

Inovando a Cópia^{*}

**Examinando estatutos partidários
através da análise de similaridade textual**

Lucas Mingardi

12º Encontro da ABCP

19 a 23 de Outubro de 2020

Evento Online


Área Temática: 3. Eleições e Representação Política

^{*}Trabalho em andamento. Favor não citar sem a expressa autorização do autor.

Resumo

Uma das formas mais elementares de se analisar a estrutura organizativa dos partidos políticos é através da leitura de seus estatutos partidários. O presente capítulo parte da concepção de texto como dados (*text as data*). Defendemos que mais do que meros veículos de registro informacional, estatutos são, antes de mais nada, documentos políticos independentes, portanto passíveis de serem estudados isoladamente. A desatenção da literatura a este problema resultou em sabermos muito pouco sobre como se dá o processo de reedição dos estatutos. O caso brasileiro é especificamente interessante para avaliarmos tal ponto. Estudos anteriores mostram que apesar da autonomia ganha com a redemocratização, estatutos brasileiros permaneceram razoavelmente idênticos entre si e em relação à estrutura organizativa imposta no período militar. Por outro lado, ao olharmos para o processo de reedição dos estatutos, vemos um cenário muito mais dinâmico, com casos extremos de partidos com dezenas de estatutos ao longo de sua formação. Nosso universo de análise é composto por 194 estatutos feitos por 35 partidos nos últimos 27 anos. Para avaliarmos a semelhança dos estatutos brasileiros, utilizamos técnicas de análise de similaridade textual. Tanto para a comparação dos estatutos com a LOPP quanto para a comparação dos estatutos entre si, percebemos que a taxa de similaridade do estatuto anterior do partido e a data de edição do partido são significativas na similaridade dos estatutos. Constatamos que, ao incluir mais casos em nossa análise, relativizamos parcialmente o argumento de que o Partido dos Trabalhadores seria distinto dos demais partidos quanto à sua estrutura estatutária.

Introdução

 PESAR de extensa a literatura acerca da organização partidária, há certa tensão na forma pela qual tratam sua fonte principal de pesquisa: os estatutos partidários. Defendemos que mais do que meros veículos de registro informacional, estatutos são, antes de mais nada, documentos políticos independentes, portanto passíveis de serem estudados como objeto próprio de pesquisa. A desatenção da literatura quanto a este problema resultou em sabermos muito pouco acerca de como se dá o processo de criação e de transformação dos estatutos partidários.

O presente capítulo parte da concepção de texto como dados (*text as data*). O desenvolvimento da análise quantitativa de texto e, em especial, a evolução de sua aplicação na ciência política têm se mostrado inovadores na forma pela qual lidamos com documentos. Parafraseando Grimmer e Stewart 2013, p.272, a linguagem é complexa, contudo, não necessitamos de toda a complexidade da linguagem para efetivamente analisar textos. Em outras palavras, assim como em áreas correlatas, a análise quantitativa de textos visa criar um modelo simplificado do mundo que seja capaz de trazer informações úteis acerca de nosso objeto de estudo.

Ao deslocarmos a análise para os estatutos partidários enquanto documentos, não estamos nos afastando da análise de estruturas de organização partidária em si. Estatutos partidários são, afinal, “um conjunto de regras escritas pelas quais partidos políticos estruturam e organizam suas atividades” (National Democratic Institute 2011). Pelo contrário, o que realçamos aqui é que a própria forma pela qual as regras organizativas são redigidas é uma variável importante para compreendermos a estrutura organizativa como um todo. A análise que propomos aqui, portanto, parte da tradicional literatura que es-

tuda a organização partidária através dos estatutos¹, propondo um esquema teórico e metodológico que facilite justamente a comparação entre os mesmos.

Utilizamos técnicas de análise de similaridade textual² para comparar nossos documentos. Apesar de relativamente recente, tal técnica vem ganhando força na ciência política como critério de comparação de textos³. Apesar de ter em muito avançado a compreensão geral da estrutura organizativa dos partidos, ainda é muito difícil de se estabelecer critérios de comparação. A análise de similaridade textual, assim, é uma forma robusta de se fazer tal avaliação⁴.

Focamos no caso brasileiro como estudo de caso para aplicar o desenho de pesquisa aqui proposto. Acreditamos que os partidos brasileiros ilustram bem nosso ponto de alteração incremental da redação dos estatutos. Durante o período autoritário, a organização dos partidos era fortemente regulada pelo estado⁵. Já durante o período de redemocratização, partidos ganharam ampla autonomia em sua regulamentação interna. Estudos anteriores (Ribeiro 2012; Ribeiro e Fabre 2019) mostram como, apesar desta autonomia, os estatutos partidários em muito mantiveram intactas as estruturas do regime militar.

¹Historicamente, podemos citar o monumental trabalho de Hartmann 1931 como precursor da área. Mais recentemente, podemos citar os trabalhos de Biezen 2000; Biezen e Piccio 2013; Calossi e Pizzimenti 2015; Freidenberg e Levitsky 2007; Hazan e Rahat 2010; Ignazi, Bardi e Massari 2010; Ignazi e Pizzimenti 2014; Katz 1992; Katz e Mair 1994; Katz 2002; Norris 2004; Ribeiro 2012; Ribeiro e Fabre 2019; Von dem Berge et al. 2013.

²Também conhecidas pela literatura especializada como análise de plágio. Ocorre que plágio usualmente é entendido como a cópia ilegal de trechos de outros documentos, o que não é o caso para nossos textos analisados. Por conveniência, passaremos a utilizar os termos 'reutilização de texto', 'similaridade textual' e 'plágio' intercambiadamente.

³Acree et al. 2016 fornecem um bom ponto de início, mostrando como diferentes campos na ciência política utilizam tal técnica.

⁴Como Grimmer e Stewart 2013 adverte, não devemos pensar na análise quantitativa de texto como um 'substituto' para a leitura humana, mas sim como seu complemento. É claro que nosso critério não é ideal para analisar pormenorizadamente os estatutos partidários. Todavia, apresenta uma forma consistente de comparar um elevado número de textos.

⁵Regulamentada pela Lei 5.672/1971, também conhecida pelo acrônimo LOPP

Talvez um tanto paradoxalmente, quando olhamos para o processo de reedição dos estatutos, encontramos um cenário muito mais dinâmico, com partidos por vezes tendo mais de dez estatutos durante o período analisado. O presente capítulo, assim, é em parte um balanço dos achados encontrados por Ribeiro. Ampliamos o escopo temporal, de 1992 a 2019, bem como o número de casos analisado, 35 partidos, totalizando a comparação de 194 estatutos distintos. Nosso objetivo, assim, é fundamentar melhor o debate, avaliando quão similares são os estatutos brasileiros.

Resumindo, antes de serem representações das estruturas organizativas dos partidos políticos, estatutos partidários são documentos escritos. Podemos encarar o próprio espaço de escrita dos estatutos partidários como um *locus* de poder, uma vez que estes determinam, em maior ou menor grau, a forma pela qual os membros do partido disputam internamente pelo poder⁶. Assim sendo, uma forma de avaliarmos a semelhança na estrutura organizativa dos partidos seria avaliar a similaridade do próprio corpo textual dos estatutos partidários. A principal vantagem de se utilizar tal técnica seria na facilidade de comparação direta dos estatutos partidários. O caso brasileiro se mostra particularmente rico para nossa análise. Na seção seguinte, enquadraremos melhor o quadro institucional dos estatutos brasileiros, melhor formulando assim nossas hipóteses de pesquisa. Em seguida, descrevemos o processo de obtenção de nossos dados e da geração de nossos índices de similaridade. Passamos por uma análise exploratória dos estatutos brasileiros e, por fim, uma avaliação da similaridade dos estatutos brasileiros entre si e para com a LOPP.

⁶Biezen e Piccio 2013; Ignazi e Pizzimenti 2014; Katz e Mair 1994; Katz 2002; Panebianco 2005.

O Contexto Institucional dos Estatutos Brasileiros

Para compreendermos a dinâmica dos estatutos partidários brasileiros, devemos analisar sua evolução a partir da transição do regime militar para o sistema democrático vigente. Durante a ditadura, os partidos políticos não tinham liberdade de estabelecer sua própria estrutura organizativa, tal matéria sendo regulamentada pela LOPP⁷. Seus estatutos partidários, portanto, meramente copiavam suas diretrizes. A partir da redemocratização, a Constituição de 1988 passou a prever a liberdade de organização dos partidos. Foi somente em 1995, contudo, com a Lei dos Partidos Políticos⁸, que os estatutos partidários passaram a ser regulamentados. Apesar da Lei dos Partidos contemplar alguns requerimentos mínimos dos estatutos partidários, pode-se dizer que é bem liberal quanto a forma que os partidos podem se organizar (Silva Ramos 2008). É importante notar que, com a Lei 9.096/1995, a criação de estatutos passou a ser obrigatória aos partidos, devendo estes enviar seus estatutos, bem como suas eventuais alterações, para o TSE⁹. Conjecturamos, assim, que o tempo seja uma variável fundamental para analisar a semelhança dos partidos brasileiros.

