

Avaliando Estatutos Partidários Brasileiros através de Análise Quantitativa de Texto*

Lucas Malta Mingardi[†]

Maio 2019

Resumo

Uma das formas mais comuns de se estudar a estrutura da organização interna dos partidos políticos é analisar seus estatutos partidários. Mas mais do que simples documentos, estatutos partidários são objetos políticos em si e, portanto, passíveis de serem estudados isoladamente. No presente artigo avaliaremos o argumento presente na literatura brasileira acerca da semelhança dos estatutos entre si e entre a Lei Orgânica dos Partidos Políticos (Lei 5.682/1971). Para testar tais hipóteses, selecionamos 195 estatutos de 35 partidos políticos brasileiros. Para dar conta de tamanho volume de estatutos, tivemos de recorrer a técnicas de análise quantitativa de texto ainda pouco exploradas na ciência política brasileira. Um segundo objetivo que temos, portanto, seria difundir tais práticas de pesquisa. Utilizamos técnicas de *optical character recognition* (OCR) para extração dos dados, bem como debatemos acerca do uso de distintas formas de análise de reutilização textual ('análise de plágio') como métrica de comparação dos estatutos. Discutimos ainda os limites e desafios de se estudar estatutos partidários, assim como buscamos investigar fatores que afetariam a taxa de semelhança textual dos estatutos.

Keywords: Estatutos Partidários, Organização Partidária, Reconhecimento Óptico de Caracteres, Análise de Reutilização de Texto.

*Trabalho preparado para apresentação no IX Seminário Discente da Pós-Graduação em Ciência Política da USP, de 6 a 10 de maio de 2019.

[†]Doutorando pelo Departamento de Ciência Política da Universidade de São Paulo.

1 Introdução

Uma forma tradicional de se sistematizar os estudos dos partidos políticos (Key et al. 1955), é feita pela diferenciação dos campos de atuação dos partidos. Podemos, assim, estudar os partidos no eleitorado, os partidos nos governos e, por fim, os partidos enquanto organizações. O presente artigo se insere neste último campo, isto é, do estudo das organizações partidárias. Para fins de definição, entendemos estatutos partidários como “um conjunto de regras escritas pelas quais partidos políticos estruturam e organizam suas atividades” (National Democratic Institute 2011).

A principal vantagem de se estudar estatutos partidários no campo da organização partidária seria seu relativo fácil acesso. Muitos estudos, de fato, partem da leitura de estatutos para então depreender características gerais da estrutura interna dos partidos políticos (e.g., Katz 2002; Freidenberg e Levitsky 2007; Hazan e Rahat 2010, dentre outros). Isso não quer dizer, todavia, que a análise dos estatutos partidários está livre de problemas. Um dos argumentos mais fortes contrários à sua análise é que estatutos seriam pouco informativos acerca da realidade das organizações partidárias. Estatutos seriam, portanto, apenas um documento pró-forma, criados apenas para que os partidos entrem em conformidade com a lei. Aos estatutos partidários escapariam não só instituições informais associadas à política (Norris 2004), mas também, mesmo quanto às instituições formais, haveria um descompasso entre a forma pela qual as regras internas de organização são redigidas nos estatutos *vis-à-vis* como a estrutura organizativa dos partidos é dada ‘de fato’. Tal descompasso seria maior, como aponta Panebianco, quanto menos institucionalizado for o partido (Panebianco 2005, pp. 110–111). Por conseguinte, o pesquisador que viesse estudar a estrutura organizativa dos partidos políticos unicamente a partir de seus estatutos estaria potencialmente enviesando sua análise (Freidenberg e Levitsky 2007).

A essa visão apresentamos dois contrapontos. Primeiramente, como apontam Freidenberg e Levitsky 2007, p. 545, para se medir a distância entre as regras formais e informais de dado partido, é preciso anteriormente ter a informação de como se operam tais regras formais.

Em segundo lugar, argumentamos que estatutos partidários são parte da realidade política, logo passíveis de serem estudados enquanto objetos independentes. É possível se desvincular, portanto, o estudo dos estatutos partidários da mensuração do grau de similitude das regras descritas nestes e as práticas realizadas dentro dos partidos, e estudá-los de forma autônoma. Deixamos de interpretar a leitura dos estatutos como meio, para se chegar à compreensão de algum fenômeno da estrutura organizativa dos partidos, para compreendê-la como um fim em si mesmo.

No presente artigo, estamos interessados em estudar especificamente os estatutos dos partidos brasileiros contemporâneos. Nosso universo de análise é composto de 195 estatutos feitos por 35 partidos distintos, redigidos ao longo do período que vai de 1995 a 2018. A nosso ver, a principal vantagem de nos centrarmos no caso brasileiro seria que temos um número suficiente de observações a serem comparadas sob um único regime político. Diferentemente de outros estudos (e.g., Norris 2004), portanto, não necessitamos isolar efeitos gerados pelo sistema-político partidário, e podemos comparar os estatutos diretamente entre si.

Assim como vemos na literatura internacional, boa parte dos estudos brasileiros partem da leitura dos estatutos partidários para assim se depreender o funcionamento da estrutura organizativa dos partidos de maneira mais geral. Vemos, por exemplo, estudos que focam em como se dá o processo interno de seleção de candidatos (e.g., Álvares 2006; Braga 2008; Braga e O. E. d. Amaral 2013; Bolognesi 2013; Dantas e Roeder 2015), o financiamento político (Braga e Bourdoukan 2009), ou mesmo a estrutura das relações de poder (Guarnieri 2009; Ribeiro 2013; Ribeiro 2014). Outros artigos se centram ainda no estudo de uns poucos partidos para compreender suas estruturas organizativas de maneira mais abrangente (Meneguello e O. Amaral 2008; O. M. E. d. Amaral 2010; Mayer 2011).

Poucos são os estudos, contudo, que tratam dos estatutos partidários enquanto objetos próprios, independentes, de análise. O estudo que mais encaixa nesse perfil, o principal impulsionador do presente artigo, é o trabalho de Ribeiro 2012. A nosso ver, Ribeiro forneceu duas importantes contribuições para a compreensão da dinâmica dos estatutos brasileiros.

Primeiramente, temos sua conclusão de que, com a exceção do PT, os estatutos de todos os partidos analisados pelo autor¹ seguem a estrutura organizativa desenhada pela Lei Orgânica dos Partidos Políticos². Como aponta o autor, até 1995 os partidos não eram obrigados a ter um estatuto, com sua estrutura interna sendo organizada pela LOPP. Com a criação da Lei dos Partidos Políticos³, a criação de estatutos passou a ser obrigatória aos partidos⁴, sendo estes autônomos para desenhar sua estrutura organizativa. Apesar dessa autonomia ganha pelos partidos, como nota Ribeiro 2012, p. 238, “os documentos de diferentes partidos não só continuaram idênticos entre si em muitos trechos: eles também seguem reproduzindo exatamente artigos inteiros fornecidos pela LOPP em 1971”. Em segundo lugar, vemos em Ribeiro uma importante preocupação em analisar os múltiplos estatutos redigidos ao longo do tempo, buscando entender o que estava em jogo em cada mudança.

Podemos, então, generalizar os achados de Ribeiro 2012, 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?; e finalmente 3) afinal, o que podemos aprender acerca dos estatutos partidários brasileiros de maneira mais abrangente? Fornecemos aqui novos métodos para avaliar tais problemas.

Como estamos lidando com um extenso volume de estatutos partidários, uma questão óbvia seria a grande dificuldade de se avaliar cada um desses estatutos individualmente. Para contornar tal problema, tivemos de recorrer a ferramentas de análise quantitativa de texto para darmos conta desse extenso volume de informações. Uma observação que devemos fazer é que, ao tratar de tais técnicas, não devemos cair na armadilha conceitual de gerarmos um embate acerca de métodos ‘quantitativos’ versus métodos ‘qualitativos’, como

¹Sendo estes: PMDB (atual MDB), PSDB, PT e PFL/DEM.

²Lei 5.672/1971, também conhecida pelo acrônimo LOPP.

³Lei 9.096/1995

⁴Para 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 lei (Art. 55 §1). Publicada em 19/03/1995, isso significa que os partidos teriam até Março de 1996 para se adequar às novas medidas.

se estes representassem lados antagônicos de um debate. Parafraseando Grimmer e Stewart 2013, p. 268, técnicas de mineração de texto não vieram para ‘substituir’ a leitura acurada feita manualmente, mas sim para complementá-la. Se, por um lado, perdemos informações geradas por *insights* que somente a leitura cuidadosa de um texto pode propiciar, por outro, conseguimos ganhar maior alcance da quantidade de textos lidos. Devemos, assim, entender esses dois modos de analisar textos como métodos complementares, capazes de ampliar e ressignificar os resultados que obtemos.

Para compararmos os estatutos partidários brasileiros, incorporamos técnicas de análise de reutilização textual ⁵ em nossa análise. Apesar de relativamente recente, a análise de plágio vem ganhando força na ciência política como critério de comparação textual. Acree et al. 2016 fornecem um bom ponto de estarte, mostrando como diferentes campos na ciência política utilizam tal técnica para avaliar seus respectivos objetos de pesquisa. Apesar de haver várias métricas distintas de análise de plágio⁶, o que há de comum entre as principais ferramentas é que seu resultado costuma cair em um intervalo que vai de 0 a 1, com 0 significando que não há nenhum elemento comum entre os textos, e 1 significando que os textos são idênticos. Garrett e Jansa 2015 mostram como é mais natural entender esse resultado como uma medida relativa de similaridade entre dois textos. Ao tratar tais dados como medidas relativas, isso significa que não precisamos nos focar tanto em um problema de distinguir similaridades ‘reais’ *versus* e similaridades ‘coincidentes’ de um artigo ⁷.

