

# Inovando a Cópia

Examinando estatutos partidários através da análise de  
similaridade textual

Lucas Mingardi

22/10/2020

## Por que Estatutos Partidários?

- ▶ Estatutos são “portas de entrada” para as instituições intrapartidárias.

# Por que Estatutos Partidários?

- ▶ Estatutos são “portas de entrada” para as instituições intrapartidárias.
  - ▶ Longa tradição de pesquisa, indo de Hartmann (1931) estudando os partidos da Tchecoslováquia, a Duverger (1954), Panebianco (1988), Norris (2004), Hazan (2010), van Biezen (2013), Ignazi (2014), etc.

## Por que Estatutos Partidários?

- ▶ Estatutos são “portas de entrada” para as instituições intrapartidárias.
  - ▶ Longa tradição de pesquisa, indo de Hartmann (1931) estudando os partidos da Tchecoslováquia, a Duverger (1954), Panebianco (1988), Norris (2004), Hazan (2010), van Biezen (2013), Ignazi (2014), etc.
- ▶ Contra-argumento: estatutos são meras formalidades, partidos não os levam a sério; estatutos não demonstram a “real” dinâmica partidária.

# Por que Estatutos Partidários?

- ▶ Estatutos são “portas de entrada” para as instituições intrapartidárias.
  - ▶ Longa tradição de pesquisa, indo de Hartmann (1931) estudando os partidos da Tchecoslováquia, a Duverger (1954), Panebianco (1988), Norris (2004), Hazan (2010), van Biezen (2013), Ignazi (2014), etc.
- ▶ Contra-argumento: estatutos são meras formalidades, partidos não os levam a sério; estatutos não demonstram a “real” dinâmica partidária.
  - ▶ Para se entender a diferença entre organizações formais e informais, é necessário se conhecer as organizações formais previamente. (Freidenberg & Levitsky, 2007)

## Por que Estatutos Partidários?

- ▶ Estatutos são “portas de entrada” para as instituições intrapartidárias.
  - ▶ Longa tradição de pesquisa, indo de Hartmann (1931) estudando os partidos da Tchecoslováquia, a Duverger (1954), Panebianco (1988), Norris (2004), Hazan (2010), van Biezen (2013), Ignazi (2014), etc.
- ▶ Contra-argumento: estatutos são meras formalidades, partidos não os levam a sério; estatutos não demonstram a “real” dinâmica partidária.
  - ▶ Para se entender a diferença entre organizações formais e informais, é necessário se conhecer as organizações formais previamente. (Freidenberg & Levitsky, 2007)
  - ▶ Brasil: caso partidos não se importassem com estatutos, não incorreriam no custo de constantemente reformá-los. Em menos de 30 anos, os atuais 35 partidos redigiram 195 estatutos.

# “Plágio” nos Estatutos Brasileiros

- ▶ **Legislação dos Estatutos Partidários no Brasil**

# “Plágio” nos Estatutos Brasileiros

- ▶ **Legislação dos Estatutos Partidários no Brasil**
  - ▶ Pré-88: estrutura delimitada pela 5681/71 (LOPP).



# “Plágio” nos Estatutos Brasileiros

- ▶ **Legislação dos Estatutos Partidários no Brasil**

- ▶ Pré-88: estrutura delimitada pela 5681/71 (LOPP).
- ▶ CF 88: autonomia partidária; previsão de regulamentação de estatutos em lei complementar.

# “Plágio” nos Estatutos Brasileiros

## ▶ **Legislação dos Estatutos Partidários no Brasil**

- ▶ Pré-88: estrutura delimitada pela 5681/71 (LOPP).
- ▶ CF 88: autonomia partidária; previsão de regulamentação de estatutos em lei complementar.
- ▶ Lei dos Partidos (9096/95): partidos já registrados no TSE tinham 6 meses para enviar seus estatutos.

# “Plágio” nos Estatutos Brasileiros

## ▶ **Legislação dos Estatutos Partidários no Brasil**

- ▶ Pré-88: estrutura delimitada pela 5681/71 (LOPP).
- ▶ CF 88: autonomia partidária; previsão de regulamentação de estatutos em lei complementar.
- ▶ Lei dos Partidos (9096/95): partidos já registrados no TSE tinham 6 meses para enviar seus estatutos.
- ▶ **Ribeiro (2012)**: Leitura manual dos estatutos do PT, PSDB, PMDB e PFL/DEM, de 1995 a 2011.