Dada esta radical alteração institucional, podemos nos perguntar como os partidos que surgiram a partir da redemocratização organizaram seus estatutos. O estudo seminal de Ribeiro 2012 vai neste sentido. Vemos em Ribeiro uma forte preocupação em analisar os múltiplos estatutos redigidos ao longo do tempo, buscando entender o que estava em jogo em cada mudança. Analisando a evolução dos

⁷Lei 5.672/1971.

⁸Lei 9.096/1995.

⁹Para os partidos já registrados definitivamente, a Lei 9096/95 colocava o dispositivo transitório de que o estatuto fosse criado e publicado em até 6 meses após a publicação dessa. Publicada em 19/03/1995, isso significa que os partidos existentes no período teriam até Março de 1996 para se adequar às novas medidas.

estatutos de quatro partidos brasileiros¹⁰, conclui o autor que, com a exceção do PT, os estatutos dos partidos continuavam a estritamente seguir a estrutura organizativa desenhada pela LOPP. Nas palavras do autor, “os documentos de diferentes partidos não só continuaram idênticos entre si em muitos trechos: eles também seguem reproduzindo exatamente capítulos inteiros fornecidos pela LOPP em 1971”¹¹.

A análise de Ribeiro 2012 norteará as principais hipóteses de pesquisa trabalhadas neste capítulo. Podemos generalizar seus achados e fundamentar as seguintes questões de pesquisa: 1) qual o grau de semelhança dos estatutos partidários brasileiros para com a LOPP; 2) qual o grau de semelhança dos partidos brasileiros entre si, subentendendo-se, com isso, 2.1) o grau de semelhança entre partidos distintos, e 2.2) o grau de semelhança intrapartidário dos estatutos redigidos ao longo dos anos.

Transformando Textos em Dados

Nosso universo de análise é composto pelos estatutos dos partidos brasileiros contemporâneos¹², coletados no *site* do Tribunal Superior Eleitoral (TSE). Incluímos não só os estatutos mais recentes de cada partido, mas também todos seus estatutos prévios. Com isso, temos 194 estatutos distintos para 35 partidos políticos¹³. Excluimos de nossa análise estatutos de partidos criados antes da redemocratização brasileira, ou de partidos pós-redemocratização que por ventura foram extinguidos ou incorporados a outros partidos¹⁴. Incluímos, contudo,

¹⁰Sendo estes: PMDB (atual MDB), PSDB, PT e PFL/DEM.

¹¹Ribeiro 2012, p. 238

¹²Sendo estes: AVANTE, DEM, MDB, NOVO, PATRI, PCB, PCdoB, PCO, PDT, PHS, PMB, PMN, PODE, PP, PPL, PPS, PR, PRB, PROS, PRP, PRTB, PSB, PSC, PSD, PSDB, PSDC, PSL, PSOL, PSTU, PT, PTB, PTC, PV, REDE e SD.

¹³Os dados foram coletados em Fevereiro de 2019.

¹⁴Tal restrição é decorrente menos por nossas próprias decisões, e mais pelo volume de informações disponibilizadas pelo TSE. Mais recentemente, conseguimos ter acesso a parte dos estatutos de

estatutos de partidos que mudaram a denominação de sua legenda¹⁵.

A mera coleção de documentos não os transforma em um banco de dados propriamente dito. Precisamos, assim, transformar nossos dados, para trabalharmos adequadamente com estes¹⁶. Conseguimos extrair o texto de 163 dos 194 de nossos estatutos através do pacote *pdftools* (Ooms 2017). Para cerca de 20% de nossos dados restantes, tivemos de usar técnicas de reconhecimento óptico de caracteres (*optical character recognition*). Decidimos utilizar o *tesseract* (Smith 2007) através do pacote homônimo do R (Ooms 2018c). Para aumentar a acurácia do reconhecimento dos caracteres, transformamos as páginas de cada estatuto em imagens, limpamo-las (Ooms 2018b), e ‘treinamos’ nosso OCR com um dicionário português (Ooms 2018a)¹⁷.

Por fim, com nosso banco de dados criado, passamos a limpar o corpo dos textos. Removemos números, sinais de pontuação (com exceção de hifens intra-palavras) e *stopwords*¹⁸, bem como tiramos a capitalização de todas as palavras. Em seguida, realizamos um *stemming*¹⁹ de nossos *corpora*. Conforme Grimmer e Stewart 2013, tal prática serve para reduzir a complexidade do texto, reduzindo o cálculo computacional, bem como facilitando a análise e comparação dos textos.

partidos extinguidos. Todavia, não conseguimos incluí-los ainda em nossa análise.

¹⁵Assim como o próprio TSE, tratamos estes casos como se fossem o mesmo partido.

¹⁶Conseguimos ler 193 dos 194 estatutos através dos métodos que descrevemos a seguir. O estatuto que removeremos de nossa análise é o do PV de 1990, dado que este foi redigido manualmente.

¹⁷Um problema que nos deparamos no tratamento de nossos dados foi que em quase todos os estatutos há carimbos e assinaturas, o que acabou criando um ruído nos nossos dados. Para a presente análise, levaremos em conta que a presença desses ruídos é distribuída aleatoriamente e que, portanto, não há um problema de viés sistemático em nossa pesquisa.

¹⁸Palavras comuns, que não possuem nenhum significado ‘próprio’, mas sim que são utilizadas para desempenhar alguma função gramatical, como, por exemplo, capítulos e preposições.

¹⁹Isto é, reduzimos as palavras ao seu radical comum. Por exemplo, ‘município’, ‘municípios’ e ‘municipal’ e ‘municipais’ são reduzidos todos para o radical comum ‘municip’.

A Análise de Similaridade Textual

Algumas observações devem ser feitas quanto às técnicas de análise de similaridade textual que utilizamos. Os resultados reportados pelos índices dependem fortemente das escolhas metodológicas feitas pelo pesquisador. Embora seja bem maleável a forma de se criar tais índices, apresentamos aqui algumas boas práticas que podem nortear o critério de similaridade escolhido.

O primeiro passo seria definir o que entendemos como nossos termos, isto é, o menor elemento constitutivo de nossos documentos. A forma mais trivial seria identificar os termos como as palavras que compõem um documento. De fato, todos os trabalhos citados por Acree et al. 2016 utilizam tal estratégia. Todavia, há operações mais eficientes do que esta. Desde meados dos anos 1990 a análise de caracteres vem ganhando força no campo. Sua principal vantagem seria que, em contraposição à análise de palavras, pode ser utilizada independentemente da linguagem do texto, não necessita necessariamente de nenhum pré-processamento de palavras²⁰, bem como tende a ser mais robusta em relação a ruídos²¹ (Barrón-Cedeño e Rosso 2009; Stamatatos 2009; Kestemont 2014).

Em seguida, definimos qual a dimensionalidade de nossos n-gramas, isto é, quantos termos analisamos consecutivamente²². Como Barrón-

²⁰Esta seria, ao nosso ver, uma das principais vantagens para nosso caso particular. Todos os dicionários de *stemming* que encontramos para o português são deficientes em alguma medida. Além disso, esta técnica não funciona bem para línguas aglutinativas, como o alemão ou o finlandês. Como afirma Kestemont 2014, apesar de parecer uma 'caixa-preta', a comparação de caracteres capta a variância linguística de uma forma muito mais natural do que a comparação de palavras.

²¹Outro ponto importante para nossa presente análise, dado que não conseguimos eliminar todos os ruídos da leitura dos documentos através do OCR.

²²Para melhor ilustrar esse ponto, vemos que a *string* `foo; bar` é composta por sete 2-gramas: 'fo', 'oo', 'o;', '; ', 'b', 'ba' e 'ar'. Observe que espaços e pontuações também são incluídos em nossos termos caso não optemos por removê-los. Enquanto que incluir pontuações pode capturar diferenças de estilo mais precisamente, acabam exigindo maior poder de computação, além de potencialmente gerar certo ruído na análise.

Cedeño e Rosso 2009 argumentam, bigramas e trigramas são os n-gramas ótimos para a comparação de *corpora*²³. A escolha por 2-gramas ou 3-gramas se resume por um *trade-off* entre *recall* e precisão. Optamos por analisar estas duas medidas conjuntamente por uma questão de robustez.