O presente artigo é dividido em 4 seções, além da Introdução e Conclusão. Na seção 2, apresentamos nossos dados e métodos. Na Seção 3, faremos uma análise exploratória dos estatutos brasileiros a partir de técnicas de *text mining*, buscando detectar alguns padrões

⁵Também 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’ e ‘plágio’ intercambiadamente

⁶Debateremos melhor acerca de como tais métricas diferem entre si na seção seguinte.

⁷Podemos imaginar que haveria uma taxa de similaridade ‘natural’ entre textos jurídicos, dado que tratam de assuntos similares, bem como utilizam uma forma bem específica de linguagem. De fato, quando olhamos para o caso específico de estatutos, vemos que mesmo a leitura manual é capaz de detectar padrões; trabalho este feito, por exemplo, por Norris 2004 para criar dados comparativos entre estatutos. Todavia, ao tratarmos as taxas de similaridade como medidas *relativas*, isso faz com que nos preocupemos mais com a variância de nossa variável do que com o resultado pontual de cada variável-resposta.

no corpo textual dos estatutos. Na seção 4 apresentamos os resultados que obtivemos da análise de reutilização textual dos estatutos em relação à LOPP. Tentamos ainda criar alguns modelos estatísticos que expliquem o que determina a similaridade entre esses textos. Já na Seção 5, aplicamos um método semelhante para analisarmos a similaridade dos estatutos entre si, sejam estes do mesmo partido ou de partidos diferentes.

2 Materiais & Métodos

Temos como universo de análise todos os estatutos dos partidos brasileiros contemporâneos⁸. Incluímos não só os estatutos mais recentes de cada partido, mas também todos seus estatutos prévios. Com isso, temos 195 estatutos distintos para 35 partidos políticos. Excluimos de nossa análise todos outros tipos de normas partidárias, tais como instruções normativas e resoluções. Excluimos também os estatutos de partidos criados antes da redemocratização brasileira, ou de partidos pós-redemocratização que por ventura foram extinguidos ou incorporados a outros partidos. Incluímos, contudo, estatutos de partidos que mudaram a denominação de sua legenda, como por exemplo o PFL, que passou a ser chamado de DEM a partir de 2007. Os estatutos partidários foram coletados no *site* do Tribunal Superior Eleitoral (TSE)⁹. As restrições em universo de análise foram impostas menos por nossas próprias decisões, e mais pelo volume de informações disponibilizadas pelo TSE .

Nosso primeiro passo foi coletar nossos dados brutos, i.e. os estatutos dos partidos brasileiros. Para tal, utilizamos de ferramentas de raspagem de dados para filtrar e baixar os dados que necessitamos. O processo seguinte foi de transformar os arquivos pdf que temos em dados. Felizmente, o R apresenta algumas soluções interessantes para esse problema¹⁰. O pacote `pdftools` (Ooms 2017) utiliza ferramentas de extração de texto para ler, ao utilizá-lo,

⁸Sendo 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.

⁹<http://www.tse.jus.br/partidos/partidos-politicos/registrados-no-tse>.

¹⁰Conseguimos ler 194 dos 195 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.

conseguimos extrair o texto de 163 dos 195 estatutos através desse método. Para cerca de 20% de nossos dados restantes, tivemos de usar técnicas de reconhecimento óptico de caracteres (*optical character recognition*). Por se tratar de uma engine open source, 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) ¹¹.

Após termos nosso banco de dados pronto, o passo seguinte foi de limpar o corpo dos textos. Assim sendo, removemos números, sinais de pontuação (com exceção de hifens intra-palavras) e *stopwords* ¹², bem como tiramos a capitalização de todas as palavras. Em seguida, realizamos um *stemming* ¹³ 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.

Algumas observações devem ser feitas quanto às técnicas de detecção de plágio que utilizamos. Há várias métricas distintas de detecção de plágio disponíveis, que em última análise dependem de como o autor classifica o que entende como ‘plágio’ em si. O primeiro passo advém de se decidir como o pesquisador tratará seus textos: se podem ser vistos como uma ‘bag of words’ (Grimmer e Stewart 2013), ou se localizar as passagens plagiadas importa, i.e., se calculamos o plágio pelo texto inteiro ou se o seccionamos por sentenças. Decidimos pelo primeiro para nossa presente pesquisa. O próximo passo seria como definimos nossos termos. A forma mais óbvia seria identificar os termos como as palavras que compõem um documento. De fato, todos os cientistas políticos citados por Acree et al. 2016 utilizam tal

¹¹Um 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. Todavia, como queremos ter uma fac-símile editável dos estatutos partidários, futuramente teremos de contornar melhor tal problema.

¹²Isto é, palavras comuns, que não possuem nenhum significado ‘próprio’, mas sim que são utilizadas para desempenhar alguma função gramatical, como, por exemplo, artigos e preposições.

¹³Isto é, reduzimos as palavras ao seu radical comum. Por exemplo, ‘município’, ‘municípios’ e ‘municipal’ e ‘municipais’ são reduzidos todos para o radical comum ‘municip’.

estratégia. Todavia, há estratégias mais efetivas do que esta. Desde meados dos anos 1990 a análise de caracteres vem ganhando força no campo. Sua principal vantagem seria que, em contraposição a análise de palavras, pode ser utilizada independentemente da língua do texto, não necessita necessariamente de nenhum pré-processamento de palavras, e tende a ser robusta em relação a ruídos (Barrón-Cedeño e Rosso 2009; Stamatatos 2009; Kestemont 2014). Após isso, decidimos acerca de quantos termos consecutivos analisamos, i.e., qual a dimensionalidade de nossos n-gramas. Como Barrón-Cedeño e Rosso 2009 argumentam, bigramas e trigramas são os n-gramas ótimos para a comparação de *corpora*. Conforme os autores (Barrón-Cedeño e Rosso 2009, p. 699), há um *trade-off* entre *recall* e precisão, caso o pesquisador opte por bigramas ou trigramas, respectivamente. N-gramas de maior dimensionalidade não apresentam ganho real de eficiência, enquanto que 1-gramas têm uma performance pior, especialmente quando são analisados caracteres e não palavras. Utilizaremos, assim, 2-gramas e 3-gramas de caracteres para nossa análise ¹⁴. A partir desses elementos conseguimos criar nossa matriz de termos (‘document term matrix’), que nada mais é do que uma matriz que sumariza a quantidade de vezes que cada termo (ou, mais especificamente, que cada n-grama) aparece em cada documento. 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:

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

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

¹⁴Para 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.

n-gramas em A e B, e não quantas vezes tais n-gramas aparecem em cada documento. Já a similaridade de cosseno é definida por:

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

Ou seja, calculamos o cosseno do ângulo de dois vetores da nossa matriz de termos. Temos, assim, 4 testes de similaridade textual, a depender do n-grama que escolhermos (2-grama e 3-grama) e do tipo de análise de similaridade (Jaccard e Cosseno). O objetivo de operarmos com esses quatro testes se dá para garantir maior robustez à nossa análise.

Para trabalhar com os modelos que analisamos nas seções 4 e 5, criamos também algumas variáveis independentes. O partido ao qual o estatuto pertence, bem como a data de redação do estatuto foi obtida a partir da leitura dos próprios estatutos. A data de fundação dos partidos políticos foi obtida através do TSE. A ideologia do partido foi criada a partir da comparação de vários estudos que trabalham com tal variável (Fernandes 1995; Coppedge 1997; Mainwaring et al. 2000; Wiesehomeier e Benoit 2007; Zucco Jr 2011; Madeira e Silva Tarouco 2013; Bolognesi e Codato 2016).

3 (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 artigo 195 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

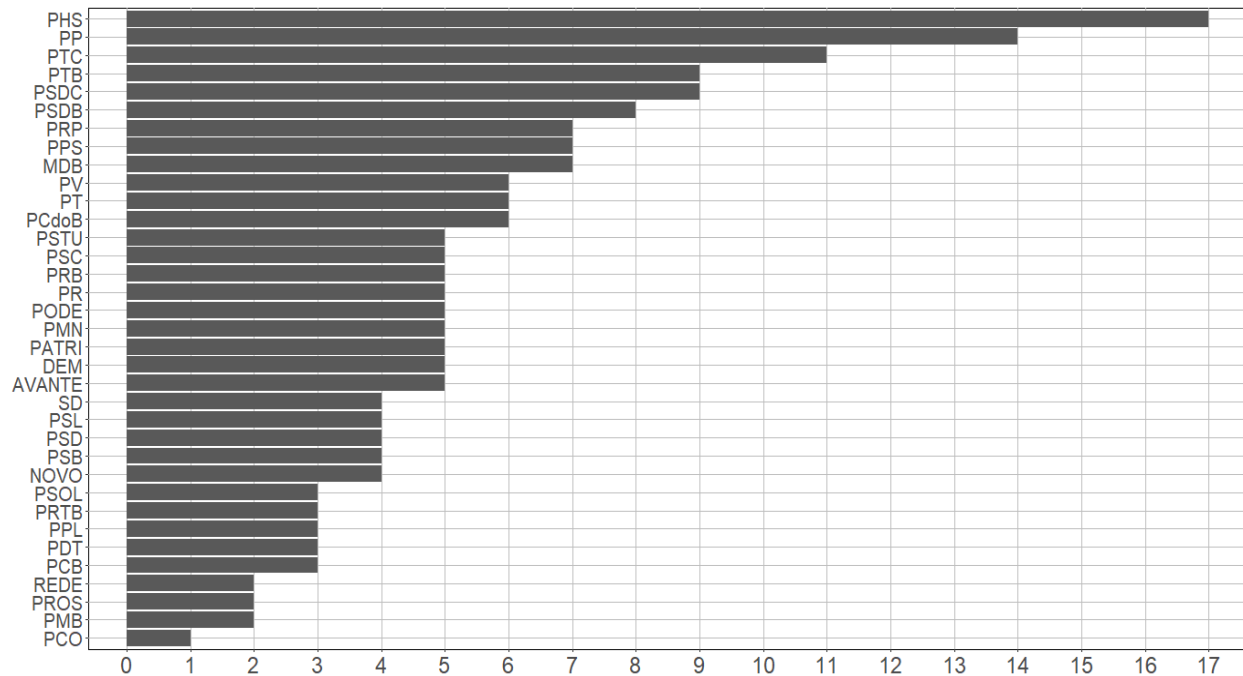


Figura 1: Número de Estatutos por Partido

mediana 5.