# “Plágio” nos Estatutos Brasileiros

## ▶ **Legislação dos Estatutos Partidários no Brasil**

- ▶ Pré-88: estrutura delimitada pela 5681/71 (LOPP).
- ▶ CF 88: autonomia partidária; previsão de regulamentação de estatutos em lei complementar.
- ▶ Lei dos Partidos (9096/95): partidos já registrados no TSE tinham 6 meses para enviar seus estatutos.
- ▶ **Ribeiro (2012):** Leitura manual dos estatutos do PT, PSDB, PMDB e PFL/DEM, de 1995 a 2011.
  - ▶ “À exceção do PT (...) 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”.

# Desenho de Pesquisa

- ▶ **Puzzle:** apesar de serem autônomos para redigir seus estatutos, e apesar de atualizarem frequentemente seus textos, partidos mantêm a estrutura da LOPP. O que está gerando esse resultado?
  - ▶ Viés de seleção de Ribeiro (2012)?
  - ▶ Path Dependence da imposição da Lei Eleitoral? (Estatutos/Partidos Estabelecidos vs Estreantes)
- ▶ **Plano de Trabalho:** generalizar a análise de Ribeiro (2012) para todos os estatutos de todos os partidos brasileiros.
- ▶ **Universo de Pesquisa:** 195 estatutos criados por 35 partidos entre 1990 (!) e 2019. Analisamos 192 desses estatutos.
- ▶ **Análise:** 1) comparação dos estatutos com a LOPP; 2) comparação dos estatutos entre si.

# Desenho de Pesquisa

- ▶ **Problema:** quantidade de textos dificulta a comparação manual.
- ▶ **Solução:** ler estatutos com optical character recognition (OCR); comparar textos com análise de similaridade textual ("plágio").
- ▶ **Limites:**
  - ▶ Text mining não deve ser pensado como substituto da leitura acurada dos textos. Tarefas complementares (Greener, 2003).
  - ▶ Carimbos, assinaturas, rubricas, etc, criando ruídos na análise.
  - ▶ Possível viés de seleção ao analisar apenas partidos contemporâneos.
  - ▶ Parte da similaridade é natural dado o tema comum dos textos. O que importa é a **variância** da distribuição, não o resultado pontual do texto A com o texto B. (Garett et al, 2015).

# Análise de similaridade de Texto - Questões a Considerar

- ▶ Dimensionalidade da Análise: **Análise de Corpora**. Textos vistos como “bag of words” e não segmentado por frases/páginas/etc (Greener & Stewart, 2003).
- ▶ Termo mínimo escolhido: **Análise de caracteres** (ao invés de palavras). Facilita análise para línguas aglutinativas e/ou com diacríticos, reduz necessidade pré-tratamento. Maior eficiência em comparação com similaridade de palavras (Solorio, 2015).
- ▶ Quantidade de termos analisados: **Bigramas e Trigramas** de termos. Cedeño & Rosso (2009): outros valores de Q não apresentam ganho significativo. 2-grama vs 3-grama: trade-off entre recall e precisão.
- ▶ Pré-tratamento: **Limpeza de texto**. Stemming, remoção de stopwords, remoção de pontuações, caixa baixa. Pró: diminui necessidade computacional. Contra: perdemos precisão. “Stopwords plagiarism” (Stamatos, 2009).

## Um exemplo prático - Índices de Similaridade

- ▶ Usamos dois índices de similaridade: Índice de Jaccard e Similaridade de Cosseno.
- ▶ Suponha que temos os seguintes documentos: D1: ('a\_banana') e D2: ('a\_batata').
- ▶ Criamos uma Matriz de Documentos a partir das regras que descrevemos anteriormente (bigrama de caracteres).

	a_	ba	na	ta	an	at	_b
D1	1	1	2	0	2	0	1
D2	1	1	0	2	0	2	1

Note que espaços ( ) são contados em nossos vetores como parte de nossos bi-gramas.



## Um exemplo prático - Jaccard

- ▶ **Jaccard:** “Intersecção pela união”.

