

## COESÃO PARTIDÁRIA NA CÂMARA DOS DEPUTADOS: UMA ABORDAGEM COM CIÊNCIA DE DADOS

*Ricardo Polato*<sup>1</sup>  
*Remis Balaniuk*<sup>2</sup>  
*Edilson Feredá*<sup>3</sup>

### Resumo

As representações partidárias têm grande relevância na política brasileira, representando as diferentes convicções políticas existentes na sociedade e lançando candidatos às eleições. Mas para que esses partidos exerçam efetiva representatividade, é necessário que os parlamentares por eles eleitos sigam seus valores e ideais. Um partido coeso consegue dar visibilidade aos interesses dos grupos sociais que representam, e não apenas aos interesses individuais do parlamentar. Neste contexto, este artigo propõe um modelo baseado em técnicas de mineração de dados, para mensurar a coesão dos partidos através da análise dos votos e pronunciamentos realizados pelos seus parlamentares da Câmara dos Deputados. O modelo proposto calcula dois indicadores de coerência da atuação dos parlamentares de cada partido, um baseado em estatísticas dos votos em proposições legislativas e outro nos principais temas abordados pelos parlamentares em seus pronunciamentos proferidos em plenário. Os resultados de ambos os indicadores mostram que os partidos têm muito a evoluir quanto a sua coesão e, mesmo que haja certa incoerência no resultado comparativo entre os dois indicadores, o modelo utilizado é promissor pois fornece *insights* e informações que servem de apoio à atuação legislativa.

**Palavras-chave:** Coesão partidária. Câmara dos Deputados. Ciência de dados. Mineração de dados.

### Abstract

Party representations have great relevance in Brazilian politics, representing the different political convictions existing in society and launching candidates for elections. But for these parties to exercise effective representation, it is necessary that the parliamentarians elected by them follow their values and ideals. A cohesive party manages to give visibility to the interests of the social groups they represent, and not just the individual interests of the parliamentarian. In this context, this article proposes a model based on data mining techniques, to measure the cohesion of parties through the analysis of votes and pronouncements made by their parliamentarians in the Chamber of Deputies. The proposed model calculates two indicators of coherence in the performance of parliamentarians from each party, one based on statistics of votes in legislative proposals and the other on the main topics addressed by parliamentarians in their pronouncements made in plenary. The results of both indicators show that parties have a lot to evolve in terms of their cohesion and, even if there is some inconsistency in the comparative result between the two indicators, the model used is promising as it provides insights and information that support legislative action.

**Keywords:** Party cohesion. Chamber of Deputies. Data Science. Data mining.

---

<sup>1</sup> Mestre em Governança, Tecnologia e Inovação pela Universidade Católica de Brasília (UCB).

<sup>2</sup> Pós-doutor em Computação pelas Universidades de Stanford e de Oxford, Doutor em Informática pelo Institut National Polytechnique de Grenoble, França. Docente do Programa de Pós-Graduação em Governança, Tecnologia e Inovação da Universidade Católica de Brasília (UCB).

<sup>3</sup> Doutor em Informática pelo Laboratoire d'Informatique, Robotique et Microélectronique de Montpellier (LIRMM/CNRS), França. Docente do Programa de Pós-Graduação em Governança, Tecnologia e Inovação da Universidade Católica de Brasília (UCB).

## Resumen

Las representaciones partidarias tienen gran relevancia en la política brasileña, representando las diferentes convicciones políticas existentes en la sociedad y lanzando candidatos para las elecciones. Pero para que estos partidos ejerzan una representación efectiva, es necesario que los parlamentarios elegidos por ellos sigan sus valores e ideales. Un partido cohesionado logra dar visibilidad a los intereses de los grupos sociales que representa, y no sólo a los intereses individuales del parlamentario. En este contexto, este artículo propone un modelo basado en técnicas de minería de datos, para medir la cohesión de los partidos a través del análisis de votos y pronunciamientos realizados por sus parlamentarios en la Cámara de Diputados. El modelo propuesto calcula dos indicadores de coherencia en el desempeño de los parlamentarios de cada partido, uno basado en estadísticas de votos en propuestas legislativas y otro en los principales temas abordados por los parlamentarios en sus pronunciamientos realizados en el pleno. Los resultados de ambos indicadores muestran que los partidos tienen mucho por evolucionar en términos de su cohesión y, si bien existe cierta inconsistencia en el resultado comparativo entre los dos indicadores, el modelo utilizado es prometedor ya que proporciona conocimientos e información que respaldan la acción legislativa.

**Palabras-clave:** Cohesión partidaria. Cámara de Diputados. Ciencia de datos. Minería de datos.

## 1. INTRODUÇÃO

No Brasil, o Poder Legislativo é exercido pelo Congresso Nacional, composto pela Câmara dos Deputados e pelo Senado e desempenha três funções essenciais para o fortalecimento da democracia: representar o povo brasileiro, legislar sobre os assuntos de interesse nacional e fiscalizar a aplicação dos recursos públicos (BRASIL, 1988, Art. 44).

A Câmara dos Deputados, detém o papel institucional de representante do povo brasileiro onde, dentre suas funções, articula e vota as proposições legislativas e também fiscaliza a utilização dos recursos arrecadados com o pagamento de tributos pela população (AMARAL, 2016). Nesta casa, é dado início à maioria das proposições legislativas e centraliza muitos dos maiores debates e decisões referentes ao âmbito nacional.

A votação das proposições é uma das funções mais importantes da atividade parlamentar. Mas, no cotidiano da Câmara, as atividades dos parlamentares vão além e estão intimamente ligadas à capacidade de diálogo de seus membros. O pronunciamento público do parlamentar na tribuna expressa sua ideologia partidária e seu posicionamento político. Para Schwartz (2018, p. 1) e Moreira (2019, p. 597), uma das principais ferramentas dos parlamentares para exercerem seu ofício é a capacidade de arranjar adequadamente as ideias, de detalhar adequadamente seus argumentos, de organizar e apresentar oralmente o pronunciamento em uma audiência. É por isso que os pronunciamentos parlamentares vão além de simples falas. Eles surgem de um contexto político, social e histórico específico e também devem ser compreendidos sob essa perspectiva.

Contudo, individualmente, parlamentares não detêm representatividade suficiente para defender e aprovar uma proposição. É neste ponto que os partidos políticos ganham notabilidade,

já que representam posições ideológicas e vontades coletivas. Cada bancada partidária, além da possibilidade de ter vários parlamentares como membros, pode realizar alianças com outros partidos a fim de formar blocos para uma determinada finalidade. Bancadas coesas, portanto, não se constroem da noite para o dia. São fruto do tempo e da oferta de incentivos, coletivos e seletivos, que possibilitam a construção de uma carreira política partidária. (MELO, 2004)

Para que os partidos políticos tenham efetiva representatividade, é necessário que os parlamentares eleitos sigam seus valores e ideais. Ou seja, é fundamental que cada um dos deputados membros, represente as correntes de pensamento e siga as orientações do seu partido.