Separando os estatutos por ideologia, criamos a Tabela 1. Partidos de esquerda, no geral, costumam criar menos estatutos, seguidos por partidos de centro e finalmente partidos de direita. Como poderia ser imaginado, as estatísticas pontuais para os partidos de direita estão sendo puxadas pelo excessivo número de estatutos do PHS e PP.

Tabela 1: Estatutos por Ideologia

Ideologia	Média	Mediana	Total
Esquerda	4.10	3.5	41
Centro	5.18	5.0	57
Direita	6.85	5.0	96

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

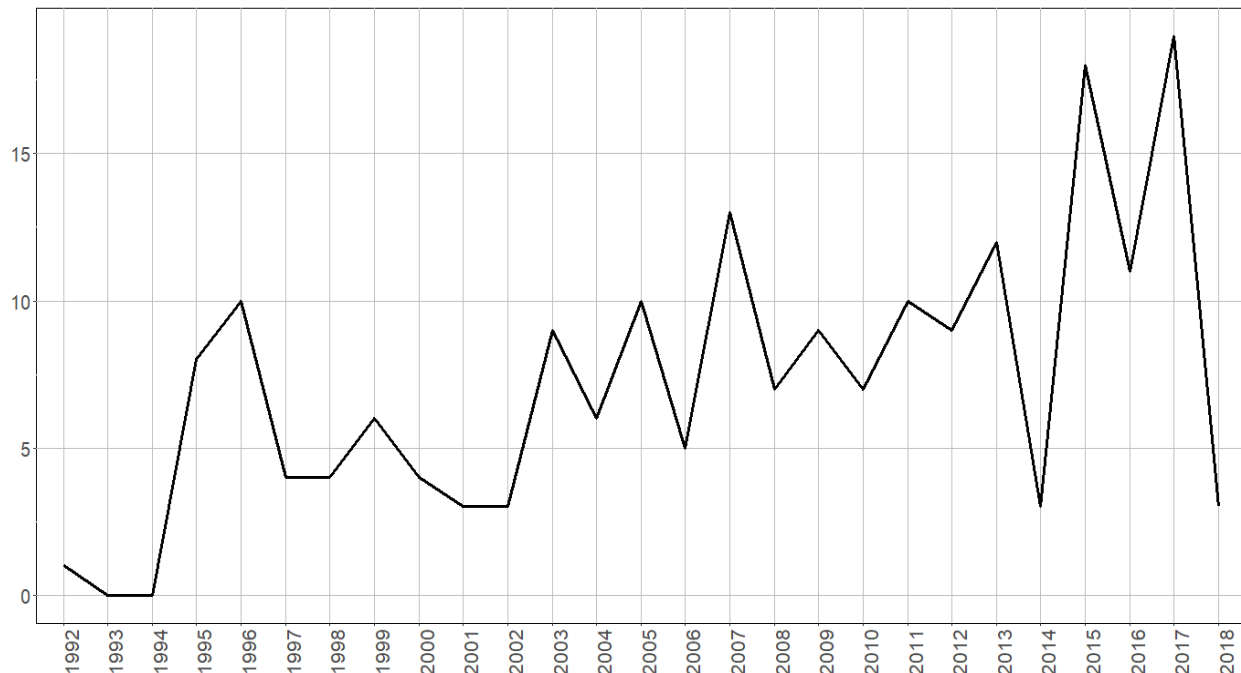


Figura 2: Número de Estatutos por Ano

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.

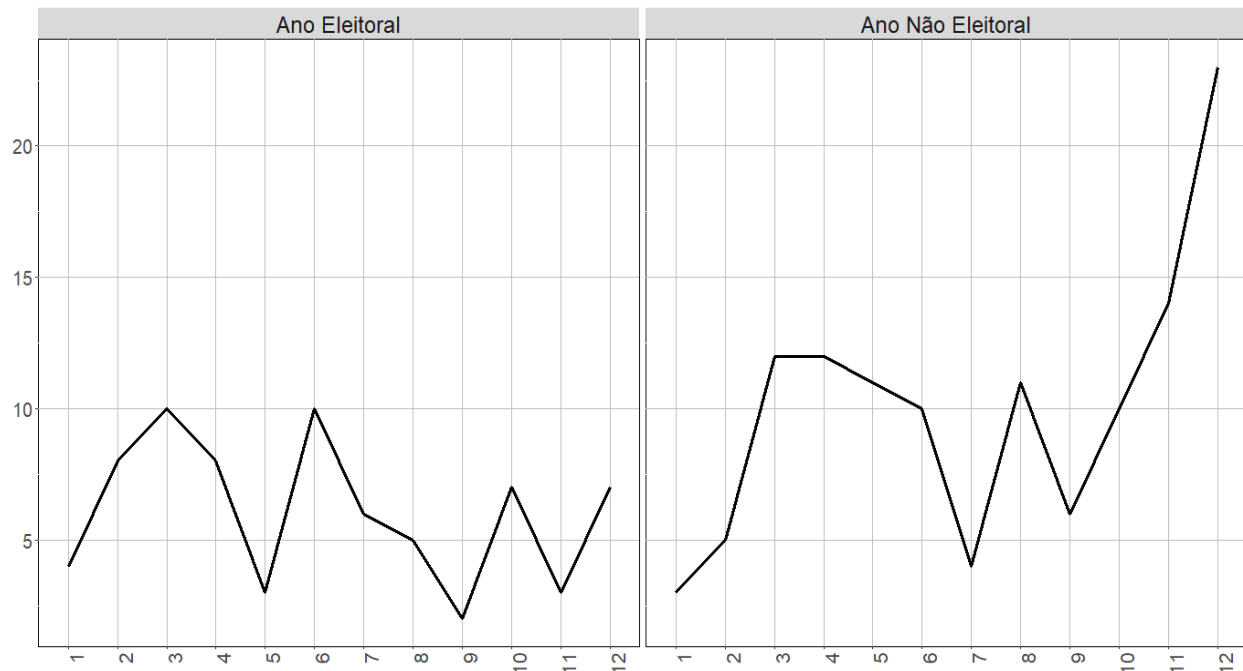


Figura 3: Número de Estatutos por Mês

Passemos agora a investigar o conteúdo propriamente dito dos estatutos brasileiros. Uma primeira incursão a esses dados seria verificar o número de palavras que cada estatuto tem. Olhando para os dados gerais, o número médio de palavras por estatuto foi de 7173.1 palavras (incluindo-se palavras repetidas), com um desvio-padrão de 3728.1 palavras. A Figura 4 mostra a densidade kernel do número de palavras dos estatutos, desagregado por partido. As cores indicam a ideologia do partido, vermelho sendo para partidos de esquerda, verde para partidos de centro, e azul para partidos de direita. Tivemos de remover desse gráfico o PCO, PMB, PROS e REDE, por não contarem com estatutos suficientes para se calcular a densidade separadamente. Quanto à ideologia, não conseguimos detectar algum padrão que distingue os políticos quanto à sua distribuição de quantidade de palavras. A distribuição varia bastante entre um partido e outro, com partidos como o PPS, PMN e SD sendo unimodais e normais, enquanto que partidos como o PRP, PATRI e PP apresentando uma distribuição multimodal.