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

- ▶ Para o Índice de Jaccard apenas contamos a ocorrência de cada termo, i.e. não nos importa a frequência dos termos.
- ▶ Comparando D1 e D2 temos:

$$J(D1, D2) = \frac{\text{Termos em Comum}}{\text{Total}} = \frac{3}{7} = 0.42$$

## Um exemplo prático - Cosseno

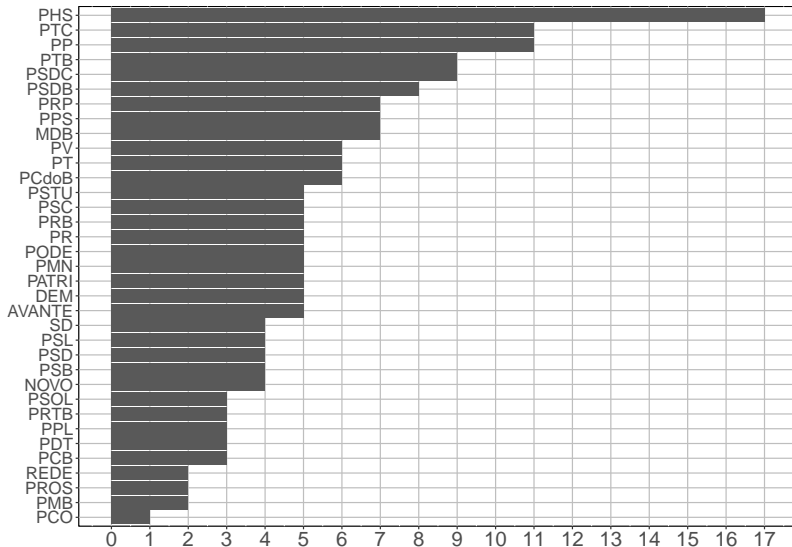
- ▶ **Cosseno:** “Cosseno da frequência dos vetores”.

$$\cos(A, B) = \frac{A \cdot B}{\|A\| \|B\|}$$

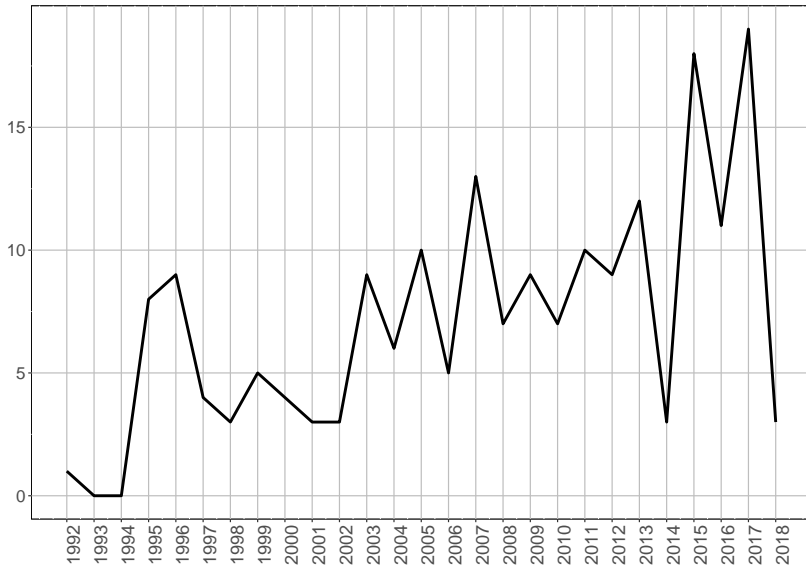
- ▶ Para a Similaridade de Cosseno a frequência de cada termo importa.
- ▶ Comparando D1 e D2 temos:

$$\cos(D1, D2) = \frac{\sum_{i=1}^n D1_i * D2_i}{(\sum_{i=1}^n D1_i^2)^{1/2} * (\sum_{j=1}^n D2_j^2)^{1/2}} = 0.27$$

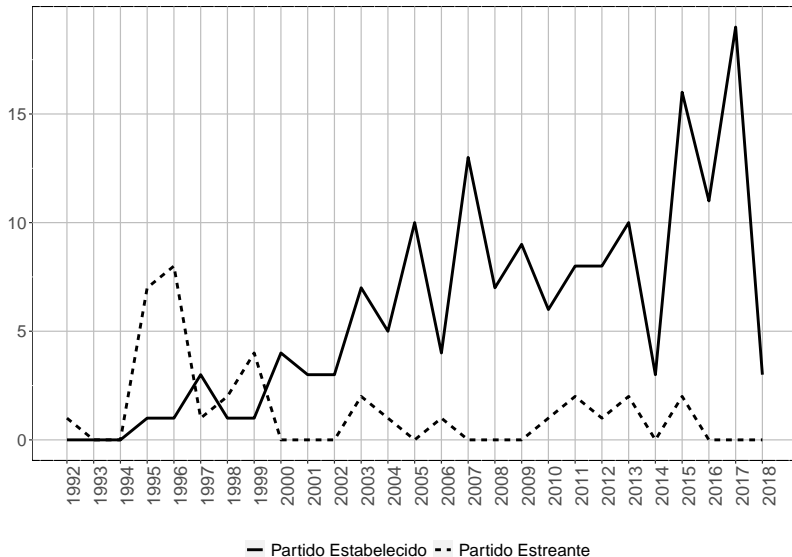
# Breve Análise Descritiva - Número de Estatutos por Partido