Por fim, resta-nos escolher como comparar os estatutos, i.e., que critério de detecção de plágio será utilizado. Optamos por comparar dois critérios comumente utilizados para este fim, o índice de Jaccard e a similaridade de cosseno. O índice de Jaccard, também conhecido como ‘índice da intersecção sobre a união’, é um critério bastante simples e, como seu outro nome indica, é definido por:

$$JA, B = \frac{A \cap \Sigma B}{\Sigma A \cup \Sigma B} \quad (1)$$

Em que A e B são o conjunto de termos dos textos A e B, respectivamente. É trivial que quanto mais próximos forem a intersecção e a união do conjunto de palavras, mais próximo de 1 será o valor, e vice-versa. Veja também que o que calculamos é a presença ou não de n-gramas em A e B, e não quantas vezes tais n-gramas aparecem em cada documento. Já a similaridade de cosseno é calculada pensando nos termos de nossa matriz como vetores, calculando-se o cosseno de seu ângulo. É definida por:

$$\cos A, B = \frac{A.B}{\|A\| \|B\|} = \frac{\sum_{i=1}^n A_i B_i}{\sqrt{\sum_{i=1}^n A_i^2} \sqrt{\sum_{i=1}^n B_i^2}} \quad (2)$$

Simplificando, podemos compreender que, grosso modo, o que varia entre um índice e outro é sobretudo a razão pela qual se está divi-

²³Os autores concluem em seu capítulo que n-gramas de dimensionalidade maior que 3 não apresentam ganho real de eficiência, aumentando contudo o tempo de análise. Similarmente, 1-gramas têm baixa performance, especialmente quando são analisados caracteres e não palavras (Barrón-Cedeño e Rosso 2009, 699)

dindo os termos. A similaridade de Jaccard divide o número comum de termos pelo número de termos que existe em pelo menos um dos documentos. Já a similaridade de Cosseno divide o número de termos comuns pelo número total de termos possíveis.

Temos, assim, 4 testes de similaridade textual, a depender do n-grama que escolhemos (2-grama e 3-grama) e do tipo de análise de similaridade (Jaccard e Cosseno)²⁴. Apesar das diferenças de formulação dos índices, todos são operacionalizados de forma razoavelmente similar: seu resultado cai em um intervalo que vai de 0 a 1, com 0 significando que não há nenhum componente comum entre os textos, e 1 significando que os textos são completamente idênticos. Garrett e Jansa 2015 mostram como é mais natural entender esse resultado como uma medida *relativa*. O que nos importa não é tanto a estimativa *pontual* da similaridade de um texto A para com um texto B, mas sim em comparar tal as estimativas dos diversos textos, isto é, sua *variância*²⁵.

Para facilitar a compreensão de como cada um de nossos índices funcionam, empregamos nossos quatro índices de similaridade comparando ambos os excertos abaixo. Não removemos stopwords, pontuações ou fizemos stemming de nossos documentos para nosso exemplo. Encontramos que $C2 = 0.94$, $C3 = 0.87$, $J2 = 0.82$ e $J3 = 0.74$. Como podemos ver, no geral, os índices de cosseno reportam valores superiores ao de Jaccard, e bigramas reportam valores superiores a trigramas.

Texto 1: Estatuto do MDB 15.5.2018.

²⁴Passaremos utilizar 'J' e 'C' para indicar, respectivamente, os índices de Jaccard e Cosseno, e '2' e '3' para bigramas e trigramas. Nossos quatro índices, portanto, são J2, J3, C2 e C3.

²⁵De fato, podemos imaginar que haveria certa taxa de similaridade 'natural' entre os estatutos partidários, visto que tratam de assuntos similares, bem como utilizam uma forma bem específica de linguagem. Em última instância, isto não representa um problema, pois o máximo que essa similaridade 'natural' poderia afetar seria elevar a *baseline* de nossos índices.

Art. 1º. O Movimento Democrático Brasileiro - MDB, partido com sede e domicílio jurídico em Brasília, Capital da República, reger-se-á por este Estatuto, definidor de sua estrutura interna, organização e funcionamento, nos termos do art. 17 da Constituição Federal, bem como, no que couber, pela legislação federal infraconstitucional em vigor.

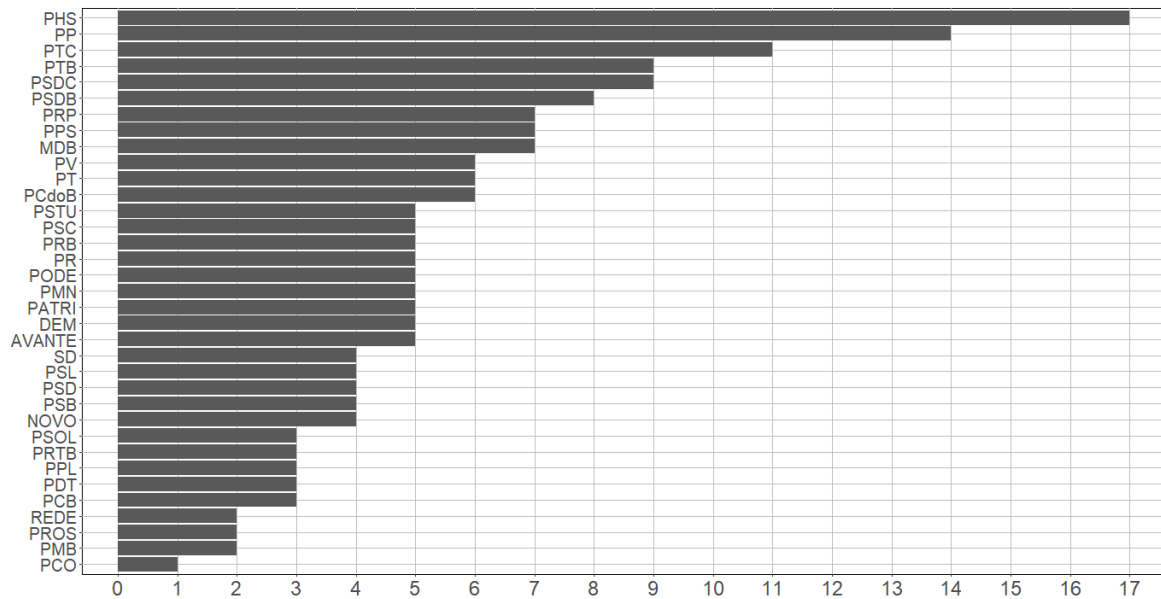
Texto 2: Estatuto MDB 18.12.2017.

Art. 1º O Partido do Movimento Democrático Brasileiro - PMDB, com sede e domicílio jurídico em Brasília, Capital da República, reger-se-á por este Estatuto, definidor de sua estrutura interna, organização e funcionamento, nos termos do art. 17 da Constituição Federal.

Breve Análise Descritiva dos Estatutos Brasileiros

Antes de passarmos para a análise da taxa de similaridade dos estatutos brasileiros, daremos um passo para trás, analisando os estatutos brasileiros de forma mais descritiva. A presente seção, pois, tem forte caráter exploratório. Como os estatutos brasileiros ainda são, em grande medida, uma caixa-preta, tal atividade se mostra útil para compreendermos melhor esse fenômeno de maneira mais abrangente. Como vimos, investigamos no presente capítulo 194 estatutos de 35 partidos distintos. Assim, um primeiro ponto que podemos analisar seria o número de estatutos por partido. A Figura 1 demonstra essa relação. O PCO é o partido com o menor número de estatutos, com apenas 1, enquanto o PHS foi o partido que mais alterou seu estatuto, tendo no total 17. A média de estatutos por partido foi de 5.54 e a mediana 5.

Figura 1: Número de Estatutos por Partido



Outra questão que podemos verificar é a data em que tais estatutos foram criados. Plotamos na Figura 2 a quantidade de estatutos criados por ano. Nossa série temporal começa com 1 único estatuto, do PV, feito em 1992. Os anos de 1993 e 1994 não tiveram nenhum estatuto registrado. Durante 1995 e 1996 salta o número de estatutos, dada a obrigatoriedade prevista pela Lei 9096/95. A geração de estatutos permanece relativamente constante até 2002. De 2003 em diante, contudo, um padrão interessante começa a ser criado. Não só o número de estatutos criados passa a ter uma tendência crescente, mas também a quantidade de estatutos criados é sempre maior em anos não-eleitorais do que em anos eleitorais. Isso pode indicar um padrão de que os partidos estariam mudando suas regras internas para se adequar às eleições seguintes. Sem verificar mais profundamente a mudança em tais estatutos, contudo, não temos como afirmar categoricamente esse ponto.

Ainda quanto a data de geração dos partidos, podemos investigar o mês que estes foram criados. Dada a possibilidade de haver padrões distintos em anos eleitorais e não-eleitorais, separamos na Figura 3 a quantidade de estatutos criados por mês pelo tipo de ano. Pela maior parte do ano, anos eleitorais e não-eleitorais são relativamente similares quanto ao número de estatutos criados. Todavia, a partir de Outubro, e sobretudo em Dezembro há uma quantidade significativamente maior de estatutos sendo redigidos em anos não-eleitorais.