Ainda quanto o PP, vemos preocupantemente que um de seus picos está quase no eixo 0. Olhando mais atentamente para estes arquivos, vemos que são os estatutos de 1996, 1998 e

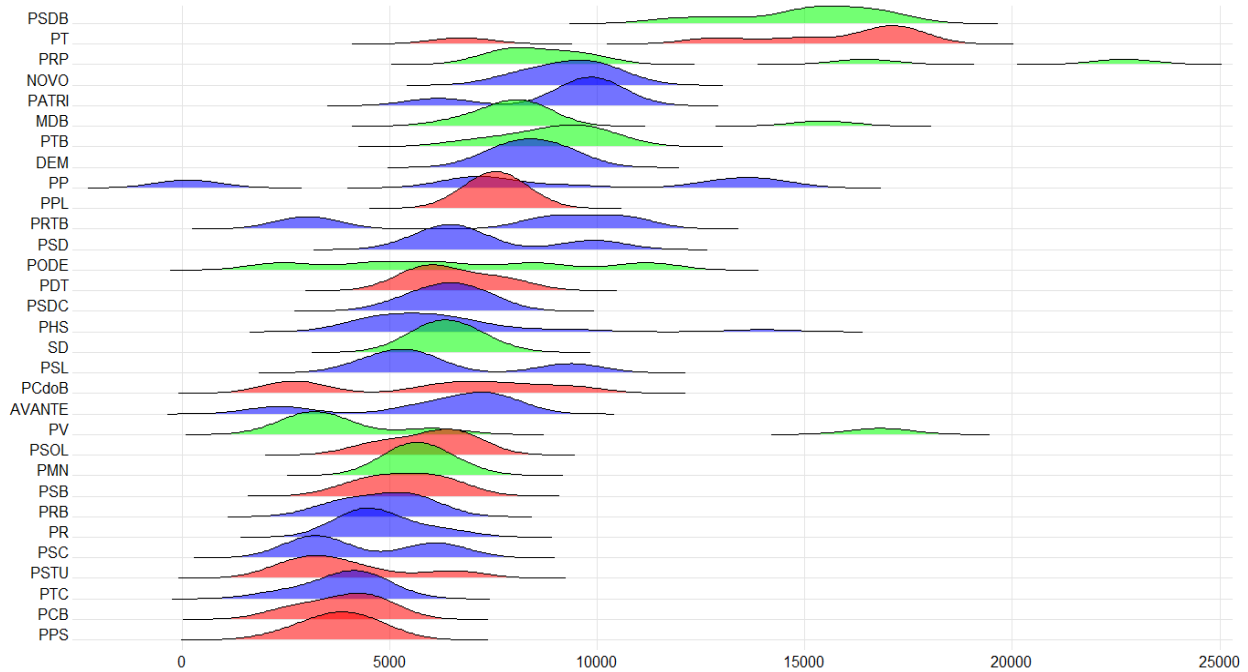


Figura 4: Densidade Kernel do Número de Palavras por Partido

1999 do então PPB, com, respectivamente, 63, 85 e 271 palavras. Analisando tais arquivos, verificamos que tais documentos de fato não são estatutos por inteiro, mas tão somente notas de que artigos de seus estatutos estavam sendo alterados após acordo em convenção. Como tais dados acabariam produzindo ruído em nossas análises futuras, por não serem de fato estatutos partidários, decidimos por remover esses documentos dos modelos apresentados nas Seções 4 e 5. Já o estatuto com o maior número de palavras foi o estatuto de 2016 do PRP, com 22703 palavras.

O processo de reedição de estatutos é incremental, assim é plausível supor que o número de palavras dos estatutos aumenta à medida que os anos passem. Testamos tal hipótese na Figura 5. 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.

Gráfico 5: Número Médio e Mediano de Palavras por Ano

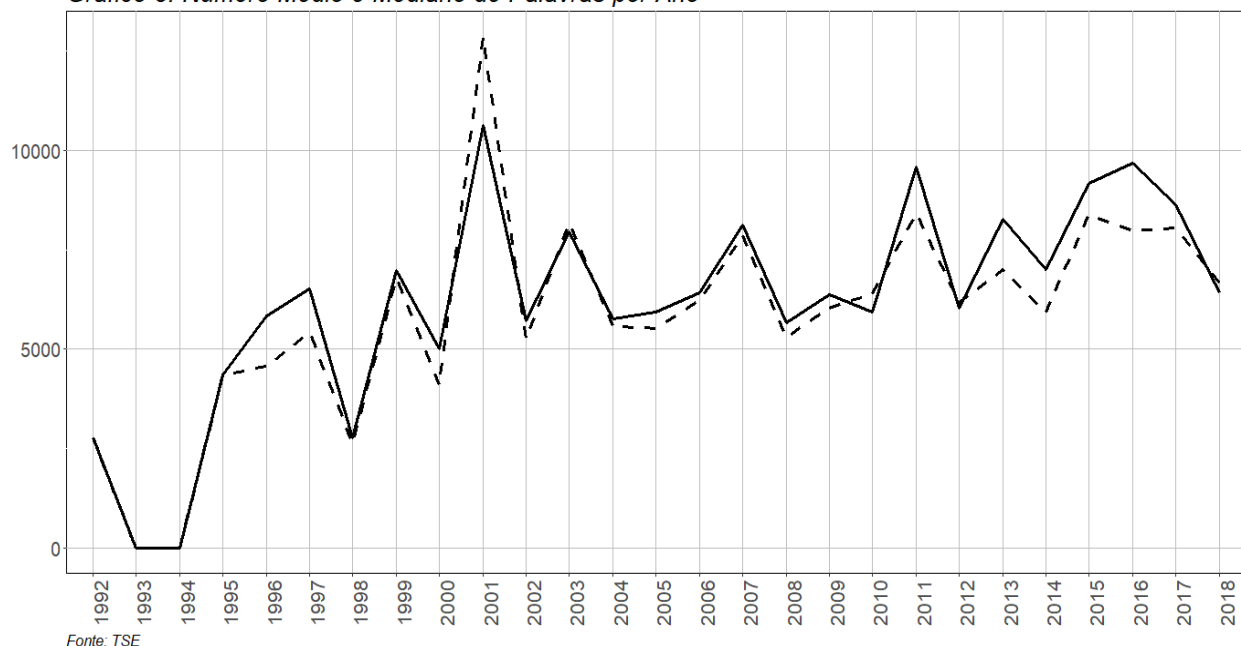


Figura 5: Número Médio e Mediano de Palavras por Ano

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 6. 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)¹⁵ 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’ (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

¹⁵Somamos 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.

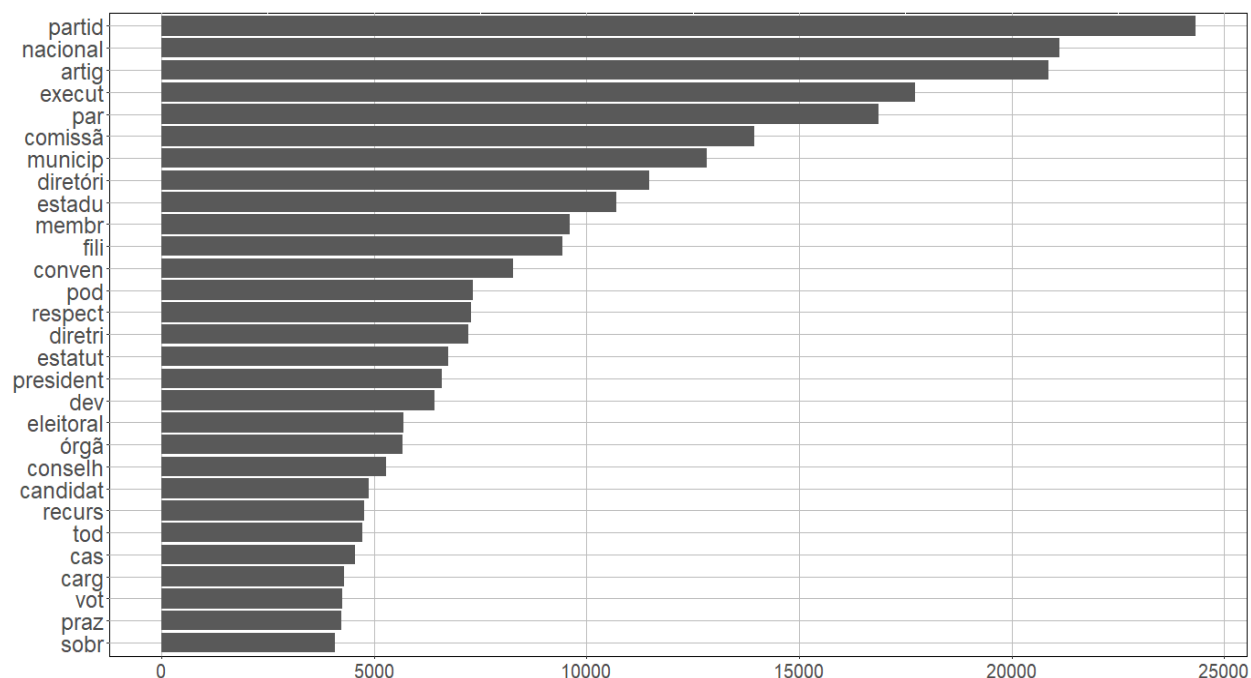


Figura 6: Palavras mais Comuns nos Estatutos

proposta por Norris 2004, o que demonstra como a análise quantitativa de texto pode ser útil para auxiliar métodos mais qualitativos de análise.

4 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. Analisamos aqui 192 estatutos, visto que removemos de nossa análise o estatuto de 1990 do PV, por não conseguirmos lê-lo; bem como os estatutos do PP de 1996, 1998 e 1999, por não serem propriamente estatutos. Utilizamos 4 taxas de similaridade de caractere, a depender do tamanho do n-grama (bigramas e trigramas) e do critério de similaridade (Jaccard ou Cosseno).