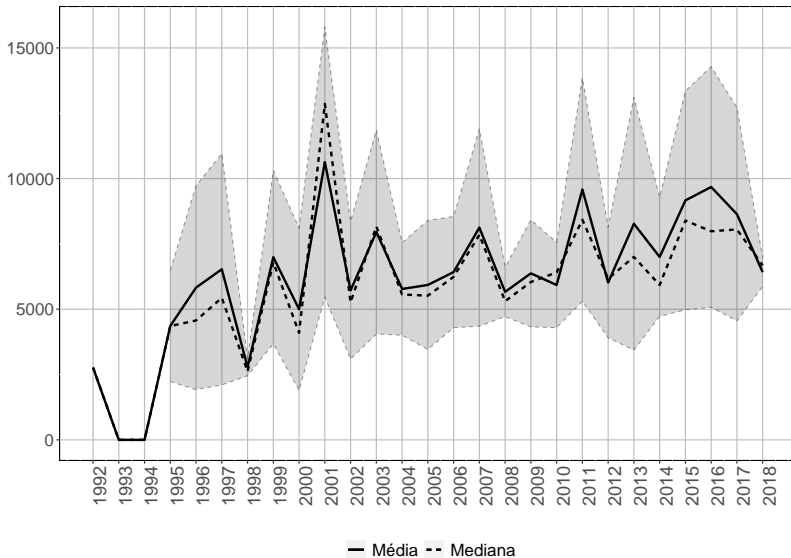
## Breve Análise Descritiva - Número de Estatutos por Ano



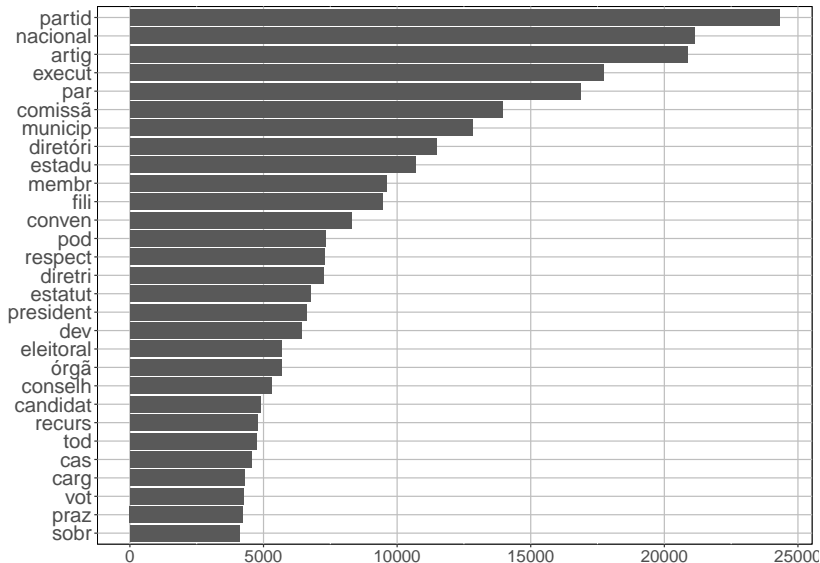
## Breve Análise Descritiva - Número de Estatutos por Ano: Partido Estreante vs Estabelecido