Neste sentido e partindo do pressuposto que são escassas as maneiras de um efetivo acompanhamento da atividade legislativa, faz-se oportuno analisar o grau de convergência dos partidos políticos no Parlamento. Neste contexto, esse trabalho tem como problema de pesquisa a mensuração do grau de coesão dos partidos políticos brasileiros. Buscando contribuir com o tema, o trabalho explora o uso de modelos analíticos para esse fim, elaborados utilizando métodos de Inteligência Artificial, e focados na análise da atuação dos membros do parlamento por meio dos discursos proferidos em plenário e votações de proposições.

Assim, tendo como base que o posicionamento parlamentar é refletido diretamente nos votos e indiretamente nos pronunciamentos em plenário de cada parlamentar e que, em um partido coeso, seus membros votam e se posicionam de forma razoavelmente homogênea o presente artigo propõe um modelo analítico visando mensurar a coesão dos partidos através dos votos e discursos realizados pelos parlamentares da Câmara dos Deputados.

Desta forma, este artigo parte da hipótese central de que é possível o uso de modelos analíticos em conjunto com técnicas de mineração de dados e processamento de linguagem natural (PLN), para medir de forma eficiente a coesão interna partidária conforme os conjuntos de dados extraídos da atividade parlamentar. Já com intuito de validar o resultado da análise realizada nos dados dos pronunciamentos parlamentares foram propostas duas hipóteses: a de que é possível agrupar discursos parlamentares em temas de cunho político e a de que partidos tendem a concentrar seus discursos em poucos temas e que é possível agrupar discursos com temas semelhantes.

Na próxima seção apresentamos uma breve discussão sobre a coesão e sobre trabalhos relacionados ao tema. Na seção três, é abordada sobre a estrutura dos dados e metodologias aplicadas. Na próxima seção, é realizada a explanação dos resultados obtidos avaliando os métodos aplicados, realizando uma discussão sobre a coesão partidária pela perspectiva dos indicadores

propostos e ainda, realizando uma análise da temática partidária através dos discursos parlamentares. Por fim, apresentamos as conclusões e aplicações do presente artigo.

## 2. REVISÃO DA LITERATURA

A atuação parlamentar na temática das proposições legislativas é amplamente discutida na literatura. Amorim Neto e Santos (2003, p. 662) apresentam uma discussão com ênfase na atuação parlamentar que adota um modelo baseado em segredo ineficiente. Pressupõe-se que os partidos são dominados por líderes que demandam um sistema eleitoral que lhes preserve a autonomia local ao mesmo tempo que se preocupam com a capacidade do governo de elaborar políticas públicas de abrangência nacional, agindo como lobistas perante o Executivo Federal na busca por maximizar as transferências de recursos para suas clientelas.

Borges (2010, p. 121) busca fundamentar o entendimento da conexão entre políticas públicas e dinâmica eleitoral onde o poder de formular e implementar propostas legislativas de políticas públicas é relativamente descentralizado. Dessa forma, ao buscar o apoio do eleitorado por meio de políticas públicas, os governos locais ou estaduais competem com políticas similares ofertadas pelo Governo Federal.

A proposição legislativa em si é definida pela Câmara dos Deputados como toda matéria sujeita a deliberação da casa, na forma de Propostas de Emenda à Constituição, projetos de lei, emendas, indicações, requerimentos, recursos, pareceres e Propostas de Fiscalização e Controle. (BRASIL, 2020). É a partir da concepção de proposições que os agentes políticos estabelecem as leis, que vão gerir os mais diversos aspectos da vida em sociedade.

Já a coesão partidária mensura o quão os membros que ocupam cargos eletivos agem conforme o direcionamento do seu partido. Tsebelis (1997), afirma que a coesão partidária "refere-se à diferenciação de posições políticas dentro de um partido antes que se proceda a discussão e votação da posição em seu interior".

Melo (2004) salienta que para captar detalhes da experiência partidária recente no Brasil, a ideia de coesão deve referir-se à existência do partido como uma instituição. Assim como qualquer outra organização, um partido político requer um consenso mínimo sobre os limites institucionais do grupo e sobre os procedimentos para resolver disputas internas.

O aumento ou redução da coesão pode ser influenciado por diversos fatores. Freitas (2015, p. 17) levanta a suposição de quatro variáveis que explicam os fatores geradores de coesão programática: 1) subordinação do comportamento partidário ao seu projeto nacional - onde, um maior nível de coordenação de um partido em torno de um projeto político nacional é um fator a

contribuir para um maior grau de homogeneidade programática de suas elites subnacionais. 2) *office X policy-seeking* - é uma importante análise que permitir a distinção entre os partidos que não guiam seu comportamento a partir de diretrizes coerentes, daqueles que buscam o cargo eletivo com o objetivo maior de implementar as políticas públicas mais caras à organização partidária. 3) ideologia - construído com base no argumento de que quanto mais à esquerda um partido se posicionar na escala ideológica, maior será o seu nível de coesão programática. 4) tempo de filiação partidária - quanto maior o tempo médio de filiação dos parlamentares à sua agremiação, maior será o grau de coesão programática interna do partido.

Já Silva (2018, p. 35) elenca uma série de variáveis político-econômicas que podem influenciar a decisão dos partidos a integrar a coalizão ou oposição ao governo. Apesar do estudo ter enfoque apenas em nível de governo e oposição, as variáveis apresentadas são motivadoras para que um parlamentar permaneça seguindo, ou não, o direcionamento apresentado por seu partido político.

Nesta pesquisa, Silva (2018, p. 67-68) conclui que as seguintes variáveis afetam positiva ou negativamente a integração entre os partidos: 1) tempo do mandato - tendo impacto no final do ciclo eleitoral com a proximidade de novas eleições, este fator se torna determinante para um aumento da coesão partidária. 2) eventos críticos - que se referem aos eventos aleatórios como escândalos políticos, divisão de partidos, crises econômicas, guerras, crises internacionais, etc. 3) popularidade presidencial - dado que os partidos buscam maximizar sua capacidade de aprovação de leis e os resultados eleitorais podem ser determinados pela avaliação dos eleitores quanto à performance do governo, então, a taxa de aprovação presidencial pode afetar a decisão de um parlamentar em estar ligado ao partido que apoia ou se opõe ao governo federal. 4) condições econômicas - altas taxas de inflação e queda no PIB podem acarretar em mudanças nos ciclos de instabilidade. O desequilíbrio monetário afeta não só o parlamentar, mas também a opinião pública sobre a capacidade da cúpula política em lidar com esse desequilíbrio.

Neste contexto, são diversos os trabalhos realizados de forma empírica e também aplicando técnicas analíticas no domínio da atividade parlamentar, principalmente para mensurar a coesão, tanto entre governo e oposição quanto aplicadas aos partidos políticos individualmente. A seguir, são detalhadas uma série de pesquisas relacionadas ao problema em questão, fazendo o uso de métodos analíticos, a fim de avaliar a metodologia empregada e os resultados obtidos.

Grimmer e Stewart (2013, p. 267-269) descrevem métodos e forneceram uma reflexão sobre os modelos automatizados para análise de coleções de textos políticos. Os autores

apresentam de forma sistemática informações sobre a utilização de métodos supervisionados e não supervisionados aplicados à mineração de textos, em especial ao uso em dados da atividade parlamentar.