Antes de passarmos para a análise de nosso modelo, contudo, olhemos rapidamente como se dá a distribuição de nossas variáveis dependentes. Na Figura 7 as curvas de linha tracejada se referem às taxas de similaridade que empregam 2-gramas, enquanto as linhas sólidas são

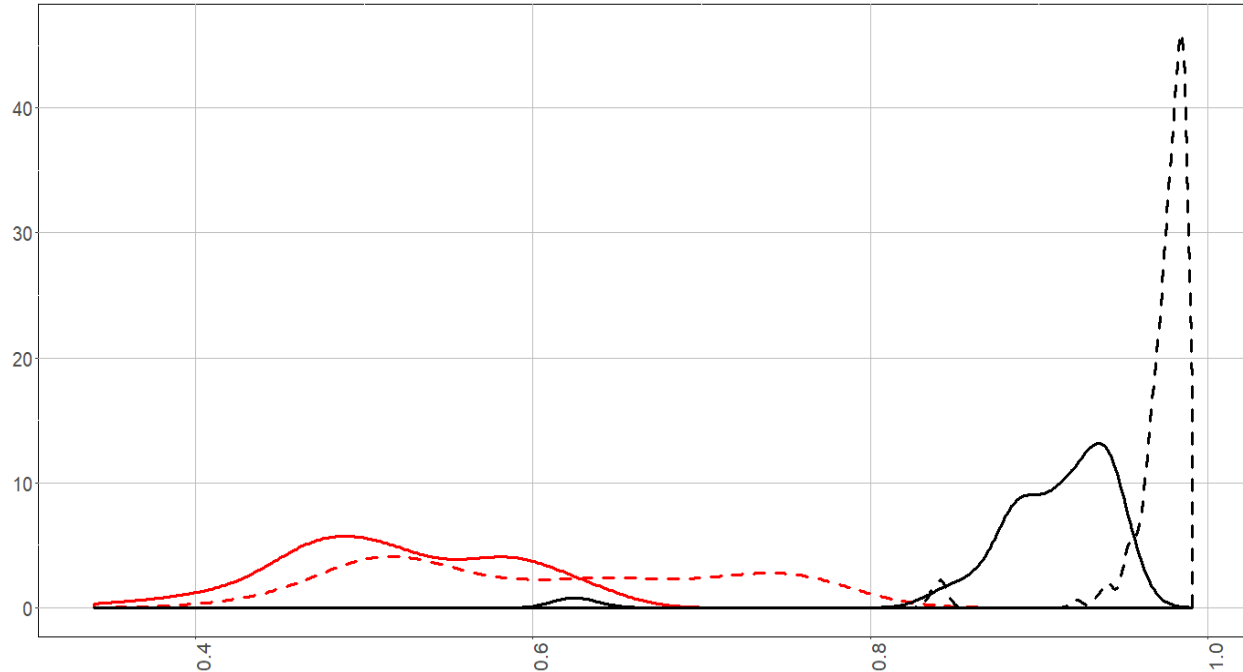


Figura 7: Densidade Kernel Estatuto x LOPP

de 3-gramas. As curvas vermelhas são dos índices de Jaccard, e as curvas pretas são a taxa de cosseno. Os índices de Jaccard sistematicamente dão taxas de similaridade menores do que os índices de cosseno. 2-gramas têm também menor obliquidade e maior curtose.

Podemos também observar a diferença dessas distribuições através de suas estatísticas pontuais. Resumimos algumas informações na Tabela 2:

Tabela 2: Medidas Pontuais das Taxas de Similaridade com a LOPP

Medida	Jaccard 2-grama	Jaccard 3-grama	Cosseno 2-grama	Cosseno 3-grama
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

Por fim, antes de analisarmos nosso modelo, podemos ver como a taxa de similaridade com a LOPP é distribuída pelos partidos. Na Figura 8, plotamos o índice 3-grama de Jaccard por partido. Escolhemos esse índice para projetar ao invés dos outros por causa de sua variância geral; apesar da forma específica das distriuições em cada índice, vemos que a relação entre os arquivos se mantém relativamente constante em cada métrica. Novamente, a cor da distribuição representa a ideologia política, de acordo com o sistema de classificação usualmente utilizado pela literatura. O padrão da taxa de similaridade entre os estatutos e a LOPP, todavia, parece depender muito mais do contexto próprio dos partidos do que com alguma forma de se classificar os mesmos.

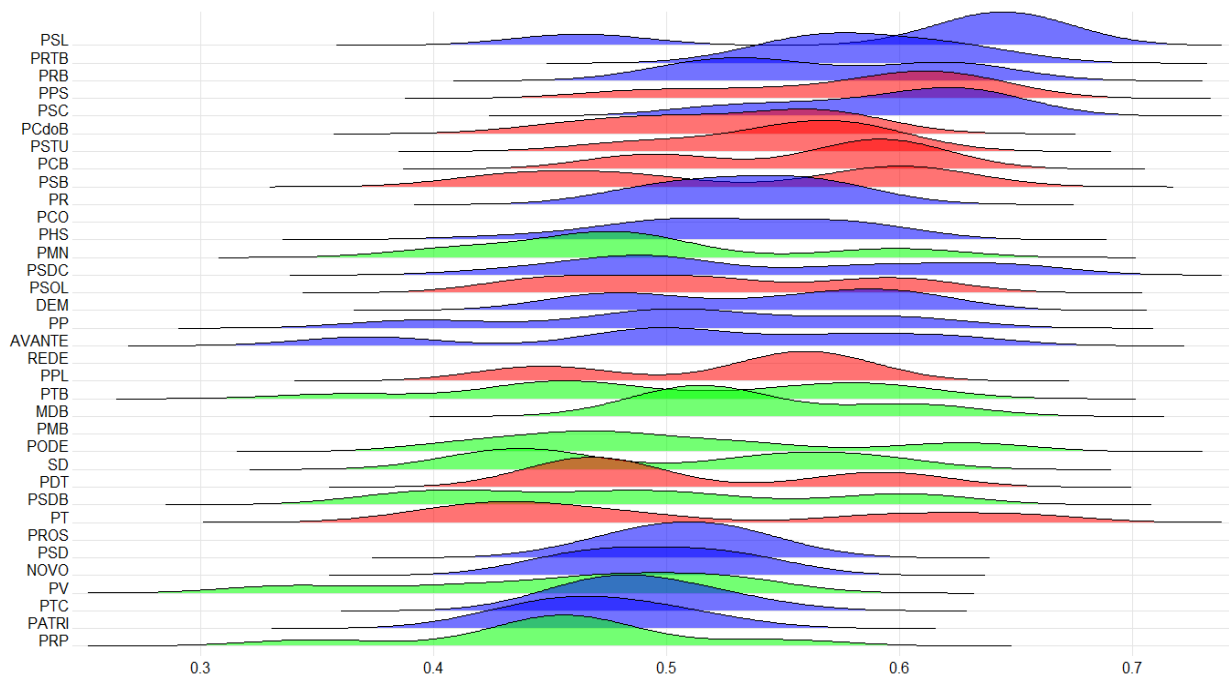


Figura 8: Densidade Kernel Índice de Jaccard N3

Operacionalizando o argumento de Ribeiro 2012, podemos afirmar que a taxa de similaridade seria uma função de ¹⁶:

- **Partido.** Variável categórica. O partido utilizado de base no intercepto é o AVANTE. Ribeiro 2012 afirma que o PT seria o único partido cuja estrutura organizativa seria diferente da LOPP. Uma interpretação mais ‘forte’ do excepcionalismo petista seria esperar que o PT seria o único partido que se mostrará estatisticamente significativo em comparação com o intercepto. Podemos relaxar tal hipótese, todavia, e simplesmente afirmar que estamos controlando nosso modelo pelos partidos.

- **Ano.** Variável contínua, sendo o ano do em que o estatuto foi redigido, estandardizado em dois desvios-padrões Gelman 2008. Como vimos, estatutos que já tinham sido registrados anteriormente à Lei 9096/95 tinham até 1996 para criar seus estatutos. Temos a expectativa de que estatutos sejam mais similares à LOPP quanto mais próximo dessa data limite tenha sido o ano de sua redação. Como não faz sentido pensar no impacto dessa legislação antes de sua criação, criamos a condicional de assumir seu valor como 0 para estatutos redigidos antes de sua criação.

A essas variáveis adicionamos um controle:

- **Razão do Número de Palavras.** Variável contínua, gerada a partir da razão do número de palavras em dado estatuto pelo número de palavras da LOPP ¹⁷. Esperamos que estatutos mais longos naturalmente terão uma taxa de similaridade menor do que documentos mais curtos. Operacionalizamos essa variável como a razão do número de palavras do LOPP para restringir a variância da mesma. A vantagem de se usar tal método e não, por exemplo, tirar o logaritmo natural seria que a razão é mais intuitiva de se compreender. Ademais, não alteramos, assim, a distribuição de nossa variável.

Neste primeiro momento, preferimos trabalhar com um modelo mais simples, adotando uma OLS. Em trabalhos posteriores poderemos sofisticá-lo, se assim se mostrar necessário.

¹⁶Não incluímos a ideologia partidária em nossos modelos dado que há uma autocorrelação serial perfeita entre a mesma e a sigla partidária. Como a sigla partidária é uma variável teoricamente interessante para o argumento de Ribeiro 2012, decidimos mantê-la ao invés da ideologia.

¹⁷Ambos os documentos já *stemmed* e sem *stop words*.

Vemos, então, que a taxa de similaridade de um estatuto com a LOPP (y_i) é dada por:

$$y_i = \beta_0 + \beta_1 \text{Partido}_i + \beta_2 \text{Ano}_i + \beta_3 \text{RazãoPalavras}_i + \epsilon_i \quad (3)$$

Na Figura 16 apresentamos os coeficientes dos quatro modelos da semelhança dos estatutos com a LOPP. Um problema geral que temos de recordar é como as distribuições de nossos índices de cosseno estão centrados muito próximo de seu valor máximo possível (1), e a maior variância em nossos índices de Jaccard.