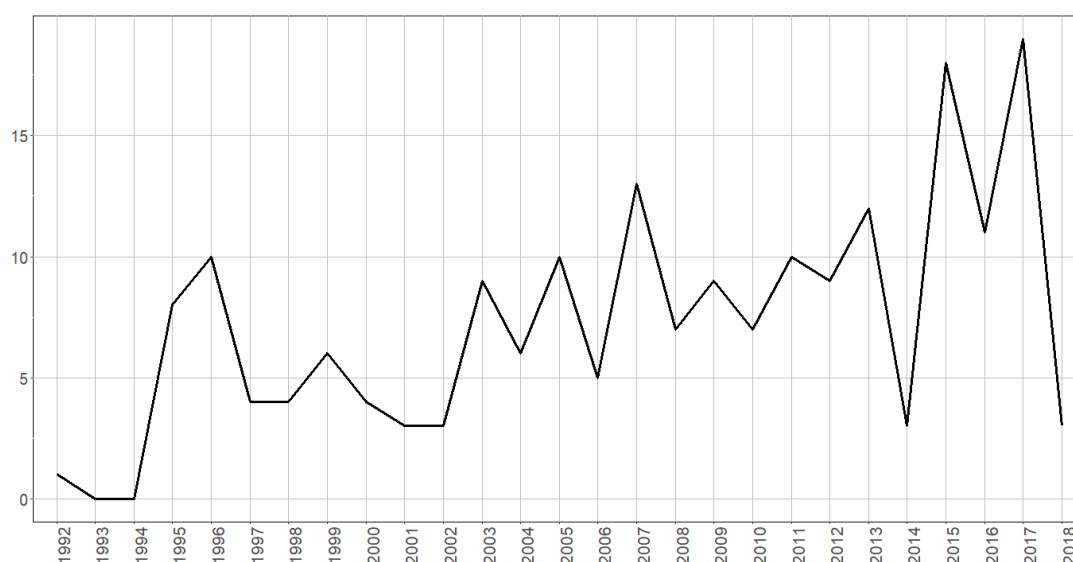


Figura 2: Número de Estatutos Criados por Ano



Figura 3: Número de Estatutos por Mês

Podemos imaginar que o processo de reedição de estatutos é incremental e supor que o número de palavras dos estatutos aumenta à medida que os anos passem. Testamos tal hipótese na Figura 4. Plotamos neste a quantidade média (linha sólida) e mediana (linha tracejada) de palavras utilizadas nos estatutos criados em dado ano. Não há grandes diferenças na média e mediana do número de palavras. É útil lembrar que nenhum estatuto foi criado durante os anos de 1993 e 1994. Como podemos ver, há, sim, uma tendência geral do aumento do número médio/mediano de palavras ao longo dos anos, sendo que o pico do número de palavras se deu em 2001, com um número médio de 10633 palavras, e mediana de 12860 palavras.

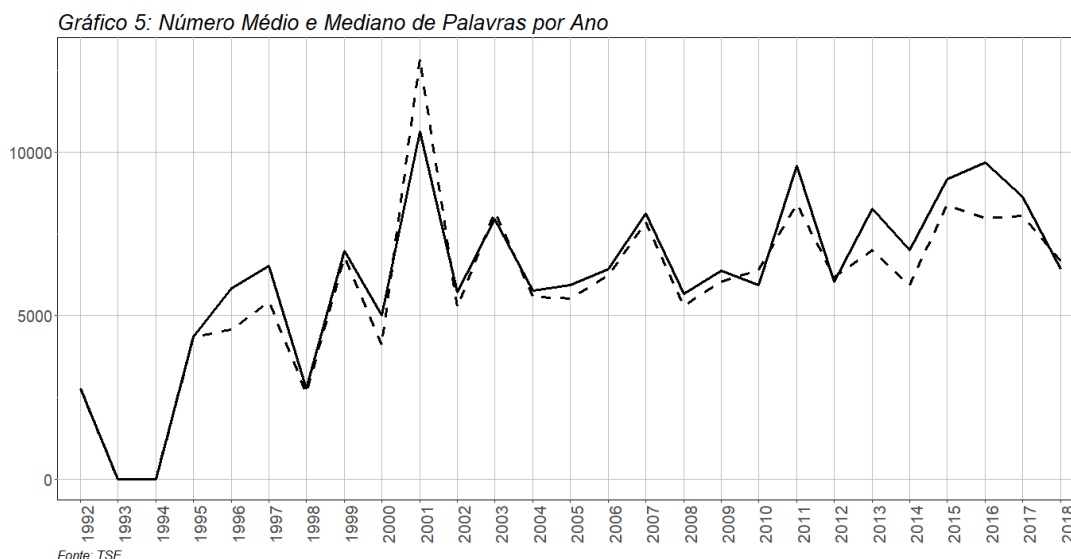


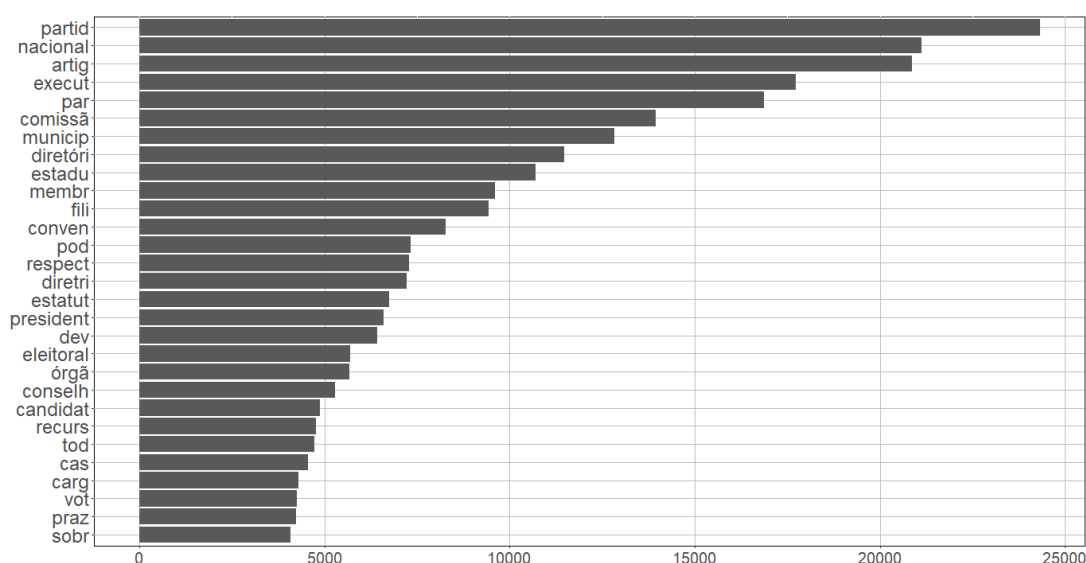
Figura 4: Número Médio e Mediano de Palavras por Ano de Redação do Estatuto

Já quanto ao número total de palavras distintas, mesmo após *stemming* e remover as *stop words*, os estatutos apresentam 10140 palavras únicas. Podemos analisar quais dessas palavras são mais comuns entre os estatutos brasileiros, representado pelo gráfico de barras na Figura 5. O termo mais comum é ‘partid’ (utilizado 23325 vezes). Um primeiro grupo de palavras que podemos destacar seria de palavras que são próprias da escrita jurídica: ‘artig’ (19921) ²⁶ e ‘par’ (16100). Em seguida, temos termos relativos à estrutura federativa: ‘nacional’ (19810), ‘municip’ (12187) e ‘estadu’ (10112). Curioso notar a preponderância de municipal sobre estadual. O terceiro grupo a destacar seria de órgãos e instituições partidários: ‘comissã’ (12819), ‘diretóri’ (11497), ‘diretri’ (6480), ‘estatut’ (6375), ‘órgã’ (5694), ‘conselh’ (4879) e ‘convençã’ (4169). Por fim, temos um grupo de palavras relativo à vida eleitoral e político-partidária: ‘membr’ (9150), ‘fili’ (9018), ‘pod’ (6996), ‘president’ (6300), ‘dev’ (6115), ‘eleitoral’

²⁶Somamos a ela também a abreviatura ‘art’, dado que em textos jurídicos ela nunca é escrita inteiramente, e que o *stemming* por si não reconheceria tal prática.

(5421), ‘candidat’ (4659), ‘recurs’ (4555), ‘praz’ (4072), ‘carg’ (4059) e ‘vot’ (4046). Esses grupos de palavras em muito se assemelha com a estrutura de classificação de estatutos proposta por Von dem Berge et al. 2013, o que demonstra como a análise quantitativa de texto pode ser útil para auxiliar métodos mais qualitativos de análise.