Schwartz (2018) buscou a partir da base textual dos discursos parlamentares definir um Índice de Identidade Ideológica, mensurando a similaridade entre os discursos realizados por parlamentares de coalizão e oposição. Nesta pesquisa, foram aplicadas técnicas de Processamento de Linguagem Natural e Aprendizado de Máquina nos discursos realizados no Pequeno e no Grande Expediente da Câmara dos Deputados, sendo possível observar a elevação ou redução do índice para coalizão e oposição conforme períodos de tempo históricos da política nacional.

Da mesma forma, utilizando dados não estruturados obtidos pelos pronunciamentos parlamentares, Moreira (2019, p. 597-598), fez uso de técnicas de aprendizagem computacional e modelos estatísticos com a aplicação de *expressed agenda model*, uma técnica que explora a estrutura textual para estimar os temas e a atenção atribuída por agentes políticos, analisando o conteúdo de mais de 127 mil pronunciamentos realizados através do Pequeno Expediente na Câmara dos Deputados Brasileira, buscando identificar a concentração temática dos Deputados Federais e a ênfase por eles atribuída às agendas econômica e social.

Já Fraussen, Graham e Halpin (2018, p. 467-468) buscaram aplicar uma abordagem supervisionada de aprendizado de máquina para identificar grupos parlamentares que se assemelham dentro do sistema político Australiano que, devido à sua alta coesão, tem maior êxito na defesa e aprovação de políticas públicas para as áreas que representam. Para isso foram utilizados métodos supervisionados de classificação atribuindo manualmente categorias predeterminadas a um grupo de textos, e após, aplicado o treinamento e teste com o uso do algoritmo *Support Vector Machine* (SVM).

Baptista (2015, p. 54-55) desenvolveu um modelo estatístico, com base em dados de votações da 50ª até a 54ª legislatura da Câmara dos Deputados, que determina a chance de um parlamentar mudar de posicionamento (governo ou oposição) com base em seu comportamento. Para isso, é avaliada a performance de classificação utilizando diversos modelos preditivos: Modelo Linear Generalizado (GLM), *Support Vector Machine* (SVM), Árvore de decisão, *Random Forest*, *Stochastic Gradient Boosting* (GBM), C5.0 e Rede Neural (NNET), comparando mudanças significativas de posicionamento de cada parlamentar ao longo do tempo.

Além da aplicação de tradicionais técnicas de aprendizado de máquina para avaliação de dados parlamentares, o Índice de Rice se mostrou um método eficiente e muito utilizado para

mensurar da coesão parlamentar baseada na votação de proposições. Nascimento (2012, p. 5), analisou a atuação parlamentar nas votações nominais ocorridas no Plenário da Câmara dos Deputados durante os Governos FHC e Lula aplicando o Índice de Rice, mensurando assim, o grau de compromisso do deputado perante seu partido. Da mesma forma, Neiva (2011, p. 298), fez uso deste índice para realizar a análise da coesão entre os partidos com base nos dados de votação, desta vez aplicado ao Senado Federal.

Por fim, Amelio e Pizzuti (2012, p. 146), exploraram os registros de votação do Parlamento Italiano utilizando a mineração de dados e métodos de análise em rede. Os autores fizeram o uso de uma técnica multidimensional para revelar coesão parlamentar. Desta forma, pode-se determinar como se estruturavam os grupos políticos durante evolução do período pesquisado (2008 a 2011). Esta técnica revelou-se uma ferramenta valiosa na detecção de tendências e desvios de posicionamento dos parlamentares.

É importante ressaltar que, apesar de um considerável número de trabalhos a respeito da coesão partidária tenha sido desenvolvido, este artigo busca como diferencial, a realização de uma análise exploratória, utilizando técnicas recentes de aprendizagem de máquina, tanto para dados obtidos através dos pronunciamentos parlamentares, quanto de votação de proposições para avaliar a coesão dos partidos políticos com representação na Câmara dos Deputados brasileira.

### 3. METODOLOGIA

A atividade parlamentar se refere a todo ato exercido pelos candidatos eleitos em relação aos interesses da sociedade. Contudo, dois atos da atividade parlamentar se destacam pela sua importância: a votação de proposições e os debates realizados em plenário.

A votação de proposições é uma das mais importantes incumbências delegadas aos parlamentares. Cada proposição define normas e estabelece a organização e a conduta necessária ao desenvolvimento coletivo. Os debates, expressos através dos pronunciamentos parlamentares, buscam defender os interesses singulares dos grupos sociais que supostamente o partido político busca representar. Diante disso, é proposto um método para mensurar a coesão partidária e expor seus resultados através de dois indicadores baseados nestes atos: (i) Indicador de Coesão Partidária sobre os Votos (ICPV), mensurado a partir dos dados de votação de proposições dos parlamentares para cada representação partidária, e (ii) Indicador de Coesão Partidária sobre os Discursos (ICPD), obtido através da similaridade dos discursos proferidos em plenário por representantes dos partidos políticos.

Para a elaboração desses indicadores foi utilizado o modelo de processos para mineração de dados CRISP-DM, seguindo as etapas detalhadas a seguir (PROVOST; FAWCETT, 2016, p. 26-33):

**Compreensão do problema** – nesta etapa, foi realizada uma revisão da literatura aprofundando em aspectos relativos à atividade parlamentar (processo legislativo; elaboração, análise e votação de proposições e a comunicação parlamentar) e à ciência de dados (aprendizado de máquina; similaridade por agrupamento e mineração de textos).

**Compreensão dos dados** – aqui, realizou-se inicialmente a extração da base de pronunciamentos parlamentares realizados em plenário e votação de proposições, para que então fosse possível realizar uma análise detalhada dos dados. Os dados de votação das proposições foram extraídos diretamente do portal de dados abertos da Câmara dos Deputados (<https://dadosabertos.camara.leg.br/>).

São diversos os tipos das proposições votadas pelos deputados: Propostas de Emenda à Constituição (PEC), Projetos de Lei Complementar (PLP), Projetos de Lei Ordinária (PL), Projetos de Decreto Legislativo (PDC), Projetos de Resolução (PRC) e Medidas Provisórias (MPV), entre outros. Neste artigo, foram utilizados os dados referentes às Propostas de Emenda à Constituição (PEC), Medidas Provisórias (MPV), Projetos de Lei Ordinária (PL) e Projetos de Lei Complementar (PLP) por se tratar das proposições com votação nominal. A votação nominal se diferencia da votação simbólica e da votação secreta por registrar em ata o posicionamento individual de cada parlamentar presente.

Os textos dos pronunciamentos realizados pelos deputados em plenário são publicados periodicamente em formato de notas taquigráficas no portal da Câmara. Esta análise foi realizada com base nos discursos relativos ao Grande Expediente para mensurar a evolução dos indicadores de coesão sobre os discursos e adicionado a esta base os discursos do Pequeno Expediente e Comunicações Parlamentares para análise temática destas falas. Foram escolhidas estas fases da comunicação na Câmara por se tratarem de momentos em que a fala do parlamentar não está exclusivamente focada na pauta e pode tratar de assuntos diversos, inclusive emitindo sua opinião pessoal sobre determinados assuntos importantes dentro do contexto nacional vivido no momento.

**Preparação dos dados** – esta fase compreendeu a seleção de variáveis consideradas relevantes à pesquisa, pré-processamento dos dados e aplicação de técnicas de PLN na base textual a qual fazem parte os pronunciamentos parlamentares.

