over the party or the party over the candidate. The literature on Brazil and the

dynamics between its political parties and their legislative and presidential candidates is yet unable to fully explain this relationship. More research on this relationship as well as its possible cross-party variations is required. Additionally, this finding brings about questions concerning other dynamics that may be at play. For example, if certain characteristics of Deputies, such as constituency, affect the extent of rhetorical cohesion with their party's presidential candidate.

Overall, this thesis has helped to uncover how previous theoretical literature on party discipline and personalisation of politics could help to explain cross-party and individual-level differences in rhetorical cohesion in social media campaigns in the context of Brazil. It is clear that social media has changed political communication and given individual candidates more autonomy over their political campaigns. While scholars have previously studied party discipline in the context of Brazil, these studies have not yet investigated party cohesion in the context of social media. At the same time, the literature on personalisation of politics has touched upon individualised campaigns in the context of social media, although these studies have been conducted mainly in the context of Western democracies.

This thesis found that the theories from pre-existing research on personalisation of politics cannot be directly applied to the context of Brazilian social media campaigns. This could be due to its relatively weak and unstable party system. However, it seems that some of the findings concerning party discipline in Brazil, particularly the cohesiveness of the Worker's Party (PT), also apply in the context of Twitter campaigns. This thesis found that the PSB, NOVO and PP parties also appeared relatively cohesive in the context of social media, which presents a possible difference concerning party cohesion in social media when compared to the findings of studies in parliamentary contexts. This presents a future research avenue for comparative studies.

Furthermore, this thesis has put forward possible individual-level political factors that could affect rhetorical cohesion in social media in Brazil. Firstly, it found that the length of one's political career could imply more individualised rhetoric in social media campaigns, as has been suggested in personalisation of politics literature. Secondly, it proposes that, in the context of Twitter campaigns, presidential candidates appear more cohesive with their political parties than what some previous studies have suggested.

Finally, this thesis has explored the use of an unsupervised scaling algorithm in measuring party cohesion, in the form of rhetorical cohesion, in social media campaigns. While taking the limitations of the thesis into consideration, the methodology has proven to be effective in measuring party cohesion based on textual data from Twitter. Further research should therefore continue to explore similar methodologies,

as such methods have the potential to explore the dynamics of party systems in the context of the new digital communication environment.

10 Bibliography

- Adam, S. and Maier, M. (2010). Personalization of Politics A Critical Review and Agenda for Research, *Annals of the International Communication Association*, 34(1), pp. 213-257, doi: 10.1080/23808985.2010.11679101.
- Arnaudo, D. (2019). Brazil: Political Bot Intervention During Pivotal Events. In: S. C. Woolley and P. N. Howard (ed.) (2019). *Computational Propaganda: Political Parties, Politicians and Political Manipulation on Social Media*. New York: Oxford University Press.
- Aydogan, A., Tuna, T. and Yildirim A. K. (2019). Ideological Congruence and Social Media Text as Data, *Representation*, 55(2), pp. 159-178, doi: 10.1080/00344893.2019.1592012.
- Borges, A., Albala, A., and Burtnik, L. (2017). Pathways to Nationalization in Multilevel Presidential Systems: Accounting for Party Strategies in Brazil and Argentina, *The Journal of Federalism*, 47(4), pp. 648–672. <https://doi.org/10.1093/publius/pjx024>.
- Borges, A. and Lloyd, R. (2016). Presidential coattails and electoral coordination in multilevel elections: Comparative lessons from Brazil, *Electoral Studies*, 43(2016), pp. 104-114. <http://dx.doi.org/10.1016/j.electstud.2016.05.009>.
- Bowler, S., Farrell, D. M. and Katz S. R. (1999). Party Cohesion, Party Discipline, and Parliaments. In: S. Bowler, D.M. Farrell and S.R. Katz (1999). *Party Discipline and Parliamentary Government*. Columbus: Ohio State University Press.
- Brito, K., Paula, N., Fernandes, M. and Meira, S., (2019). Social Media and Presidential Campaigns – Preliminary Results of the 2018 Brazilian Presidential Election, *Proceedings of the 20th Annual International Conference on Digital Government Research*. <https://doi.org/10.1145/3325112.3325252>.
- Close, C. and Gherghina, S. (2019). Rethinking intra-party cohesion: Towards a conceptual and analytical framework, *Party Politics*, 25(5), pp. 652–663. doi: 10.1177/1354068819836044.
- Figueiredo, A. C., and Limongi, F. (2000). Presidential Power, Legislative Organization, and Party Behavior in Brazil. *Comparative Politics*, 32(2), pp. 151–170. <https://doi.org/10.2307/422395>.
- Gauja, A. (2015). The Individualisation of Party Politics: The Impact of Changing Internal Decision-Making Processes on Policy Development and Citizen Engagement, *The British Journal of Politics and International Relations*, 17(1), pp. 89–105. doi: 10.1111/1467-856X.12035.
- Giannetti, D. and Benoit, K. (2009) Intra-party politics and coalition governments