Figura 5: Palavras mais Comuns nos Estatutos



Caso 1: Similaridade Textual com a LOPP

Passemos agora a avaliar a primeira parte do argumentado levantado por Ribeiro 2012, da semelhança textual entre os estatutos partidários brasileiros e a LOPP. Dos 195 estatutos coletados tivemos de excluir 4, totalizando 191 estatutos analisados²⁷. Utilizamos 4 taxas de similaridade de caractere, a depender do tamanho do n-grama (bigramas

²⁷Foram estes: o estatuto do PV de 1992 que por ter sido redigido manualmente não foi possível ser lido por nosso OCR, bem como os estatutos de 1996, 1998 e 1999 do PPB (atual PP), por não serem estatutos partidários próprios, mas sim excertos anunciando os capítulos alterados do estatuto anterior.

e trigramas) e do critério de similaridade (Jaccard ou Cosseno).

A Figura 6 mostra a densidade Kernel dos índices de similaridade dos partidos com a LOPP. As curvas de linha tracejada se referem às taxas de similaridade que empregam 2-gramas, enquanto as linhas sólidas são de 3-gramas. Constatamos aqui o mesmo padrão encontrado na aplicação dos índices nossos excertos dos estatutos do MDB, com os valores de cosseno sendo, na média, maiores que os de Jaccard, e bigramas superiores a trigramas. Olhando para estatísticas pontuais, a média de foi $J2 = 0.61$, $J3 = 0.5$, $C2 = 0.97$ e $C3 = 0.91$. Estas e outras estimativas pontuais podem ser vistas na Tabela 1.

Figura 6: Densidade Kernel Estatuto x LOPP

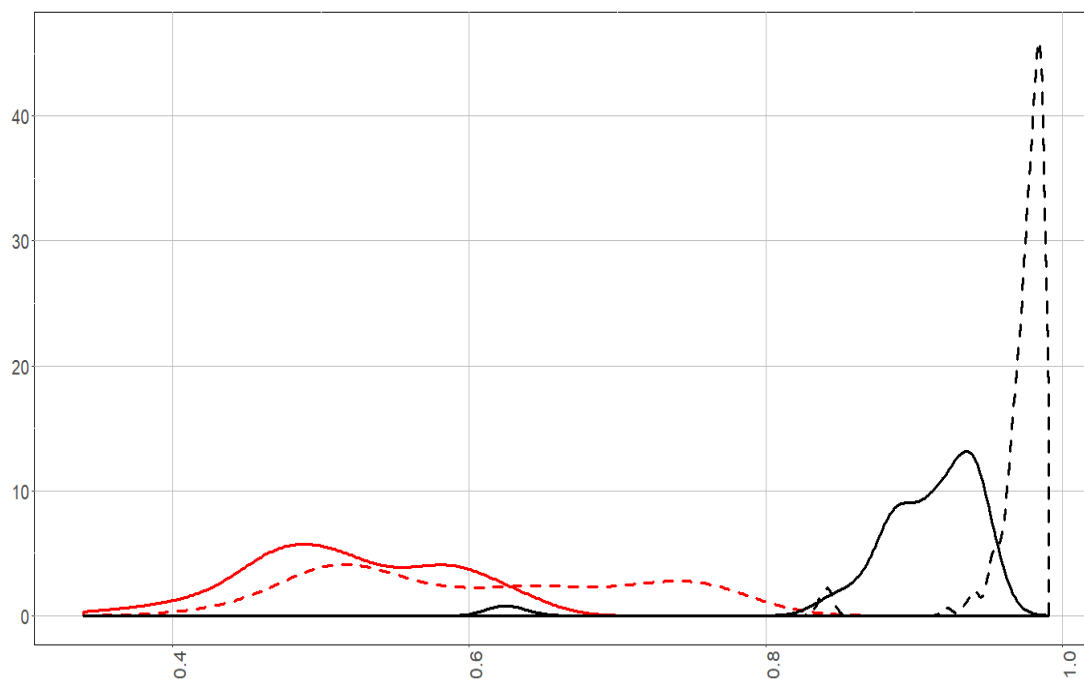


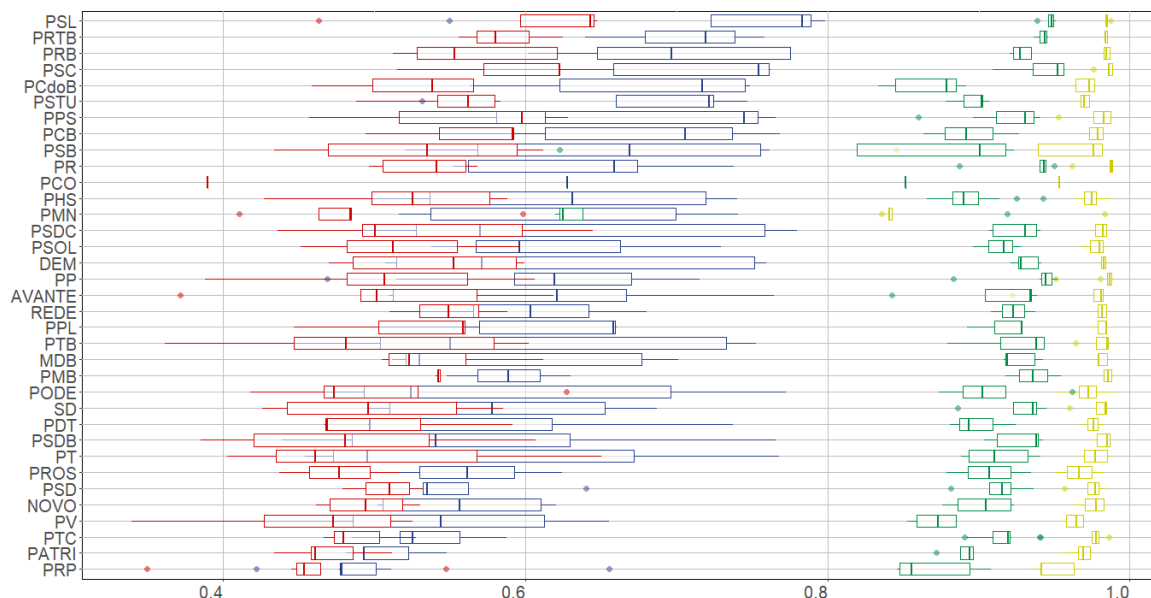
Tabela 1: Medidas Pontuais das Taxas de Similaridade com a LOPP

Medida	J2	J3	C2	C3
Min	0.42	0.34	0.83	0.61
Q1	0.51	0.47	0.97	0.89
Mediana	0.59	0.51	0.98	0.88
Média	0.61	0.50	0.97	0.91
Q3	0.70	0.57	0.98	0.93
Max	0.79	0.65	0.99	0.96
Desvio-Padrão	0.11	0.07	0.03	0.56

O argumento original de Ribeiro 2012 inclui a questão de como os partidos diferem em relação à sua similaridade com a LOPP. Contudo, não temos um número suficiente de estatutos por partido para nos dar grau de liberdade para modelar esta hipótese. Uma alternativa seria avaliar graficamente tal afirmação. O argumento original de Ribeiro 2012 inclui a questão de como os partidos diferem em relação à sua similaridade com a LOPP. Contudo, não temos um número suficiente de estatutos por partido para nos dar grau de liberdade para modelar esta hipótese. Uma alternativa seria avaliar graficamente tal afirmação. A Figura 7 mostra os boxplots dos nossos quatro índices de similaridade. Em vermelho representamos o Jaccard trigramma (J3), em azul o Jaccard bigrama (J2), em verde o Cosseno trigramma (C3) e em amarelo o Cosseno bigrama (C2). Como podemos ver, encontramos um padrão semelhante de ao encontrado por Ribeiro em sua análise substantiva da similaridade da LOPP com os estatutos por ele analisado (PT, PSDB, MDB, DEM). Todavia, quando incluimos mais casos em nossa análise, relativamos os achados de Ribeiro. Se, de fato, o PT é um dos partidos menos similares com a LOPP, vemos que partidos como o PV e o NOVO, ou mesmo partidos como o PSD e o PRP são ainda mais dissimilares. Similarmente, a distribuição da similaridade do PSDB não se mostrou radicalmente distinta da distribuição do

PT.

Figura 7: Boxplots da Similaridade dos Estatutos e a LOPP por Partido



Passemos a especificar nossas hipóteses acerca da similaridade dos estatutos brasileiros com a LOPP. Retomando nossa interpretação do esquema teórico de Ribeiro (Ribeiro 2012; Ribeiro e Fabre 2019), defendemos que o ano de criação do estatuto é fundamental para compreender o processo de criação de estatutos no Brasil. A Lei dos Partidos Políticos criou um constrangimento institucional aos partidos, forçando-os a criar estatutos sob pena de não poder concorrer nas eleições de 1996. É possível de se imaginar, assim, que estatutos redigidos temporalmente mais próximos desta data sejam mais similares à LOPP do que estatutos mais distantes no tempo. Por outro lado, podemos pensar em uma hipótese mais relacionada ao *path dependence* das organizações partidárias e, por conseguinte, de seus estatutos. Enquanto que partidos mais antigos, ligados ao período militar, teriam herdado as estruturas organizativas do sistema anterior,

partidos criados mais recentemente não teriam este vínculo, potencialmente tendo estruturas organizativas distintas²⁸. O processo de edição dos estatutos é cumulativo. Portanto, podemos imaginar que a taxa de similaridade do estatuto anterior do partido se correlaciona com a taxa de seu estatuto atual. Por fim, O Partido dos Trabalhadores teria o único partido analisado por Ribeiro 2012 cuja estrutura organizativa seria distinta da LOPP. Podemos generalizar este argumento e supor que isto se trata de uma característica dos partidos de esquerda no geral²⁹.