Para os textos extraídos dos discursos, foi necessária aplicação de um pré-processamento específico, com objetivo de, a partir dos textos, inerentemente não estruturados, dar-lhes uma representação estruturada para que pudessem ser manipulados por algoritmos de agrupamento (SINOARA; ANTUNES; REZENDE, 2017, p. 7). Foram realizadas atividades de padronização do corpus, remoção de palavras desnecessárias, seleção dos termos mais significativos e, por fim, representação da coleção textual em um formato estruturado que preserve as principais características dos dados possibilitando a identificação de termos mais significativos.

A extração dos dados de votação proporcionou a obtenção de informações de sessão da votação, a saber: número da proposição, parlamentar votante, partido político e o voto em si.

**Modelagem** - nesta etapa, foram gerados os modelos a serem aplicados para atingir os objetivos propostos.

A partir da base de dados pré-processada, foi possível realizar o agrupamento dos discursos conforme sua similaridade. O modelo de agrupamento utilizado foi o *k-means*, algoritmo que particiona um conjunto de observações dentro de grupos mais próximos, definido por uma medida de distância entre observações e o centro (ou centroide) do grupo (também chamado de *cluster*).

Para este tipo de agrupamento, a definição da quantidade ideal de *K* (*clusters*) pode ser obtida através da utilização do método de cotovelo (*Elbow Method*). Conforme Grus (2016, p. 230), este método se baseia em marcar a soma dos erros ao quadrado (entre cada ponto e a média de seu agrupamento) como uma função de *k* e através de uma visualização gráfica verificar o ponto em que o gráfico "dobra".

Para mensurar a coesão partidária sobre os dados de votação de proposições, foi utilizado o Índice de Rice (RICE. 1925) um método amplamente utilizado, considerado intuitivo e facilmente interpretável onde seu valor é calculado com base na diferença entre o percentual de votos majoritários sobre o percentual de votos minoritários em cada votação (NEIVA, 2011, p. 298).

Diversas pesquisas, nacionais e internacionais fazem o uso deste índice para mensurar a coesão de grupos políticos. Neiva (2011) fez o uso do índice de Rice para avaliar o grau de coesão dos partidos com representantes no Senado brasileiro. Nascimento (2012) avalia a atuação

parlamentar nas votações nominais comparando as bases governista e de oposição com o índice de Rice, segundo o autor, devido ser uma métrica eficientemente capaz de avaliar a disciplina partidária em votações nominais. Já Dewan e Spirling (2011, p. 337) avaliam a coesão entre a base governista e oposição americana também com o uso do índice de Rice justificando que o índice é assertivo ao capturar a unicidade de um partido político em qualquer votação.

Os resultados prévios foram então analisados no que tange a utilização dos métodos propostos e, por vezes, estes métodos foram revisados a fim de gerar um resultado robusto e confiável. O resultado foi expresso através de dois indicadores que mensuram a coesão de cada partido: o ICPV e o ICPD.

O valor do ICPV foi obtido pela média aritmética do índice de Rice para cada votação nominal dos deputados federais agrupados por partido político, tendo como base cada ano entre o período de 2010 a 2019.

$$ICPV = \frac{R_i}{n}$$

sendo,  $R_i$  a soma dos índices obtidos de cada votação por partido e ano e  $n$  a quantidade de votações computadas para o partido e ano correspondente.

O cálculo do ICPD foi realizado após o agrupamento (clusterização) dos discursos por sua similaridade, mensurando o resultado para cada partido. A suposição aqui é a de que o algoritmo de agrupamento é capaz de separar discursos em temas diferentes e agrupar discursos em torno de um mesmo tema. Neste indicador, foram utilizados os dois *clusters* mais significativos, ou seja, os dois agrupamentos (ou temas) com maior quantidade de discursos associados a eles por partido. Nossa suposição aqui é a de partidos coesos e com identidade própria tendem a concentrar seus discursos num número pequeno de temas que consideram relevantes.

Desta forma, o indicador conta com a seguinte fórmula para seu cálculo:

$$ICPD = \frac{(n * (k_1 + k_2) - 2 * t)}{(n * t - 2 * t)}$$

sendo  $k_1$  e  $k_2$  as quantidades de discursos agrupadas no primeiro e segundo *cluster* mais significativos para o partido,  $n$  a quantidade de *clusters* utilizada para o agrupamento e  $t$  a quantidade total de discursos realizada pelo partido.

Para cada partido, foram selecionados os agrupamentos aos quais se obteve a maior quantidade de discursos e aplicados à fórmula a fim de extrair um valor entre 0 e 1, sendo 0 correspondente a uma distribuição completamente homogênea e aleatória dos discursos entre todos agrupamentos identificados e 1 a uma total concentração dos discursos nos dois *clusters* mais frequentes do partido. Isso significa que, quanto mais o indicador se aproximar de 1, maior tende a ser a concentração de temas presentes nos discursos dos parlamentares e, conseqüentemente maior a coesão no Partido como um todo.

**Avaliação** – Nesta fase, os resultados da mineração de dados são estimados de forma rigorosa e validados a fim de se obter um resultado válido e confiável. (PROVOST; FAWCETT, 2016, p. 31). Para isso, no caso dos dados obtidos através dos discursos, foi realizada inicialmente a clusterização de uma amostra aleatória de dados para observar o comportamento do algoritmo ao agrupar os discursos por temas distintos. Após isto, foi possível aplicar o teste de hipótese a fim de verificar a concentração de temas para cada partido.

Após a validação do modelo utilizado é proposto a discussão dos resultados pela perspectiva dos indicadores de coesão sobre os discursos e votação, bem como uma análise da temática partidária através dos discursos.

**Implantação** – Nesse caso, por se tratar de uma iniciativa de cunho acadêmico, sua “implantação” se materializa na sua publicação.

#### 4. RESULTADOS

O modelo proposto foi implementado por meio de algoritmos e ferramentas de análise de dados. Todo o código fonte e resultados estão disponíveis publicamente em <https://nbviewer.org/github/ricardopolato/coesaopartidaria/tree/main/>. Especificamente para a análise cujos os resultados são apresentados nessa seção foram utilizados dados do último ano da 53ª legislatura (2010) até o primeiro ano da 56ª legislatura (2019). Os indicadores foram processados e analisados buscando sua validação estatística e sua interpretabilidade.

Nesta sessão é apresentada uma avaliação e validação dos métodos propostos e também um detalhamento dos resultados em duas vertentes: (i) a discussão sobre a evolução dos indicadores de coesão interna dos principais partidos tendo como base os anos de 2010 a 2019; (ii) uma análise

temática com base em termos mais relevantes e recorrentes nos pronunciamentos proferidos em plenário no ano de 2019.

#### 4.1 AVALIAÇÃO E VALIDAÇÃO DOS MÉTODOS

O indicador ICPV, por ter sido obtido diretamente do índice de Rice, que é bastante intuitivo e largamente aceito na avaliação da coesão partidária, não requer uma validação específica. Já o ICPD foi construído a partir de duas hipóteses: a de que é possível agrupar discursos parlamentares em temas de cunho político e a de que partidos tendem a concentrar seus discursos em poucos temas. O objetivo dessa seção é testar essas hipóteses.