- in parliamentary democracies. In: D. Giannetti and K. Benoit. ed. (2009). *Intra-party politics and Coalition governments*. Oxon: Routledge.
- Huang, L. (2017) *Use WORDFISH for Ideological Scaling: Unsupervised Learning of Textual Data Part I*. [online] Available at: <https://sites.temple.edu/tudsc/2017/11/09/use-wordfish-for-ideological-scaling/> [Accessed 2nd May 2022].
- Huneus, S. (2021). Quantitative Analysis of Political Texts. In: F. Urdinez, and A. Cruz (ed.) (2021). *R for Political Data Science: A Practical Guide*. Chapman and Hall/CRC. Available through Lund University Library website: <https://lubcat.lub.lu.se/> [Accessed 9th May 2022].
- Inter-Parliamentary Union (2022). *Global data on national parliaments: Brazil*. [online] Available at: https://data.ipu.org/content/brazil?chamber_id=13349 [Accessed 25th May 2022].
- Júca, I. C. (2022). *Explaining the Puzzle of Weak Parties with Strong Legislative Leaders in Brazil: Towards a Theory of Clientelistic Party Discipline*, Ph. D. University of Minnesota.
- Kam, C. (2014). Party Discipline. In: S. Martin, T. Saalfeld, and K. W. Strøm (ed.) (2014). *The Oxford Handbook of Legislative Studies*. [online] Available at: <https://www.oxfordhandbooks.com/view/10.1093/oxfordhb/9780199653010.001.0001/oxfordhb-9780199653010> [Accessed: 10th May 2022].
- Karlsen, R. and Enjolras, B. (2016) Styles of Social Media Campaigning and Influence in a Hybrid Political Communication System: Linking Candidate Survey Data with Twitter Data, *The International Journal of Press/Politics*, 21(3), pp. 338–357. doi: 10.1177/1940161216645335.
- Karlsen, R., Rütz, G. and Öhberg, P. (2020). Candidate Campaigns in Comparative Perspective. In: L. De Winter, R. Karlsen, H. Schmitt (ed.) (2020). *Parliamentary Candidates Between Voters and Parties: A Comparative Perspective*. Oxon: Routledge.
- Karvonen, L. (2010). *Personalisation of politics: a study of parliamentary democracies*. Colchester: ECPR Press.
- Kruikemeier, S., Gattermann, K. and Vliegenthart, R. (2018). Understanding the dynamics of politicians’ visibility in traditional and social media, *The Information Society*, 34(4), pp. 215-228, doi: 10.1080/01972243.2018.1463334.
- Laver, M. (1999). Divided Parties, Divided Government, *Legislative Studies Quarterly*, 24(1), pp. 5–29. <https://doi.org/10.2307/440298>.
- Leme, L. and Healy, M. (2017). *Explainer: Brazil’s Political Parties to Watch in 2018*. [online] Available at: <https://www.as-coa.org/articles/explainer-brazils-political-parties-watch-2018> [Accessed 24th May 2022].