A Tabela 2 mostra os resultados de nossos modelos OLS para cada um dos índices de similaridade. Nossa principal variável teórica é o Ano de Edição do Estatuto. Controlamos também pelo Ano de Fundação do Partido, a Ideologia do Partido e uma dummy para o PT. Por fim, para cada modelo, adicionamos a taxa de similaridade do estatuto anterior dos partidos respectiva ao índice de similaridade sendo avaliado (Lag J2, Lag J3, Lag C2, Lag C3, respectivamente).

Avaliando nossos resultados, vemos que os Lags das taxas de similaridade são as variáveis substantivamente mais significativas, sendo positivamente correlacionadas com nossas variáveis dependentes. Nossa variável de interesse, o Ano de Edição do Estatuto, é significativo para os modelos de Jaccard (modelos (1) e (2)), mas não significativa para os modelos de Cosseno (modelos (3) e (4)). Como esperávamos, o

²⁸Um fato relativamente comum no Brasil é a troca de legenda partidária. Para fins legais, partidos sob nova denominação são considerados pelo TSE como novos partidos. Como queremos testar a relação dos partidos com o regime militar, decidimos considerar partidos que trocam de legenda como o mesmo partido. Desta forma, a data de fundação será a data de fundação da primeira legenda deste.

²⁹Utilizamos três categorias, esquerda, direita e centro. A classificação da ideologia dos partidos brasileiros não é um processo trivial. Não há um estudo sistemático que cubra suficientemente todos os partidos brasileiros, logo algumas adaptações tiveram de ser tomadas. Criamos nossa própria classificação a partir da combinação de distintas métricas utilizadas pela literatura (Fernandes 1995; Coppedge 1997; Mainwaring et al. 2000; Wiesehomeier e Benoit 2007; Zucco Jr 2011; Madeira e Silva Tarouco 2013; Bolognesi e Codato 2016; Silva Tarouco e Madeira 2015). Assim como reportado por Silva Tarouco e Madeira 2015, encontramos uma alta taxa de correlação entre as distintas mensurações da ideologia, o que, ao menos, mostra um sinal de robustez.

ano do estatuto é associado negativamente com nossas variáveis de interesse. Não podemos rejeitar a hipótese nula para o Ano de Fundação do Partido em todos nossos modelos. Curiosamente, partidos de esquerda se mostraram associados positivamente com nossos índices dos modelos (1) e (2), quando comparados aos partidos de centro. Já os partidos de direita também se mostraram associados positivamente para os modelos (2), (3) e (4). Por fim, há uma associação negativa significativa do PT para o modelo (1), o que parece corroborar a leitura de Ribeiro 2012 acerca dos partidos por ele analisado.

Caso 2: Similaridade Textual entre Estatutos

O segundo ponto do argumento de Ribeiro 2012 seria da taxa de similaridade dos estatutos brasileiros entre si. Novamente, analisamos quatro índices de similaridade textual. O número de estatutos analisados, todavia, diminui mais uma vez mais (vide a análise da Figura 8, abaixo). Temos, assim, 190 partidos. Como estamos comparando estatutos entre si, isto implica que o número total de comparações de documentos é de $190 \times 190 - 1 = 35910$ observações.

Vejamos como nossas variáveis dependentes se distribuem. A Figura 8 mostra a densidade Kernel de nossas quatro taxas de similaridade, enquanto que a Tabela 3 mostra suas estatísticas pontuais. As curvas de linha tracejada se referem às taxas de similaridade que empregam 2-gramas, enquanto as linhas sólidas são de 3-gramas. As curvas vermelhas são dos índices de Jaccard, e as curvas pretas são a taxa de cosseno. Similarmente à taxa de semelhança com a LOPP, vemos que 3-gramas se mostram menos dispersos que 2-gramas, e que a taxa de cosseno fornece uma taxa de semelhança consistentemente superior à de Jaccard. Vemos ainda que em todos os índices há casos de semelhança igual a 1. Tal fato implica uma semelhança perfeita

Tabela 2: Modelos OLS para Similaridade Textual dos Partidos com a LOPP

	<i>Variável Dependente:</i>			
	J2	J3	C2	C3
	(1)	(2)	(3)	(4)
Lag J2	0.145** (0.069)			
Lag J3		0.169*** (0.063)		
Lag C2			0.533*** (0.071)	
Lag C3				0.495*** (0.067)
Ano de Edição do Estatuto	-0.044*** (0.009)	-0.021*** (0.006)	-0.001 (0.002)	-0.002 (0.005)
Ano de Fundação do Partido	0.008 (0.008)	0.005 (0.006)	0.00002 (0.002)	-0.001 (0.005)
Ideologia de Esquerda	0.058*** (0.022)	0.035** (0.015)	0.005 (0.005)	0.005 (0.012)
Ideologia de Direita	0.026 (0.017)	0.025** (0.012)	0.010** (0.004)	0.023** (0.010)
Partido dos Trabalhadores	-0.075* (0.043)	-0.037 (0.030)	0.001 (0.011)	0.005 (0.024)
Constante	0.495*** (0.044)	0.411*** (0.033)	0.447*** (0.069)	0.442*** (0.061)
Observações	156	156	156	156
R ²	0.280	0.209	0.327	0.321
R ² Ajustado	0.251	0.178	0.300	0.294

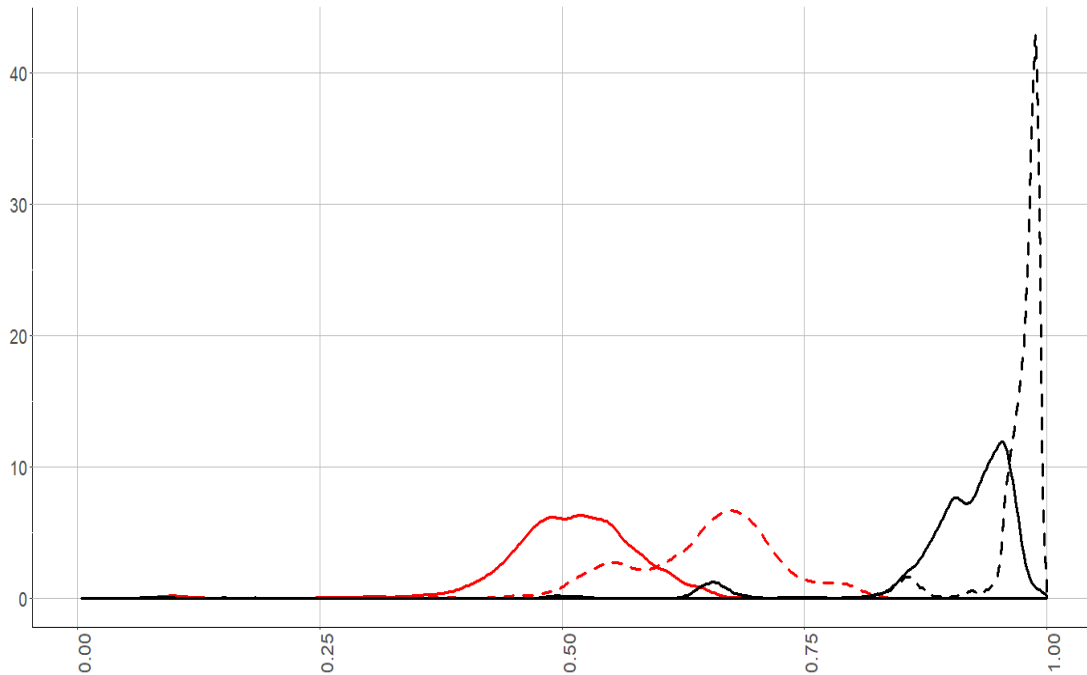
Nota: *p<0.1; **p<0.05; ***p<0.01

entre os documentos. Ao verificar os dados, vemos que isso se deu para os estatutos do PRP de 2003 e 2005. Analisando o histórico do partido, descobrimos que apesar deste ter sido fundado apenas em 2005, seu estatuto tinha sido registrado ainda em 2003. Quando foi finalmente oficializado o registro do partido, o TSE publicou ambas as versões de seu estatuto.