Para avaliar a capacidade do algoritmo *k-means* de agrupar os discursos por temas, uma amostra dos dados de discursos parlamentares do ano de 2019 foi submetida ao pré-processamento e ao algoritmo de clusterização. Para a definição da quantidade ideal de *clusters* foi aplicada o método de cotovelo mensurando entre 1 e 30 agrupamentos até que não fosse encontrada uma mudança significativa na variância calculada, que se deu na formação do cotovelo no décimo *cluster*.

Com o agrupamento dos discursos, buscou-se interpretar os temas subjacentes por meio de análise estatística dos termos mais frequentes (nuvem de palavras) tendo como base a hipótese que é possível agrupar discursos com temas semelhantes.

Na Figura 1 são mostradas as nuvens de palavras<sup>4</sup> criadas com base nos quatro maiores *clusters* obtidos. O primeiro *cluster* apresenta uma grande quantidade de termos. Contudo, algumas palavras se destacam devido a frequência elevada. Termos como previdência, reforma, aposentar, direito, sistema, salário, trabalhador, aposentadoria e trabalhadores são representativos de discursos correspondentes ao tema Previdência. Já o segundo *cluster* apresenta uma série de termos gerais entre os discursos. Contudo, se destacam pela frequência as palavras referentes aos temas educação e saúde. No terceiro *cluster* nota-se uma concentração de palavras altamente correlatas aos temas segurança e meio ambiente. E no quarto *cluster* se destacam termos referentes à economia e sistema fiscal: economia, dinheiro, orçamento, governo, renda, salário, dívida, crise e fiscal. Esta análise evidencia que o método *k-means* é capaz de agrupar os pronunciamentos em grandes temas.

---

<sup>4</sup> Essa ferramenta de visualização realça os termos mais frequentes por meio do tamanho da fonte. Termos mais frequentes aparecerão maiores na imagem.



A partir dos resultados, observa-se que em todos os partidos a hipótese nula de que os parlamentares pertencentes a este partido concentram suas falas em determinados temas foi validada pois todos os resultados de  $p$  apresentam valores superiores ao nível de significância 0,05.

#### 4.2 DISCUSSÃO SOBRE A COESÃO PARTIDÁRIA PELA PERSPECTIVA DOS INDICADORES PROPOSTOS

Nessa seção apresentamos os resultados obtidos a partir do processamento dos indicadores propostos para o período indo de 2010 a 2019. Foi utilizado o período entre 2010 a 2019 a fim de se obter uma série temporal significativa e ainda, para que se possa avaliar uma maior diversidade de legislaturas já que se inicia a partir do último ano da 53<sup>a</sup> até o primeiro ano da 56<sup>a</sup> legislatura.

Esta seção não visa a discussão dos fatores que influenciam a coesão, mas a apresentação dos resultados obtidos através da análise. Os fatores que influenciam a coesão foram levantados pela da bibliografia existente na seção 2, através dos trabalhos de Freitas (2015, p. 17)<sup>5</sup> e Silva (2018, p.35)<sup>6</sup>.

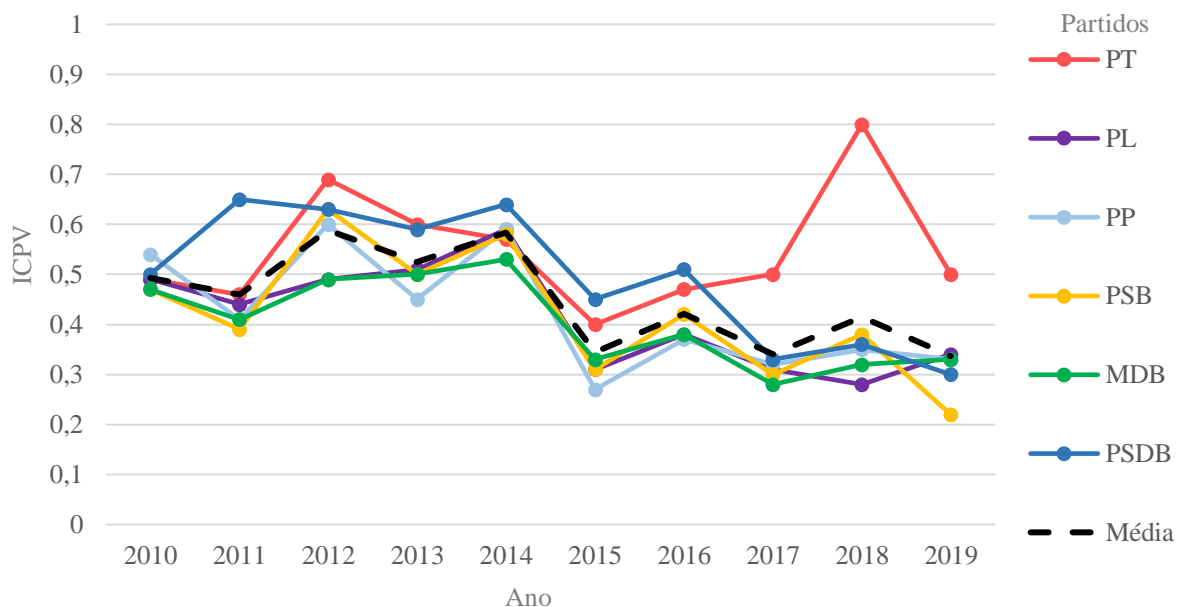
Os gráficos 1 e 2 representam respectivamente, os resultados dos Indicadores de Coesão Partidária Sobre os Votos (ICPV) e Discursos (ICPD). Apesar da apuração ocorrer para todos os partidos, foi representado o graficamente apenas o resultado dos seis partidos com maiores bancadas na última legislatura e que continham representação na Câmara dos Deputados em todos os anos apurados.

---

<sup>5</sup> Fatores que podem influenciar a coesão: subordinação do comportamento partidário ao seu projeto nacional, office x policy-seeking, ideologia e tempo de filiação partidária (FREITAS, 2015, p. 17).

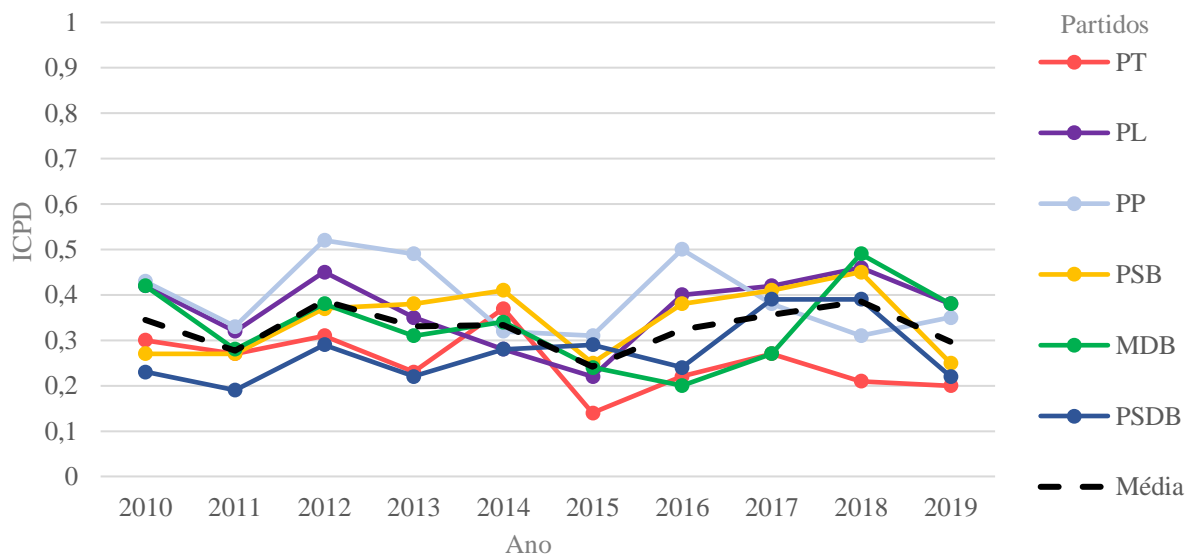
<sup>6</sup> Fatores que podem influenciar a coesão: tempo de mandato, eventos críticos, popularidade presidencial e condições econômicas (SILVA, 2018, p. 35).