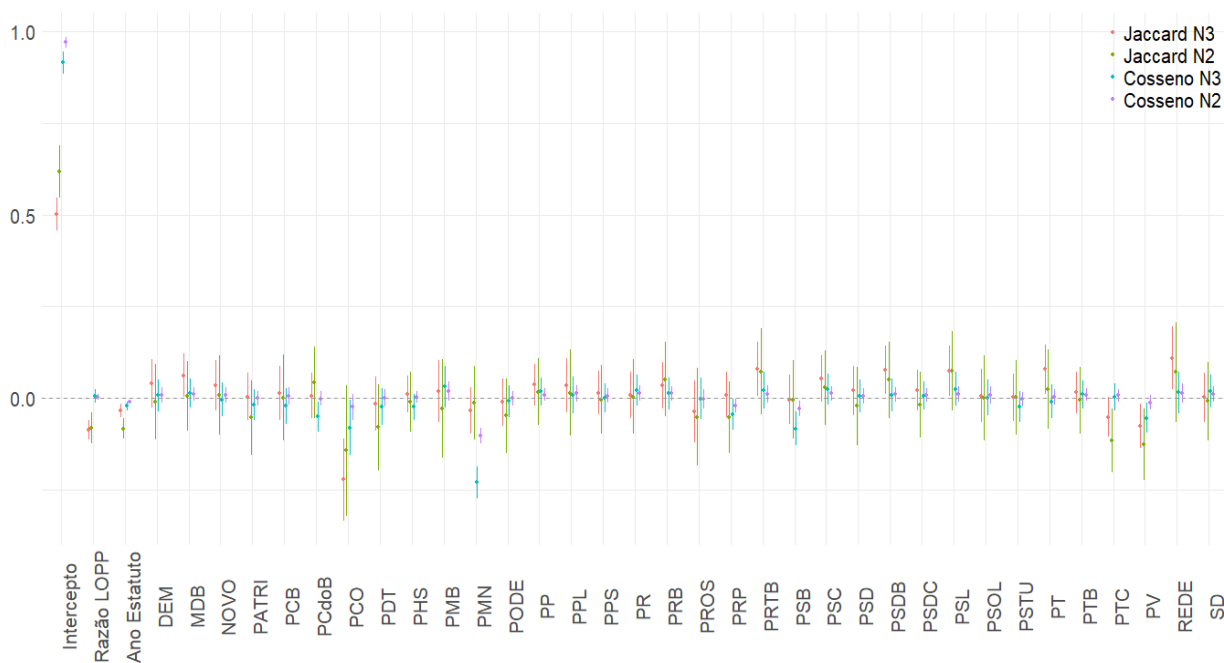


Figura 9: Modelo OLS Estatuto x LOPP

Nosso controle, a razão de palavras, mostra-se estatisticamente significativo para ambos os modelos Jaccard, mas não para os modelos de Cosseno. Assim como esperávamos, o Ano de criação do estatuto é estatisticamente significativo, contudo seu efeito marginal é relativamente baixo. Conforme observado na análise de Ribeiro 2012, não encontramos diferença estatisticamente significativa entre o MDB, PSDB e DEM em relação ao nosso intercepto (AVANTE). Não foi possível, contudo, observar uma diferença estatisticamente significativa dos estatutos do PT. Por um lado, isto se deve mais à forma que definimos

a taxa de similaridade textual, que não necessariamente é completamente relacionada ao conteúdo substantivo do estatuto. Por outro, vemos que outros partidos como o PCO e o PV são significativos em ao menos um de nossos modelos.

Para melhor verificar o efeito do Ano do estatuto, estimamos nas Figura 10 a Figura 13 o valor predito de nossas taxas de similaridade em relação à nossa variável explicativa de interesse. Como podemos ver, nos modelos de Jaccard (Figura 10 e Figura 11), o Ano possui um claro impacto na taxa de similaridade, com o intervalo de confiança dos valores superiores sendo claramente distinto do intervalo dos valores inferiores. Novamente, tal resultado é condicionado pela forma que criamos nossas taxas de similaridade e como geramos nossos modelos, e tal efeito é diminuído quando observamos os modelos de cosseno.

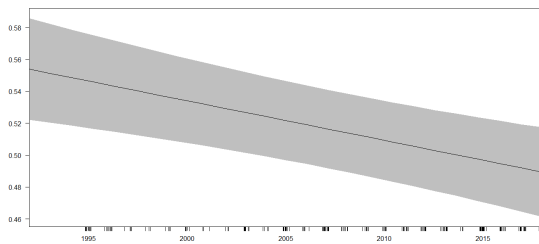


Figura 10: Diff 96 Jaccard 2-grama

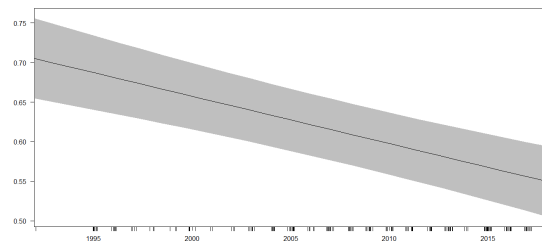


Figura 11: Diff 96 Jaccard 3-grama

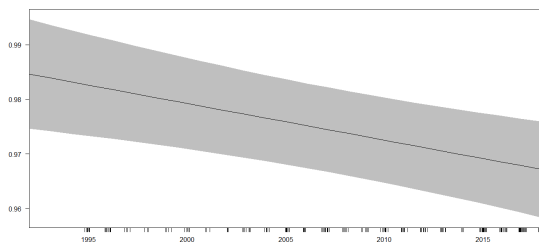


Figura 12: Diff 96 Cosseno 2-grama

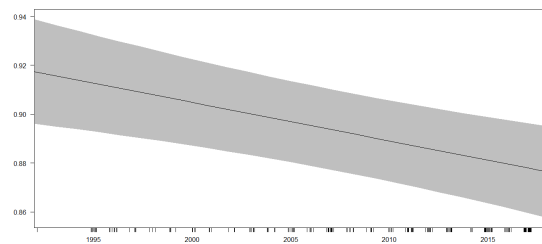


Figura 13: Diff 96 Cosseno 3-grama

5 Caso 2: Similaridade Textual entre Estatutos

Um segundo ponto do argumento de Ribeiro 2012 seria da taxa de similaridade dos estatutos brasileiros entre si. Especificamente, vimos que os estatutos partidários não pareciam mudar tanto ao longo do tempo. Dado que os estatutos são similares à LOPP, podemos esperar que seriam igualmente parecidos entre si. Novamente, analisamos 192 estatutos através de 4 taxas de similaridade textual. Como comparamos os estatutos entre si, e descartamos as comparações de estatutos idênticos, temos $(192) \times (192 - 1) = 36672$ casos.

Assim como fizemos na seção anterior, vejamos como nossas variáveis dependentes se distribuem. A Figura 14 mostra a densidade Kernel de nossas quatro taxas de similaridade, enquanto que a Tabela 3. 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 a quando analisamos a 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 de 100%. Ao verificar os dados, vemos que isso se deu para os estatutos do PRP de 2003 e 2005. Olhando para a história do partido, vemos que o mesmo foi fundado somente em 2005, contudo o TSE analisou e publicou ambos os estatutos apresentados pelo mesmo. Tais documentos são, como se pode esperar, exatamente idênticos. O lado positivo que podemos aproveitar disso é que podemos verificar que tanto a importação dos documentos quanto nossos índices de semelhança de fato funcionam como deveriam.

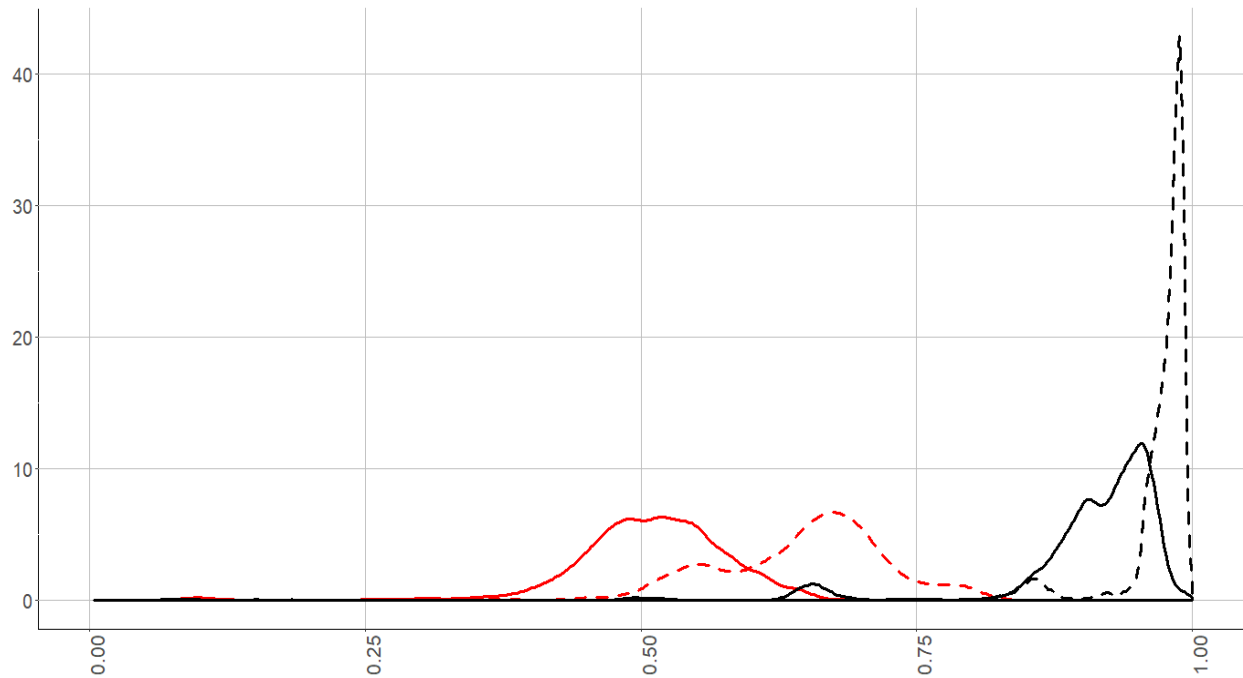


Figura 14: Densidade Kernel Estatuto x Estatuto

Tabela 3: Medidas Pontuais das Taxas de Similaridade entre Estatutos

Medida	Jaccard 2-grama	Jaccard 3-grama	Cosseno 2-grama	Cosseno 3-grama
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

Podemos comparar como a taxa de similaridade varia através dos partidos. Construímos um *heatmap* na Figura 15 mostrando a taxa de similaridade média entre os partidos, calculada pelo índice de Jaccard 3-grama. 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, pois como vimos 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 mais diferentes do que os demais. É o caso do PCO, do PV e, em menor medida, do PTC e SD.