Tabela 3: Medidas Pontuais das Taxas de Similaridade entre Estatutos

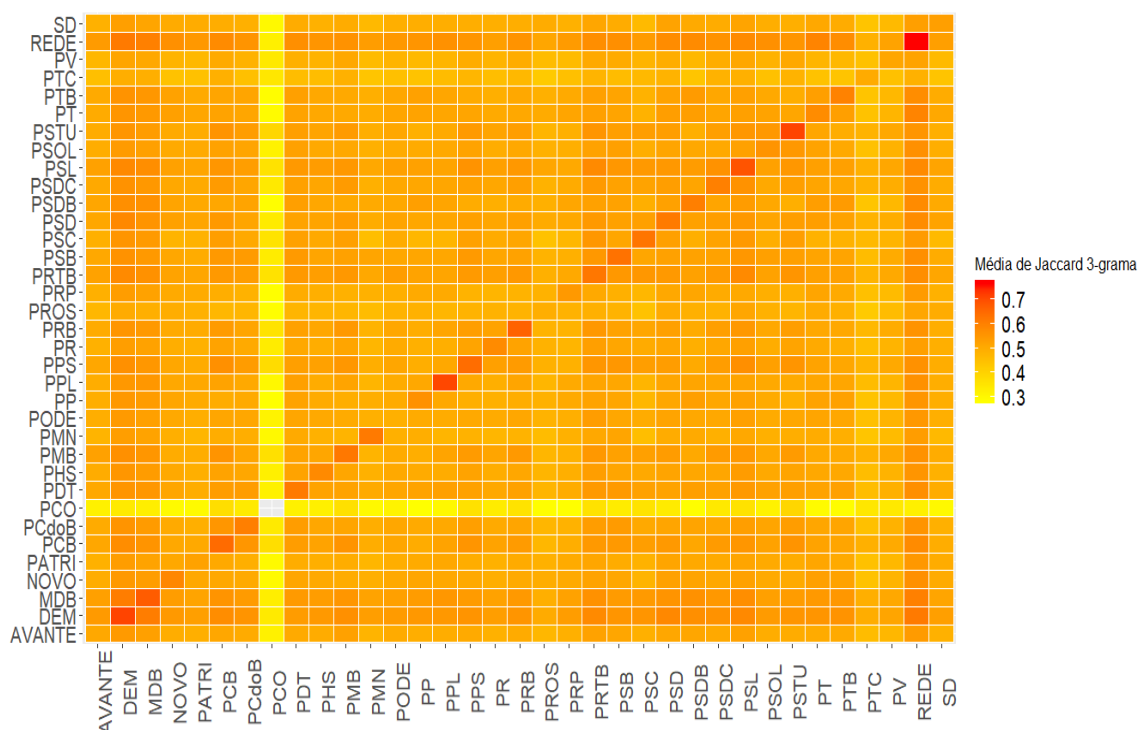
Medida	J2	J3	C2	C3
Min	0.16	0.03	0.01	0.01
Q1	0.59	0.47	0.96	0.89
Mediana	0.66	0.51	0.98	0.92
Média	0.64	0.51	0.95	0.88
Q3	0.69	0.55	0.98	0.95
Max	1.00	1.00	1.00	1.00
Desvio-Padrão	0.13	0.14	0.17	0.22

Figura 8: Densidade Kernel Estatuto x Estatuto



Podemos comparar como a taxa de similaridade varia através dos partidos. Construímos um *heatmap* na Figura 9 mostrando a taxa de similaridade média entre os partidos, calculada pelo índice de Jaccard trígama. Seleccionamos esse índice dentre os demais unicamente pela sua dispersão, que torna mais simples a leitura do mapa de calor. O padrão geral de relação entre os partidos é complexo, contudo conseguimos tirar algumas observações gerais. Primeiramente, vemos que o quadrante PCO x PCO é vazio, visto que tal partido possui apenas um estatuto. Olhando para a diagonal secundária de nossa matriz, percebemos que os partidos possuem taxas de similaridade maiores entre eles mesmos, sendo a REDE, o PPL e o DEM os partidos que possuem maior semelhança interna. Nota-se, igualmente, que alguns partidos sistematicamente são menos similares em comparação aos demais. É o caso do PCO, do PV e, em menor medida, do PTC e SD.

Figura 9: Heatmap de Jaccard 3-Grama Estatuto x Estatuto



Tivemos de ajustar relativamente nossas variáveis de controle para avaliar a similaridade entre estatutos. Para o ano do estatuto, analisamos a Diferença Absoluta no Ano do Estatuto. Similarmente, controlamos pela Diferença Absoluta no Ano do Estatuto. Controlamos a ideologia por uma interação entre a Ideologia do partido “plagiador”(partido A) e do partido “plagiado”(partido B). Similarmente, vimos por nosso heatmap que os estatutos partidários são altamente correlacionados dentro do mesmo partido. A dummy Mesmo Partido captura este efeito. Por fim, controlamos pela taxa de similaridade do estatuto do partido A com o lag do estatuto do partido B.

Assim como vimos em nossos modelos da LOPP, o lag da taxa de similaridade é a variável que melhor explica a similaridade entre es-

tatutos. Tanto a Diferença no Ano do Estatuto quanto a Diferença no Ano de Fundação do Partido se mostraram estatisticamente significativos em nossos quatro modelos. Seu efeito marginal, todavia, é baixo, além de ter o sinal invertido ao esperado nos modelos (7) e (8).

A dummy do mesmo partido é estatisticamente significativa, confirmando o padrão encontrado na Figura 9. A interação entre a ideologia dos partidos revela um padrão interessante. O par esquerda-esquerda é significativo nos modelos (5), (6) e (8), já o par direita-direita para os modelos (5) e (6) somente. Os pares antagônicos, contudo, são significativos e associados positivamente às nossas variáveis dependentes, contrariando nossas expectativas. É possível que isto se deva a um problema de construção da mesma.

Conclusão & Passos Futuros

No presente capítulo buscamos testar o argumento feito por Ribeiro 2012 de que os estatutos partidários brasileiros são muito similares entre si, em parte dado que copiaram trechos inteiros da LOPP. Para tal, coletamos todos os 194 estatutos que os 35 partidos contemporâneos redigiram ao longo de sua existência. Quanto à nossa perspectiva teórica, argumentamos que mais do que simples ‘portas de entrada’ à estrutura organizacional dos partidos políticos, estatutos partidários são objetos em si da realidade política e, portanto, passíveis de análise independente.

Para compararmos nossos textos, utilizamos distintas métricas de detecção de plágio. A utilização de tal técnica está se tornando popular na ciência política, contudo ainda há um debate acerca de que métricas de reutilização textual devem ser empregadas. Argumentamos a favor do uso de n-gramas de caracteres ao invés de n-gramas de palavras. Muito embora o resultado particular da comparação en-

Tabela 4: Modelos OLS para Similaridade Textual dos Estatutos com Estatutos

	<i>Variável Dependente:</i>			
	J2	J3	C2	C3
	(5)	(6)	(7)	(8)
Lag J2	0.348*** (0.005)			
Lag J3		0.286*** (0.005)		
Lag C2			0.326*** (0.005)	
Lag C3				0.301*** (0.005)
Ideologia de Esquerda A	-0.003 (0.002)	-0.003 (0.002)	-0.001 (0.002)	-0.005* (0.003)
Ideologia de Direita A	-0.004** (0.002)	-0.005*** (0.002)	0.0003 (0.002)	0.005** (0.003)
Ideologia de Esquerda B	-0.002 (0.002)	-0.003 (0.002)	0.006** (0.002)	-0.001 (0.003)
Ideologia de Direita B	-0.007*** (0.002)	-0.008*** (0.002)	-0.002 (0.002)	0.005** (0.003)
Mesmo Partido	0.037*** (0.003)	0.084*** (0.002)	0.016*** (0.003)	0.053*** (0.003)
Diff. Ano Estatuto	-0.0004*** (0.0001)	-0.0004*** (0.0001)	0.0004*** (0.0001)	0.0004*** (0.0001)
Diff. Ano Fundação dos Partidos	-0.0002*** (0.0001)	0.0003*** (0.00005)	0.001*** (0.0001)	0.001*** (0.0001)
Esquerda * Esquerda	0.017*** (0.004)	0.015*** (0.003)	0.002 (0.004)	0.014*** (0.005)
Direita * Esquerda	0.007** (0.003)	0.004 (0.003)	0.001 (0.003)	-0.0004 (0.004)
Esquerda * Direita	0.008*** (0.003)	0.006** (0.003)	0.002 (0.003)	-0.0003 (0.004)
Direita * Direita	0.005* (0.002)	0.006** (0.002)	-0.001 (0.002)	-0.001 (0.003)
Constante	0.428*** (0.004)	0.365*** (0.003)	0.638*** (0.005)	0.606*** (0.005)
Observações	34,990	34,990	34,990	34,990
R ²	0.183	0.120	0.112	0.106
R ² Ajustado	0.133	0.120	0.112	0.106

Nota:

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

tre dois documentos varie fortemente conforme a métrica utilizada, argumentamos que este ponto é secundário. Apoiamos o argumento de Garrett e Jansa 2015, que devemos nos preocupar menos com o resultado exato da comparação de apenas um par de documentos, mas sim compreender o índice de maneira relacional, através da variância gerada por todos os documentos analisados.