**Gráfico 1** – Comparativo do Score do ICPV entre os anos de 2010 a 2019



Fonte: Elaborado pelos autores

**Gráfico 2** – Comparativo do Score do ICPD entre os anos de 2010 a 2019



Fonte: Elaborado pelos autores

Durante o período de 2010 a 2019, alguns partidos mudaram de nomenclatura. É o caso do PMDB que passou a ser representado como MDB. Para uma melhor avaliação da série histórica de resultados, cada partido será tratado pelo nome/sigla atual.

A avaliação iniciou-se pelo ano de 2010, o score obtido pelos partidos avaliados variou entre 0,47 e 0,54 para o ICPV e entre 0,23 e 0,43 para o ICPD. Em ambos os indicadores, o partido PP obteve o maior índice: 0,54 para o ICPV e 0,43 no ICPD e os partidos PSB e MDB obtiveram o menor resultado para o ICPV (0,47). Já no ICPD o partido com menor resultado foi o PSDB (0,23).

Contudo, esse alto índice de coesão não perdurou após 2010. Na comparação entre os anos de 2010 e 2011, é unânime a redução dos índices de coesão entre os partidos. A média dos valores obtidos para os partidos analisados reduziu de 0,49 para 0,46 em comparação ao ano anterior. Com exceção do PSDB, todos os demais partidos obtiveram uma redução na coesão, tanto no cálculo baseado nas votações, quanto nos discursos.

Em 2011 os valores mensurados para o ICPV ficaram entre 0,39 e 0,65, com o maior índice obtido pelo PSDB, seguido pelo PT com 0,46, PL com 0,44, PP e MDB com 0,41 e o menor resultado foi obtido pelo PSB, que ficou com 0,39. No ICPD, o resultado variou entre 0,19 e 0,33, sendo o PP com o índice de 0,33, PL com 0,32, MDB com 0,28, PT e PSB com 0,27 e o PSDB com 0,19.

No ano seguinte, 2012, nota-se que a grande maioria dos partidos conseguiu aumentar seu índice de coesão em pelo menos um dos indicadores. Apenas o PSDB, que obteve um acréscimo substancial no ICPV do ano anterior, não conseguiu elevar ainda mais seu score. A média dos valores para o ICPV aumentou de 0,46 para 0,59 e para o ICPD obteve uma média de 0,39, superior à média do ano anterior (0,28).

Os valores obtidos neste ano para o ICPV ficaram entre 0,49 e 0,69, com destaque para o partido PT que obteve o maior índice. Os partidos PL e MDB obtiveram os menores resultados para esse ano no ICPV. No Indicador de Coesão Partidária sobre os Discursos os valores variaram entre 0,29 e 0,52, sendo que, o partido PP obteve o valor de 0,52 e o PSDB obteve o menor valor calculado.

Em 2013 ocorreu uma nova queda dos índices de coesão para a maioria dos partidos. Com exceção de partidos como MDB e PL, que conseguiram aumentar ainda mais sua coesão sobre votos e do PSB que obteve um acréscimo no indicador de coesão sobre os discursos. Neste ano, para os partidos analisados, os valores do ICPV variaram de 0,45 até 0,60, com destaque para o PT que obteve o maior resultado, seguido pelo PSDB com 0,59, PL com 0,51, PSB e MDB com 0,50 e PP com 0,45. O resultado médio do ICPV reduziu de 0,59 para 0,53 em comparação ao período anterior.

No Indicador de Coesão sobre os Discursos, o score obtido variou entre 0,22 e 0,49, com o partido PP obtendo 0,49, seguido pelo PSB com 0,38, PL com 0,35, MDB com 0,31, PT com 0,23 e PSDB com 0,22. Assim como o ICPV, a média dos valores obtidos para o ano de 2013 também foi inferior em comparação ao ano anterior reduzindo de 0,39 para 0,33.

O ano de 2014 foi intenso para a política nacional, marcado por uma série de fatores políticos, econômicos e sociais que impactaram diretamente no cotidiano dos partidos. Estes fatos, afetaram a coesão em cada partido de forma distinta. Devido a isso, houve um acréscimo na coesão em pelo menos um dos indicadores para a maioria dos partidos. Apenas reduziram os valores para o partido PT na coesão sobre os votos e, PL e PP na coesão sobre os discursos. A média dos valores obtidos para os partidos analisados aumentaram de 0,53 para 0,58 no ICPV e obteve um valor semelhante ao ano anterior para o ICPD (0,33).

Dessa forma, os valores para o ICPV em 2014 variaram entre 0,53 e 0,64 como destaque para o PSDB que obteve o maior resultado, seguido pelos partidos PL e PP com 0,59, pelo PSB com 0,58 e PT com 0,57. O menor resultado ficou com o MDB que obteve 0,53. Já no ICPD o score obtido foi entre 0,28 e 0,41, com o partido PDT obtendo o melhor resultado. O partido PT obteve 0,37, MDB ficou com 0,34, PP com 0,32 e PL e PSDB com 0,28.

Em 2015, todos os partidos tiveram uma redução na coesão sobre as votações. Da mesma forma, houve uma redução generalizada na coesão sobre os discursos, com exceção do PP e PSDB que obtiveram resultado próximo ao ano anterior neste indicador.

O partido que obteve o melhor resultado no ICPV de 2015 foi o PSDB com score de 0,45, seguido pelo PT com 0,40, MDB com 0,33, PSB e PL com 0,31 e PP com 0,27. Neste ano, o ICPD obteve os seguintes resultados para cada partido analisado: PP com 0,31, PSDB com 0,29, PSB com 0,25, MDB com 0,24, PL com 0,22 e PT com 0,14.

Mesmo com as divergências políticas e a crise econômica, em 2016 houve um acréscimo na coesão para diversos partidos, com redução apenas no ICPD para os partidos PSDB e MDB. A média dos valores obtidos foi de 0,35 em 2015 para 0,42 em 2016 para o ICPV e 0,24 para 0,32 em 2016 no ICPD.