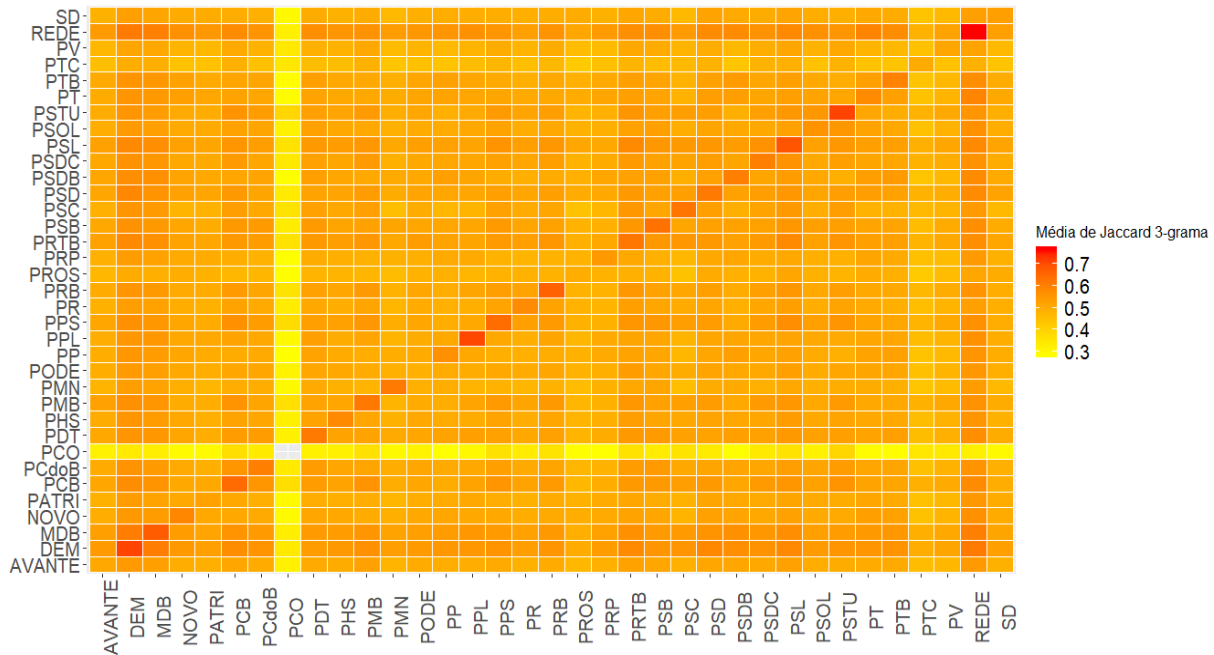


Figura 15: Modelo OLS Estatuto x Estatuto

Operacionalizando tal argumento, temos as seguintes variáveis:

- **Partido Plagiador**, ou Partido A. Variável categórica. Partido do estatuto que ‘plagiou’ o estatuto de outro.

- **Partido Plagiado**, ou Partido B. Variável categórica. Partido do estatuto que teve o estatuto ‘plagiado’.

- **Mesmo Partido**. Variável dicotômica. Indica se Partido A = Partido B. Podemos conjecturar que estatutos do mesmo partido são, na média, mais similares entre si do que entre estatutos de partidos diferentes. Partimos da intuição de que a construção de novos estatutos por um mesmo partido é incremental, logo maior sua similaridade. O argumento de Ribeiro 2012 de que partidos não alteram tanto seus estatutos ao longo do tempo também reforça a criação dessa variável.

- **Diferença Ano do Estatuto**. Variável discreta, definida como a diferença absoluta entre o ano de criação de dois estatutos. Novamente, esperamos que estatutos criados em anos mais próximos tenham maior probabilidade de serem mais semelhantes. Podemos esperar que tal semelhança será maior entre estatutos do mesmo partido do que estatutos de partidos distintos. Assim, moderamos o efeito dessa variável pela variável ‘Mesmo Partido’.

- **Razão Número de Palavras**. Variável contínua. A razão do número de palavras de ambos os estatutos. Assim como para nosso modelo da LOPP, assumindo que estatutos com páginas semelhantes serão mais similares entre si.

A taxa de similaridade entre estatutos (y_{2i}) é dada por:

$$y_{2i} = \beta_0 + \beta_1 \text{Partido1}_i + \beta_2 \text{Partido2}_i + \beta_3 \text{RazãoPalavras}_i + \beta_4 \text{DiffAno}_i + \beta_5 \text{MesmoPartido}_i + \beta_6 \text{DiffAno}_i * \text{MesmoPartido}_i + \epsilon_i \quad (4)$$

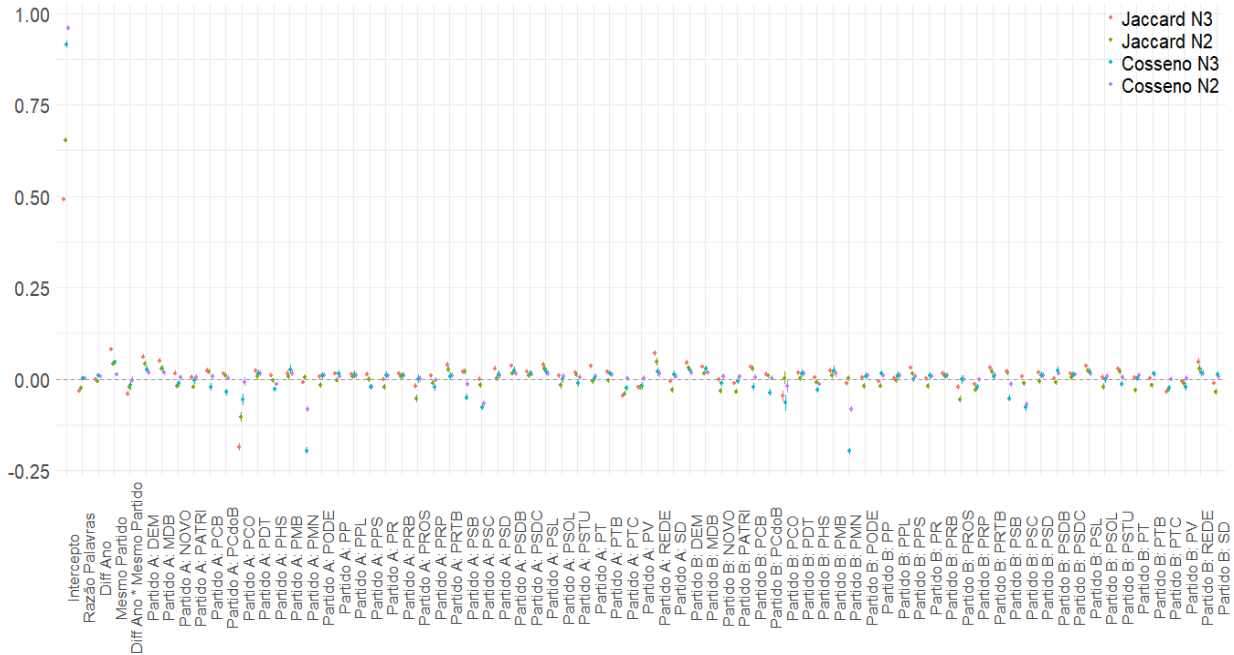


Figura 16: Modelo OLS Estatuto x Estatuto

Assim como vimos em nosso modelo para a similaridade com a LOPP, os coeficientes do modelo da similaridade entre estatutos variam sobretudo em relação ao tipo de índice utilizado, se de Jaccard ou cosseno. A Razão de Palavras, nosso controle, mostra-se significativa, contudo com efeitos moderados, para os modelos de Jaccard, e com efeitos marginalmente nulos para os modelos de cosseno. A Diferença do Ano, nossa variável explicativa de interesse, possui efeitos marginalmente nulos para os presentes modelos. Assim como esperado, o fato dos estatutos serem do Mesmo Partido aumenta significativamente a taxa de similaridade entre estatutos. Tal efeito é reduzido quando controlamos pela Diferença do Ano.

Olhando para os partidos plagiadores,

DEM, MDB, PDT, PMB, PP, PPL, PRTB, PSDB, PSDB, PSDC, PSL e REDE se mostraram consistentemente significativos como preditores de uma maior taxa de similaridade, enquanto que PCO, PMN, PSC (em menor medida) têm uma covariação negativa com a taxa de similaridade. Os coeficientes dos partidos plagiados possuem um padrão semelhante aos partidos plagiadores. Assim, foi possível detectar a similaridade do DEM, MDB e PSDB

já indicada por Ribeiro 2012, contudo não conseguimos verificar um padrão de distinção do PT sobre os demais estatutos e, de fato, para todos os modelos tal associação é levemente positiva.