- Lisi, M. and Santana Pereira, J. (2020). Campaign individualization in a comparative perspective: does the context matter? *História Revista da Faculdade de Letras da Universidade do Porto*, 9(2), pp. 104-1028. doi: 10.21747/0871164X/hist9_2a6.
- Lucas, K. and Samuels, D. (2010). The Ideological “Coherence” of the Brazilian Party System, 1990-2009, *Journal of Politics in Latin America*, 2(3), pp. 39–69. doi: 10.1177/1866802X1000200302.
- Mainwaring, S., & Liñán, A. P. (1997). Party Discipline in the Brazilian Constitutional Congress. *Legislative Studies Quarterly*, 22(4), pp. 453–483. <https://doi.org/10.2307/440339>.
- Metz, M., Kruikemeier S. & Lecheler, S. (2020). Personalization of politics on Facebook: examining the content and effects of professional, emotional and private self-personalization, *Information, Communication & Society*, 23(10), pp. 1481-1498. <https://doi.org/10.1080/1369118X.2019.1581244>.
- Nunes, F. and Melo, C. (2015). Political Parties: The Case of Brazil. In: G. Passarelli, ed. (2015). *The Presidentialization of Political Parties: Organizations, Institutions and Leaders*. London: Palgrave Macmillan.
- Owens, J. E. (2003) Part 1: Cohesion, *The Journal of Legislative Studies*, 9(4), pp. 12-40. doi: 10.1080/1357233042000306236.
- Passarelli, G. ed. (2015). *The Presidentialization of Political Parties: Organizations, Institutions and Leaders*. London: Palgrave Macmillan.
- Pedersen, H. H. and Rahat, G. (2021) Political personalization and personalized politics within and beyond the behavioural arena, *Party Politics*, 27(2), pp. 211–219. doi: 10.1177/1354068819855712.
- Ribeiro, P. F. and Fabre E. (2020). Multilevel party organizations in a fragmented presidential system: The case of Brazil, *Regional & Federal Studies*, 30(4), pp. 525-555, doi: 10.1080/13597566.2019.1591375.
- Roberts, K. M., and Wibbels, E. (1999). Party Systems and Electoral Volatility in Latin America: A Test of Economic, Institutional, and Structural Explanations, *The American Political Science Review*, 93(3), pp. 575–590. <https://doi.org/10.2307/2585575>.
- Samuels, D. (1999) Incentives to Cultivate a Party Vote in Candidate-centric Electoral Systems: Evidence from Brazil, *Comparative Political Studies*, 32(4), pp. 487–518. doi: 10.1177/0010414099032004004.
- Samuels, D. (2000). Concurrent Elections, Discordant Results: Presidentialism, Federalism, and Governance in Brazil. *Comparative Politics*, 33(1), pp. 1–20. <https://doi.org/10.2307/422421>.
- Santos, F. G. (2020). Social media, disinformation, and regulation of the electoral

- process: a study based on 2018 Brazilian election experience. *Journal of Constitutional Research*, 7(2), pp. 429-449. doi: 10.5380/rinc.v7i2. 71057.
- Slapin, J. B., and Proksch, S.O. (2008). A Scaling Model for Estimating Time-Series Party Positions from Texts. *American Journal of Political Science*, 52(3), pp. 705–722.
- Wilkerson, J. and Casas, A. (2017). Large-Scale Computerized Text Analysis in Political Science: Opportunities and Challenges, *Annual Review of Political Science*, 20(1), pp. 529-544. <https://doi.org/10.1146/annurev-polisci-052615-025542>.
- Williams, M.L., Burnap, P., Sloan, L., Jessop, C. and Lepps, H. (2017), Users' Views of Ethics in Social Media Research: Informed Consent, Anonymity, and Harm. In: K, Woodfield (ed.) (2017) *The Ethics of Online Research (Advances in Research Ethics and Integrity, Vol. 2)*. Emerald Publishing Limited. [online] Available through Lund University Library website: <https://lubcat.lub.lu.se/> [Accessed 10th May 2022].
- Woodfield, K. and Iphofen, R. (2017), Introduction to Volume 2: The Ethics of Online Research, K. Woodfield, (ed.) *The Ethics of Online Research (Advances in Research Ethics and Integrity, Vol. 2)*. Bingley: Emerald Publishing Limited. [online] Available through Lund University Library website: <https://lubcat.lub.lu.se/> [Accessed 10th May 2022].
- Zittel, T. (2009). Lost in Technology? Political Parties and the Online Campaigns of Constituency Candidates in Germany's Mixed Member Electoral System, *Journal of Information Technology & Politics*, 6(3-4), pp. 298-311. doi: 10.1080/19331680903048832.
- Åström, J. and Karlsson, M. (2013). Blogging in the Shadow of Parties: Exploring Ideological Differences in Online Campaigning, *Political Communication*, 30(3), pp. 434-455, doi: 10.1080/10584609.2012.737430.