No ICPV, os valores ficaram entre 0,37 e 0,51, sendo o PSDB com o maior índice. O PT ficou com 0,47, PSB com 0,42, MDB e PL com 0,38 e PP com 0,37. Já no ICPD, os valores ficaram entre 0,20 e 0,50 com destaque para o PP com 0,50, seguido pelo PL com 0,40, PSB com 0,38, PSDB com 0,24, PT com 0,22 e MDB com 0,20.

O contexto político e econômico de 2017 afetou cada um dos partidos políticos de forma distinta. No ICPV, todos os partidos obtiveram uma grande redução, com exceção do PT que conseguiu um resultado um pouco maior neste ano em comparação ao ano anterior. O partido com melhor resultado foi o PT com 0,50, PSDB com 0,33, PP com 0,32, PL com 0,31, PSB com 0,30 e MDB com 0,28.

Já no ICPD a disparidade foi maior já que partidos como MDB, PSDB, PSB, PT e PL obtiveram um acréscimo e o partido PP obteve uma grande redução no seu score para este indicador. O partido com maior resultado foi o PL com 0,42, seguido pelo PSB com 0,41, PSDB com 0,39, PP com 0,38 e PT e MDB com 0,27.

No ano seguinte, 2018, em especial pelo período eleitoral, acarretou um crescimento nos indicadores de coesão partidária. A média dos valores calculados para os partidos subiram de 0,34 para 0,42 no ICPV e de 0,36 para 0,39 no ICPD de 2018.

Com exceção do PL, todos os demais partidos apresentaram um aumento significativo no ICPV. Os valores deste indicador variaram entre 0,28 e 0,80 sendo que o maior resultado ficou com o partido PL, seguido por PSB com 0,38, PSDB com 0,36, PP com 0,35, MDB com 0,32 e PL com 0,28. Para o ICPD, o aumento se deu apenas para os partidos PSB, PL e MDB. O partido PSDB obteve um resultado semelhante ao ano anterior. Os valores calculados ficaram entre 0,21 e 0,49, com o partido MDB obtendo o maior resultado e o partido PT com o menor.

Finalizado o período eleitoral, em 2019, semelhante aos demais anos pós-eleição, houve uma redução nos índices de coesão para a maioria dos partidos. Com exceção do PL, que aumentou seu ICPV, e do PP, com aumento no ICPD. O cálculo da média dos resultados obtidos pelos partidos analisados reduziu de 0,42 para 0,34 no ICPV e de 0,39 para 0,30 no ICPD de 2019.

Neste ano, o ICPV variou entre 0,22 e 0,50, com maior resultado para o partido PT que obteve um score de 0,50. O partido PL obteve um valor 0,34, PP e MDB ficaram com 0,33, PSDB com 0,30 e PSB com 0,28. Já para o ICPD, os resultados ficaram entre 0,20 e 0,38, com destaque para os partidos MDB e PL com o melhor resultado, seguido pelo PP com 0,35, PSB com 0,25, PSDB com 0,22 e PT com 0,20.

No contexto geral, nota-se que fatores como o período eleitoral, momento econômico e pressão popular afetam diretamente na atuação dos parlamentares da Câmara, aumentando ou diminuindo a coesão conforme o envolvimento de representantes do partido nestes fatos. Apesar dos índices aqui medidos terem relação com a similaridade das ações dos parlamentares, estes



Progressistas (PP)



Movimento Democrático Brasileiro (MDB)



Partido Liberal (PL)



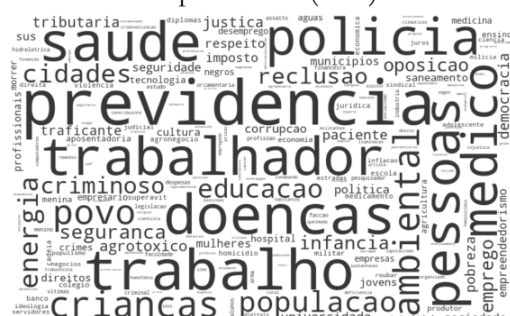
Partido Social Democrático (PSD)



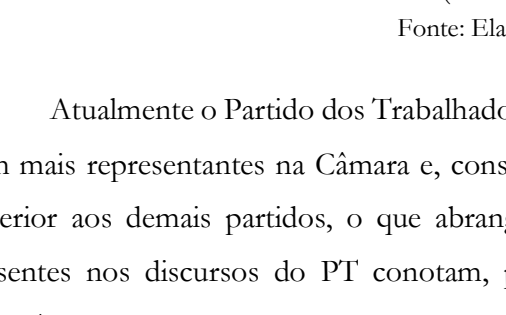
Partido Socialista Brasileiro (PSB)



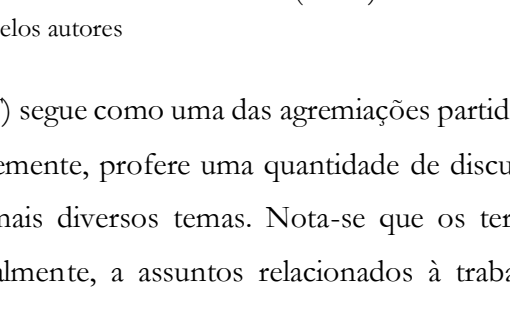
Republicanos (PRB)



Partido da Social Democracia Brasileira (PSDB)



Democratas (DEM)



Fonte: Elaborado pelos autores

Atualmente o Partido dos Trabalhadores (PT) segue como uma das agremiações partidárias com mais representantes na Câmara e, conseqüentemente, profere uma quantidade de discursos superior aos demais partidos, o que abrange os mais diversos temas. Nota-se que os termos presentes nos discursos do PT conotam, principalmente, a assuntos relacionados à trabalho, previdência, educação, pobreza e saúde, sem uma maior centralidade em um tema em específico.

O ano de 2019 foi marcado por grandes discussões a respeito da reforma da previdência. Diante disso, termos relativos a este assunto são comuns nos discursos para a maioria dos partidos.

Apesar dos discursos proferidos pelos deputados do PSL conterem uma grande diversidade de termos que expressam preocupações com os diversos problemas econômicos e sociais, nota-se uma ênfase em termos relacionados à segurança pública. Termos estes: polícia, segurança, militar,

crimes, violência, reclusão, entre outros, sendo as palavras polícia e segurança as com maior frequência dentre os termos analisados.

Os pronunciamentos do PP, como diferencial, enfatizam assuntos ligados ao agronegócio, contendo termos como: produtor, rural, agricultura, agronegócio, agropecuária, agricultor, palavras estas pouco utilizadas nos discursos dos demais partidos.

Os pronunciamentos dos parlamentares do MDB apresentam termos que refletem assuntos diversos, mas sem maior ênfase em um tema específico. Pela nuvem de palavras nota-se uma maior presença das palavras: municípios e cidades, povo e pessoas, trabalho e trabalhador, reclusão e segurança, saúde, educação, entre outros. Esta variação de palavras relacionadas a diferentes assuntos mostra que o partido busca a discussão de problemas diversos que afetam a população em geral.

Apesar dos discursos proferidos pelos parlamentares do PL apresentarem grande ocorrência em temas diversos, como educação e segurança, nota-se uma maior frequência de palavras relacionadas à saúde. Isso se dá devido à alta frequência do termo saúde e pela presença de outras palavras correlatas como: médico, hospital, medicamento, doenças, entre outras.