6 Conclusão & Passos Futuros

No presente artigo 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 195 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 relativamente 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 entre 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 a data de criação dos estatutos, vimos que estes estão concentrados em Dezembro de anos não eleitorais. Não conseguimos detectar um padrão no número de palavras por estatuto, com cada partido apresentando uma realidade distinta. O volume de palavras por estatuto teve uma tendência crescente desde o começo da série temporal. 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 (Norris 2004).

Em nossos modelos, testamos dois argumentos principais, seguindo o argumento de Ribeiro 2012: o partido que criou o estatuto e o ano que o estatuto foi criado. Testamos tais parâmetros para dois tipos de modelo, a semelhança entre estatutos e a LOPP, por um lado, e os estatutos entre si, por outro. No geral, vimos que o ano que o estatuto foi criado, embora estatisticamente significativo, possui um efeito marginal baixo. Já quanto ao partido, embora Ribeiro afirme que o PT possui uma estrutura organizativa sistematicamente diferente dos outros principais partidos brasileiros, não observamos tal distinção quanto aos nossos índices de similaridade. Vimos que outros partidos como o PDT, o PV e o PATRI, todavia, sistematicamente se mostram distintos dos demais.

A presente pesquisa ainda está em estágios preliminares, e há múltiplas frentes em que ela deve ser melhorada. Primeiramente, quanto ao texto dos estatutos partidários, devemos sistematizar a revisão dos documentos, para que corrijamos eventuais distorções geradas quando extraímos informações dos pdfs originais. Em segundo lugar, temos de gerar de forma mais efetiva nossa variável dependente, i.e., a taxa de semelhança dos estatutos. Por ora, trabalhamos com múltiplos índices, contudo futuramente trabalharemos para integrá-los de maneira mais eficiente. Outro ponto a se investir seria na comparação dos estatutos dos partidos contemporâneos com partidos já extintos, tenham sido eles criados no atual período democrático ou em períodos anteriores. Acreditamos que com isso mapearemos melhor a variância da estrutura organizativa dos partidos do Brasil. Outro ponto que gostaríamos de avançar seria em uma análise mais profunda do conteúdo dos estatutos, ainda pouco explorados no geral. Enfim, é preciso avaliar o potencial da taxa de similaridade dos estatutos como variável explicativa para questões da estrutura organizativa dos partidos brasileiros.

Referências

- [Acr+16] Brice Acree et al. Comparing and Evaluating Cosine Similarity Scores, Weighted Cosine Similarity. Rel. téc. Working Paper, 2016.
- [Álv06] Maria Luzia Miranda Álvares. “Determinantes da seleção de candidaturas aos cargos parlamentares e regras estatutárias na formação do selecionador brasileiro, nas eleições de 2002”. Em: V Encontro da Associação Brasileira de Ciência Política (2006).
- [Ama10] Oswaldo Martins Estanislau do Amaral. “As transformações na organização interna do Partido dos Trabalhadores entre 1995 e 2009”. Tese de doutorado. [Universidade de Campinas], 2010.
- [BA13] Maria do Socorro Sousa Braga e Oswaldo Estanislau do Amaral. “Implicações do processo de seleção de candidatos na competição partidária: o caso brasileiro”. Em: Revista de Sociologia e Política 21.46 (2013).
- [BB09] Maria do Socorro Sousa Braga e Adla Bourdoukan. “Partidos políticos no Brasil: organização partidária, competição eleitoral e financiamento público”. Em: Perspectivas: revista de ciências sociais 35 (2009).
- [BC16] Bruno Bolognesi e Adriano Codato. “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.
- [Bol13] Bruno Bolognesi. “Caminhos para o poder: a seleção de candidatos a Deputado Federal nas eleições de 2010”. Tese de doutorado. Universidade de São Carlos, 2013.
- [BR09] Alberto Barrón-Cedeño e Paolo Rosso. “On automatic plagiarism detection based on n-grams comparison”. Em: European Conference on Information Retrieval. Springer. 2009, pp. 696–700.

- [Bra08] Maria do Socorro Sousa Braga. “Organizações partidárias e seleção de candidatos no estado de São Paulo”. Em: Opinião Pública 14.2 (2008), pp. 454–485.
- [Cop97] Michael Coppedge. “A classification of Latin American political parties”. Em: (1997).
- [DR15] Eric Gil Dantas e Karolina Mattos Roeder. “Partido, gênero e cor: a seleção de candidatos no PSTU nas eleições gerais de 2014”. Em: (2015).
- [Fer95] Luis Fernandes. “Muito barulho por nada? O realinhamento político-ideológico nas eleições de 1994”. Em: Dados 38.1 (1995), pp. 107–144.
- [FL07] Flavia Freidenberg e Steven Levitsky. “Organización informal de los partidos en América Latina”. Em: Desarrollo Económico 184 (2007), p. 539.
- [Gel08] Andrew Gelman. “Scaling regression inputs by dividing by two standard deviations”. Em: Statistics in medicine 27.15 (2008), pp. 2865–2873.
- [GJ15] Kristin N Garrett e Joshua M Jansa. “Interest group influence in policy diffusion networks”. Em: State Politics & Policy Quarterly 15.3 (2015), pp. 387–417.
- [GS13] Justin Grimmer e Brandon M Stewart. “Text as data: The promise and pitfalls of automatic content analysis methods for political texts”. Em: Political analysis 21.3 (2013), pp. 267–297.
- [Gua09] Fernando Henrique Eduardo Guarnieri. “A força dos partidos fracos-um estudo sobre a organização dos partidos brasileiros e seu impacto na coordenação eleitoral”. Tese de doutorado. Universidade de São Paulo, 2009.
- [HR10] Reuven Y Hazan e Gideon Rahat. Democracy within parties. Oxford University Press, 2010.
- [Kat02] Richard S Katz. “The internal life of parties”. Em: Political challenges in the new Europe (2002), pp. 87–118.

- [Kes14] Mike Kestemont. “Function words in authorship attribution. From black magic to theory?” Em: 3rd Workshop on Computational Linguistics for Literature (CLFL). 2014, pp. 59–66.
- [Key+55] Valdimer Orlando Key et al. “Politics, parties, and pressure groups”. Em: (1955).
- [MA08] Rachel Meneguello e Oswaldo Amaral. “Ainda novidade: uma revisão das transformações do Partido dos Trabalhadores no Brasil”. Em: BSP Occasional Papers 2 (2008), pp. 1–25.
- [Mai+00] Scott Mainwaring et al. Partidos conservadores no Brasil contemporâneo. Paz e Terra, 2000.
- [May11] R Mayer. “Os partidos como organizações: um estudo comparado do PSDB & PT. 2011. 86 f”. Tese de doutorado. Dissertação de mestrado. Programa de Pós-Graduação em Ciência Política, UFPR . . . , 2011.
- [MS13] Rafael Machado Madeira e Gabriela da Silva Tarouco. “Esquerda e direita no sistema partidário brasileiro: análise de conteúdo de documentos programáticos”. Em: Revista Debates (UFRGS) (2013).
- [Nat11] NDI National Democratic Institute. A compilation of political party statutes. 2011.
- [Nor04] Pippa Norris. “Building political parties”. Em: International IDEA (2004).
- [Oom17] Jeroen Ooms. “pdftools: Text extraction, rendering and converting of pdf documents”. Em: Computer software manual (2017).
- [Oom18a] Jeroen Ooms. hunspell: High-Performance Stemmer, Tokenizer, and Spell Checker. R package version 3.0. 2018. URL: <https://CRAN.R-project.org/package=hunspell>.
- [Oom18b] Jeroen Ooms. magick: Advanced Graphics and Image-Processing in R. R package version 2.0. 2018. URL: <https://CRAN.R-project.org/package=magick>.

- [Oom18c] Jeroen Ooms. tesseract: Open Source OCR Engine. R package version 4.0. 2018.
URL: <https://CRAN.R-project.org/package=tesseract>.
- [Pan05] Angelo Panebianco. Organização e poder nos partidos políticos. 2005.
- [Rib12] Pedro Floriano Ribeiro. “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 (2012), pp. 225–265.
- [Rib13] Pedro Floriano Ribeiro. “El modelo de partido cartel y el sistema de partidos de Brasil”. Em: Revista de ciencia politica (Santiago) 33.3 (2013), pp. 607–629.
- [Rib14] Pedro Floriano Ribeiro. “Em nome da coesão: parlamentares e comissionados nas executivas nacionais dos partidos brasileiros”. Em: Revista de Sociologia e Política 22.52 (2014), pp. 121–158.
- [Smi07] Ray Smith. “Tesseract ocr engine”. Em: Lecture. Google Code. Google Inc (2007).
- [Sta09] Efstathios Stamatatos. “Intrinsic plagiarism detection using character n-gram profiles”. Em: threshold 2.1,500 (2009).
- [WB07] Nina Wiesehomeier e Kenneth Benoit. “Presidents and Parties in Policy Space: Expert surveys of policy positions in 18 Latin American countries”. Em: 2007.
- [Zuc11] Cesar Zucco Jr. “Esquerda, direita e governo: a ideologia dos partidos políticos brasileiros”. Em: O Congresso por ele mesmo: autopercepções da classe política brasileira. UFMG, 2011, pp. 37–60.