# Breve Análise Descritiva - Número Médio de Palavras nos Estatutos por Ano



# Breve Análise Descritiva - Palavras mais comuns nos Estatutos

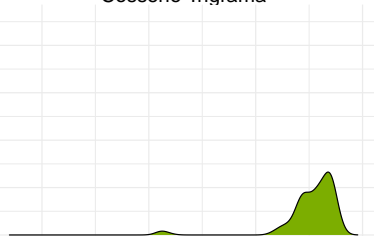


# Modelo 1 Estatuto x LOPP - Densidade Kernel Taxa Similaridade

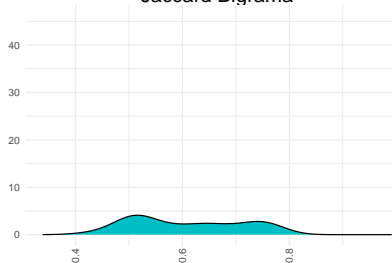
Cosseno Bigrama



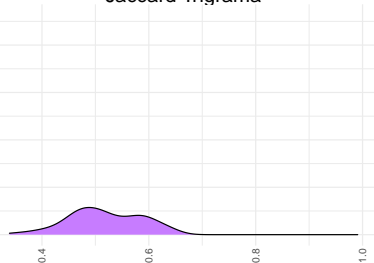
Cosseno Trigrama



Jaccard Bigrama

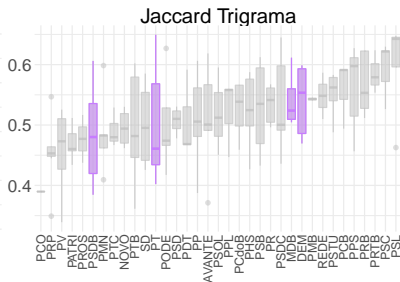
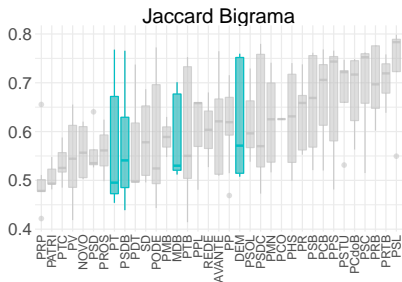
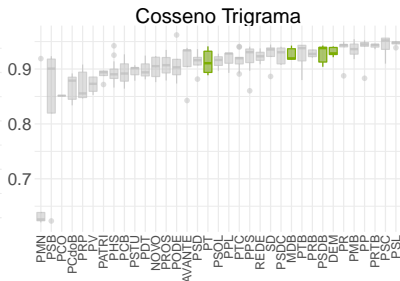
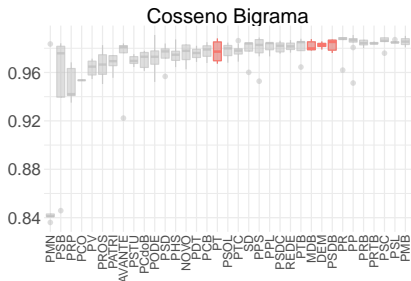


Jaccard Trigrama





## Modelo 1 Estatuto x LOPP - Boxplots Taxa Similaridade por Partido



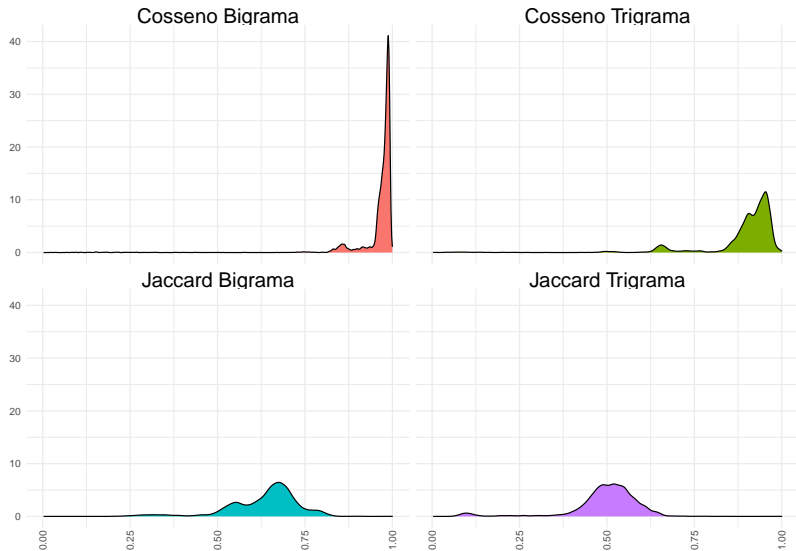
## Modelo 1 Estatuto x LOPP - Estimação OLS