Ganhamos *insights* importantes acerca da organização dos partidos brasileiros. Quanto à data de criação dos estatutos, vimos que estes estão concentrados em Dezembro de anos não eleitorais. Isto pode implicar que ao alterar os estatutos os dirigentes dos partidos têm em mente as convenções partidárias, contudo mais estudos devem ser feitos para comprovar tal conjectura. A quantidade de palavras por estatuto, igualmente, tem uma tendência crescente ao longo de todo o período analisado. Analisar mais pormenorizadamente as mudanças dos estatutos em cada período, assim, pode ser uma chave para se compreender a dinâmica interna de poder dos partidos brasileiros. Já quanto às palavras específicas utilizadas, verificamos que os termos mais usados têm um bom grau de correspondência com esquemas de classificação manual de estatutos partidários (Von dem Berge et al. 2013).

Expandimos a análise de Ribeiro 2012 para incluir mais casos em um maior período de tempo. Testamos seu argumento da semelhança dos estatutos entre si e com a LOPP. Avaliamos 4 modelos para cada uma dessas análises, variando pelo número de termos analisados (bigramas ou trigramas) e pelo tipo de índice de similaridade utilizado (Jaccard ou cosseno). Para nossos 8 modelos, a similaridade com o estatuto anterior (lag da similaridade) é a variável que melhor prevê a similaridade textual. Confirmamos parcialmente que quanto mais recente for o ano de edição do estatuto menor a similaridade deste com a LOPP. É importante frisar que nossos modelos ainda estão em uma fase muito inicial e seus resultados podem vir a ser alterados

no futuro. Por exemplo, um dos problemas presentes seria da especificação de nossas variáveis (e.g., a legenda partidária) que precisam ser melhor investigados. Outrossim, nosso desenho de pesquisa é intrinsecamente hierárquico. Devemos, portanto, analisar como a performance de modelos multiníveis se dá em relação a estes iniciais modelos lineares.

Referências

- Acree, Brice et al. (2016). *Comparing and Evaluating Cosine Similarity Scores, Weighted Cosine Similarity Scores, & Substring Matching*. Rel. técn. Working Paper, 2016.
- Barrón-Cedeño, Alberto e Paolo Rosso (2009).
“On automatic plagiarism detection based on n-grams comparison”.
Em: *European Conference on Information Retrieval*. Springer. 2009, 696–700.
- Biezen, Ingrid van (2000). “On the Internal Balance of Party Power Party Organizations in New Democracies”. Em: *Party Politics* 6.4, 395–417.
- Biezen, Ingrid van e Daniela Romée Piccio (2013). “Shaping Intra-Party Democracy: On the Legal Regulation of Internal Party Organizations”. Em: 2013, 27–48.
- Bolognesi, Bruno e Adriano Codato (2016).
“Is there a new right coming up in Brazil?: transformations of recruitment patterns in the national political representation”.
Em: *VIII Congresso Associação Portuguesa de Ciência Política, Lisbon*. 2016.
- Calossi, Enrico e Eugenio Pizzimenti (2015).
“Party Organizational Change: Formal Distribution of Power between National and Regional Levels in Italian Political Parties (1991-2012)”.
Em: *Partecipazione e Conflitto* 8.1, 167–189.
- Coppedge, Michael (1997). “A classification of Latin American political parties”.
Em:
- Fernandes, Luis (1995). “Muito barulho por nada? O realinhamento político-ideológico nas eleições de 1994”. Em: *Dados* 38.1, 107–144.
- Freidenberg, Flavia e Steven Levitsky (2007).
“Organización informal de los partidos en América Latina”.
Em: *Desarrollo Económico* 184, 539.

- Garrett, Kristin N e Joshua M Jansa (2015).
 “Interest group influence in policy diffusion networks”.
 Em: *State Politics & Policy Quarterly* 15.3, 387–417.
- Grimmer, Justin e Brandon M Stewart (2013). “Text as data: The promise and pitfalls of automatic content analysis methods for political texts”.
 Em: *Political analysis* 21.3, 267–297.
- Hartmann, Paul (1931).
Die politische partei in der Tschechoslowakischen republik; eine juristische studie.
 Vol. 2. RM Rohrer, 1931.
- Hazan, Reuven Y e Gideon Rahat (2010). *Democracy within parties.*
 Oxford University Press, 2010.
- Ignazi, Piero, Luciano Bardi e Oreste Massari (2010).
 “Party organisational change in Italy (19912006)”.
 Em: *Modern Italy* 15.2, 197–216.
- Ignazi, Piero e Eugenio Pizzimenti (2014).
 “The Reins of Intra-Party Power in The Italian Political Parties (1990-2011)”.
 Em: *Rivista italiana di scienza politica* 3.3, 223–246.
- Katz, Richard S (1992). *Party organizations: a data handbook on party organizations in western democracies, 1960-90.* Vol. 1.
 Sage Publications Ltd, 1992.
- (2002). “The internal life of parties”.
 Em: *Political challenges in the new Europe*, 87–118.
- Katz, Richard S. e Peter Mair (1994). *How Parties Organize: Change and Adaptation in Party Organizations in Western Democracies.* 1994.
- Kestemont, Mike (2014).
 “Function words in authorship attribution. From black magic to theory?”
 Em: *3rd Workshop on Computational Linguistics for Literature (CLFL)*. 2014, 59–66.
- Madeira, Rafael Machado e Gabriela da Silva Tarouco (2013).
 “Esquerda e direita no sistema partidário brasileiro: análise de conteúdo de documentos programáticos”. Em: *Revista Debates (UFRGS)*.
- Mainwaring, Scott et al. (2000). *Partidos conservadores no Brasil contemporâneo.*
 Paz e Terra, 2000.
- National Democratic Institute, NDI (2011). *A compilation of political party statutes.*
 2011.
- Norris, Pippa (2004). “Building political parties”. Em: *International IDEA*.

- Ooms, Jeroen (2017).
 “pdftools: Text extraction, rendering and converting of pdf documents”.
 Em: *Computer software manual*.
- (2018a). *hunspell: High-Performance Stemmer, Tokenizer, and Spell Checker*.
 R package version 3.0. 2018.
<https://CRAN.R-project.org/package=hunspell>.
- (2018b). *magick: Advanced Graphics and Image-Processing in R*.
 R package version 2.0. 2018.
<https://CRAN.R-project.org/package=magick>.
- (2018c). *tesseract: Open Source OCR Engine*. R package version 4.0. 2018.
<https://CRAN.R-project.org/package=tesseract>.
- Panebianco, Angelo (2005). *Organização e poder nos partidos políticos*. 2005.
- Ribeiro, Pedro Floriano (2012). “Organização e poder nos partidos brasileiros: uma análise dos estatutos (Organization and Power within Brazilian Political Parties: An Analysis of Party Statutes)”.
 Em: *Revista Brasileira de Ciência Política* 10, 225–265.
- Ribeiro, Pedro Floriano e Elodie Fabre (2019). “Multilevel party organizations in a fragmented presidential system: The case of Brazil”.
 Em: *Regional Federal Studies*, 1–31.
- Silva Ramos, Elival da (2008). “O delineamento do estatuto dos partidos políticos na jurisprudência do Supremo Tribunal Federal”.
 Em: *Revista de Direito Administrativo* 248, 54–72.
- Silva Tarouco, Gabriela da e Rafael Machado Madeira (2015).
 “Os partidos brasileiros segundo seus estudiosos. Análise de um expert survey”.
 Em: *Civitas-Revista de Ciências Sociais* 15.1, e24–e39.
- Smith, Ray (2007). “Tesseract ocr engine”. Em: *Lecture. Google Code. Google Inc.*
- Stamatatos, Efstathios (2009).
 “Intrinsic plagiarism detection using character n-gram profiles”.
 Em: *threshold* 2.1,500.
- Von dem Berge, Benjamin et al. (2013). *Measuring intra-party democracy*.
 Springer, 2013.
- Wiesehomeier, Nina e Kenneth Benoit (2007). “Presidents and Parties in Policy Space: Expert surveys of policy positions in 18 Latin American countries”. Em: 2007.
- Zucco Jr, Cesar (2011).
 “Esquerda, direita e governo: a ideologia dos partidos políticos brasileiros”.
 Em: *O Congresso por ele mesmo: autopercepções da classe política brasileira*.
 UFMG, 2011, 37–60.