- ▶ Modelo:  $Similaridade_{(a,lopp)} = \beta_0 + \beta_1 Lag(Similaridade_{(a,lopp)}) + \beta_2 Ano\ do\ Estatuto + \beta_3 Ano\ Fundação\ do\ Partido + \beta_4 Ideologia + \beta_5 PT + \epsilon$
- ▶ 4 VDs: 2-Jaccard, 3-Jaccard, 2-Cosseno, 3-Cosseno
- ▶ Modelo ainda básico, para ver projeções mais gerais. Testaremos futuramente com modelos hierárquicos. Dummy provisória para o PT.
- ▶  $Lag(Similaridade_{(a,lopp)})$ :
- ▶ Data Fundação do Partido: partidos mais antigos com maior probabilidade de ter path dependence da estrutura partidária anterior.
- ▶ Data criação do estatuto: estatutos criados mais próximos de 1995 tinham maior pressão de entregar um estatuto para competir eleitoralmente

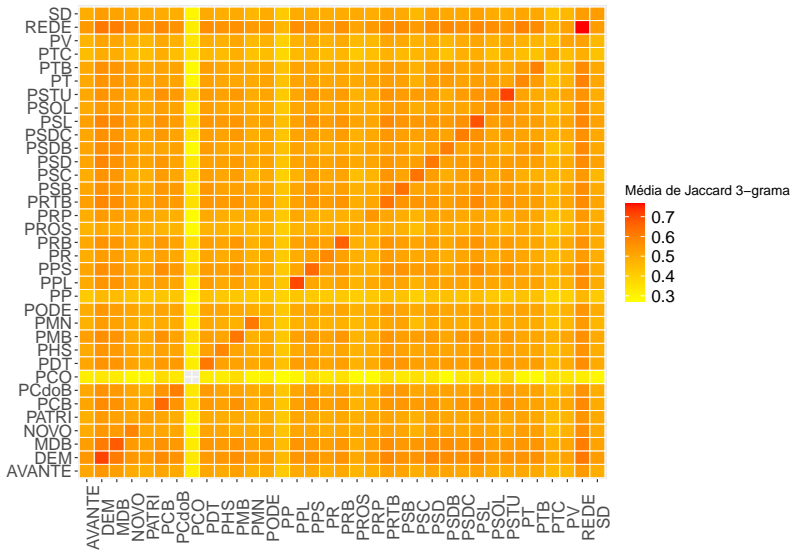
# Modelo 1 Estatuto x LOPP - Estimação OLS

	<i>Dependent variable:</i>			
	J2	J3	C2	C3
	(1)	(2)	(3)	(4)
Lag J2	0.146* (0.074)			
Lag J3		0.189** (0.077)		
Lag C2			0.613*** (0.073)	
Lag C3				0.643*** (0.069)
Ano de Edição do Estatuto	-0.043*** (0.010)	-0.018*** (0.006)	0.001 (0.002)	0.003 (0.005)
Ano de Fundação do Partido	0.008 (0.008)	0.005 (0.006)	-0.0002 (0.002)	-0.001 (0.004)
Partido de Esquerda	0.058*** (0.022)	0.034** (0.015)	0.005 (0.005)	0.007 (0.011)
Partido de Direita	0.026 (0.017)	0.023* (0.012)	0.008* (0.004)	0.017* (0.009)
PT	-0.075* (0.043)	-0.037 (0.030)	-0.00004 (0.011)	-0.0004 (0.023)
Constant	0.494*** (0.047)	0.400*** (0.041)	0.370*** (0.071)	0.309*** (0.063)
Observations	156	156	156	156
R <sup>2</sup>	0.278	0.203	0.373	0.412
Adjusted R <sup>2</sup>	0.249	0.171	0.347	0.389

# Modelo 2 Estatuto x Estatuto - Densidade Kernel



# Modelo 2 Estatuto x Estatuto - Heatmap Jaccard Trigramma



## Modelo 2 Estatuto x Estatuto - Estimação OLS

- ▶ Modelo:  $Similaridade_{(a,b)} = \beta_0 + \beta_1 Diff\ Ano\ Estatuto_{(a,b)} + \beta_2 Diff\ Fundação\ Partido_{(a,b)} + \beta_3 Ideologia_a + \beta_4 Ideologia_b + \beta_4 Ideologia_a * Ideologia_b + \epsilon$
- ▶ 4 VDs: 2-Jaccard, 3-Jaccard, 2-Cosseno, 3-Cosseno
- ▶  $\beta_2 Diff\ Fundação\ Partido_{(a,b)}$ : Partidos criados em épocas semelhantes podem ser organizacionalmente mais próximos entre si.
- ▶  $\beta_1 Diff\ Ano\ Estatuto_{(a,b)}$ : estatutos criados em épocas semelhantes podem sofrer influências semelhantes.

# Modelo 2 Estatuto x Estatuto - Estimação OLS

	<i>Dependent variable:</i>			
	J2	J3	C2	C3
	(1)	(2)	(3)	(4)
Lag J2	0.324*** (0.005)			
Lag J3		0.382*** (0.005)		
Lag C2			0.274*** (0.005)	
Lag C3				0.336*** (0.005)
Ideologia de Esquerda A	-0.001 (0.003)	-0.002 (0.003)	-0.0005 (0.002)	-0.004 (0.003)
Ideologia de Direita A	-0.010*** (0.002)	-0.011*** (0.002)	-0.001 (0.002)	0.0002 (0.003)
Ideologia de Esquerda B	-0.001 (0.003)	-0.002 (0.003)	0.004 (0.002)	-0.002 (0.003)
Ideologia de Direita B	-0.011*** (0.002)	-0.013*** (0.002)	-0.002 (0.002)	-0.0001 (0.003)
Mesmo Partido	0.021*** (0.003)	0.051*** (0.003)	0.013*** (0.003)	0.036*** (0.003)
Diff. Ano Estatuto	-0.001*** (0.0001)	-0.001*** (0.0001)	0.0004*** (0.0001)	0.0001 (0.0001)
Diff. Ano Fundação dos Partidos	0.00002 (0.0001)	0.001*** (0.0001)	0.001*** (0.0001)	0.001*** (0.0001)
(cont.)				

Note:

\*  $p < 0.1$ ; \*\*  $p < 0.05$ ; \*\*\*  $p < 0.01$

## Modelo 2 Estatuto x Estatuto - Estimação OLS

	<i>Dependent variable:</i>			
	J2	J3	C2	C3
	(1)	(2)	(3)	(4)
(cont.)				
Diff. Ano Fundação dos Partidos	0.00002 (0.0001)	0.001*** (0.0001)	0.001*** (0.0001)	0.001*** (0.0001)
Esquerda * Esquerda	0.015*** (0.004)	0.012*** (0.004)	0.002 (0.004)	0.011** (0.005)
Direita * Esquerda	0.007** (0.003)	0.004 (0.003)	0.001 (0.003)	−0.0003 (0.004)
Esquerda * Direita	0.007** (0.003)	0.005 (0.003)	0.001 (0.003)	−0.001 (0.004)
Direita * Direita	0.006** (0.003)	0.007** (0.003)	0.0001 (0.002)	0.001 (0.003)
Constant	0.444*** (0.004)	0.318*** (0.003)	0.688*** (0.005)	0.579*** (0.005)
Observations	36,121	36,121	36,121	36,121
R <sup>2</sup>	0.123	0.178	0.082	0.126
Adjusted R <sup>2</sup>	0.123	0.178	0.081	0.126

Note:

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



## Conclusões Provisórias

- ▶ Estatutos importam: quantidade de estatutos redigidos; regularidade da redação com o período eleitoral.
- ▶ Tendência geral é de estatutos ficarem cada vez mais diferentes e mais extensos com o passar do ano.
- ▶ O Ano em que o estatuto foi redigido importa para ambos os modelos. Seu poder preditivo, contudo, é bem baixo.
- ▶ Excepcionalismo petista? PT não é o único diferente da LOPP. Há partidos novos/velhos, de esquerda/direita, pequenos/grandes que possuem índices de similaridade menores que o PT.
- ▶ Lição mais geral para estudos de partidos no Brasil: evitar recortes a partir dos “partidos mais relevantes”.

## Próximos Passos

- ▶ Diminuir ruídos do OCR.
- ▶ Melhorar índices de plágio: unificar índices? mais índices?
- ▶ Modelo muito primário: reespecificar nosso modelo. Explorar modelos hierárquicos por partido.
- ▶ Retrabalhar variáveis derivadas de partido: até agora parecem pouco significativas teórica e empiricamente.
- ▶ Incluir partidos extintos na análise, testar recodificação da legenda para nomes antigos (eg: PFL e DEM).
- ▶ Investigar melhor relação temporal dos estatutos.