O partido PSD apresenta em seus discursos uma série de termos que tratam de assuntos em geral, sem grande frequência em algum tema específico. Os termos mais recorrentes são: municípios, energia, previdência, cidades, reclusão, pessoas e povo. A alta frequência destas palavras revela a intenção de defesa de interesses comuns, não direcionando seus discursos a demandas ou grupos sociais específicos.

De forma semelhante à maioria dos demais partidos, o PSB detém uma alta frequência de termos de refletem problemas sociais como trabalho, saúde, economia e educação. Contudo, é um dos poucos partidos com grande representação na Câmara cujos discursos contêm termos relacionados ao meio ambiente como: Amazônia e ambiental.

Os pronunciamentos do PRB trazem uma grande frequência de termos que denotam assuntos de interesse social, o que pode ser notado pelos termos: mulheres, povo, trabalho, previdência, segurança, saúde e educação.

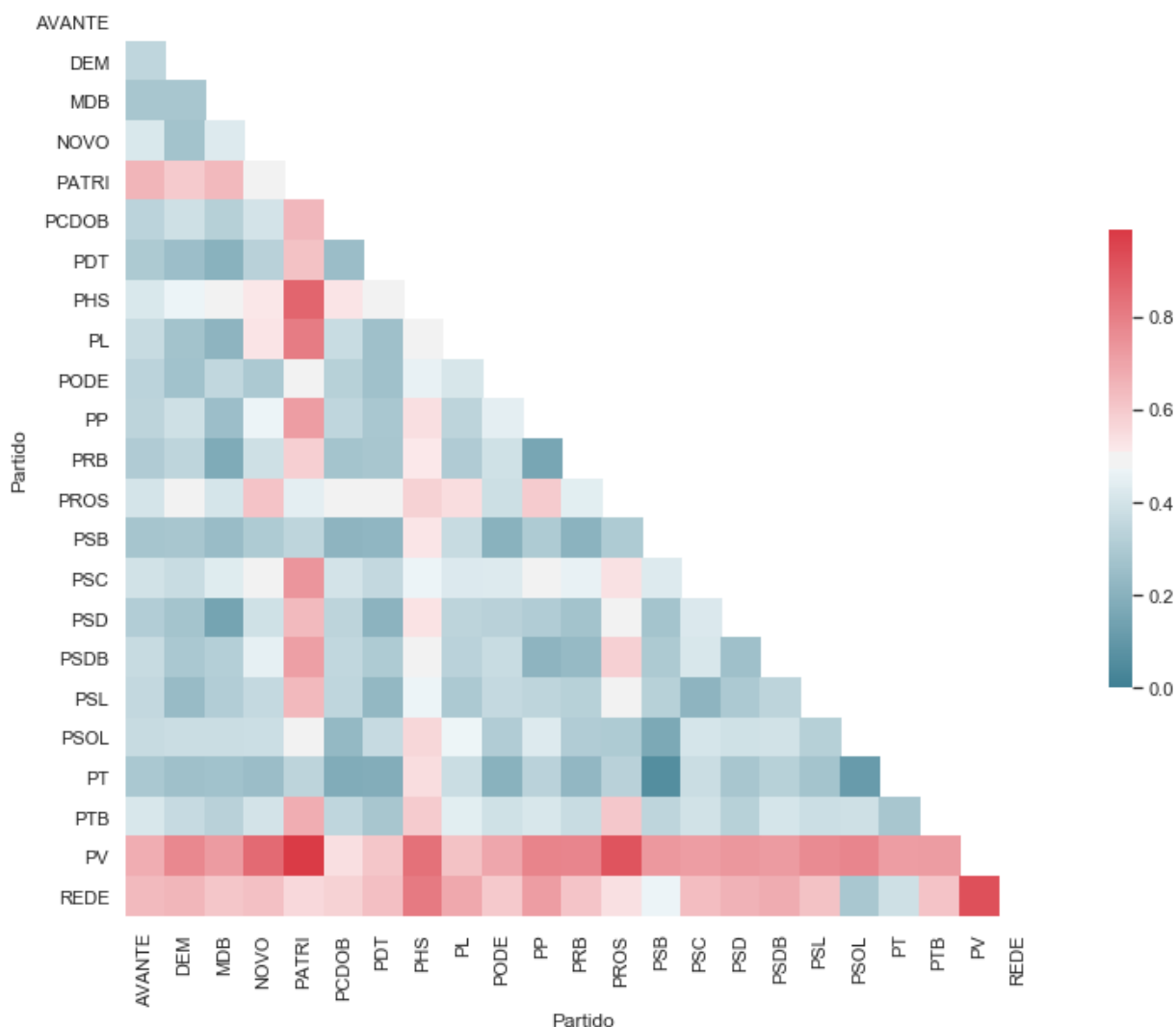
Os discursos do partido PSDB, apesar do termo energia ter grande ênfase, as demais palavras como mulheres, pessoas, população, produtor, saúde, educação, trabalho, trabalhador e previdência têm destaque. O conjunto destes e dos demais termos não apresenta uma proeminência de algum tema em especial, mas sim uma busca pela defesa de interesses gerais da população, sem priorizar alguma pauta social em específico.

Assim como muitos dos demais partidos, a reforma da previdência foi um assunto de grande discussão para o DEM, sendo que o termo previdência foi o mais frequente dentre as principais palavras abordadas nos discursos. Contudo, nota-se que demais termos que inferem pautas sociais estão da mesma forma presentes. Palavras como trabalho e trabalhador como referência ao tema trabalho; saúde, doenças e médico para o tema saúde pública; polícia, criminoso e segurança para segurança pública; além de termos gerais como pessoas, crianças, população e cidades que expressam a discussão e defesa dos direitos da sociedade em geral.

Diante deste conjunto de nuvens de palavras apresentadas, constata-se que a maioria dos partidos avaliados buscam uma maior amplitude na discussão de pautas econômicas e sociais, não focando em um tema único, mas sim nos mais diversos assuntos que impactam a sociedade em geral. Mesmo os partidos que apresentam destaque para um assunto em específico, também trazem em suas pautas os demais temas que referenciam os problemas sociais que afetam o País.

Contudo, cada partido detém termos e frequências distintas. Na Figura 3 é apresentado um gráfico obtido pelo cálculo da média entre o ângulo do Cosseno gerado para cada vetor em relação à origem. Ou seja, essa métrica é utilizada como uma medida de similaridade entre o corpus em que se mede a distância de todos os pontos presentes no espaço-vetorial podendo variar entre 0 e 1 onde, quanto menor for a distância entre estes vetores no espaço, maior a similaridade.

**Figura 3** – Similaridade de discursos entre os partidos



Com base nesta métrica, os partidos com discursos mais semelhantes são o PT com o PSB, que obtiveram uma média do cosseno de 0,06, seguido pelo PSOL também com o PT, com 0,11, e MDB com PSD e PP com PRD, que obtiveram cosseno de 0,15.

O partido com menor semelhança foi o PV, que obteve um valor calculado para o cosseno de 0,99 para o Patriota; 0,93 para REDE e 0,91 para o PROS. A comparação entre os partidos com maior e menor semelhança é detalhada na Tabela 1.

**Tabela 1** – Similaridade através da distância do Cosseno

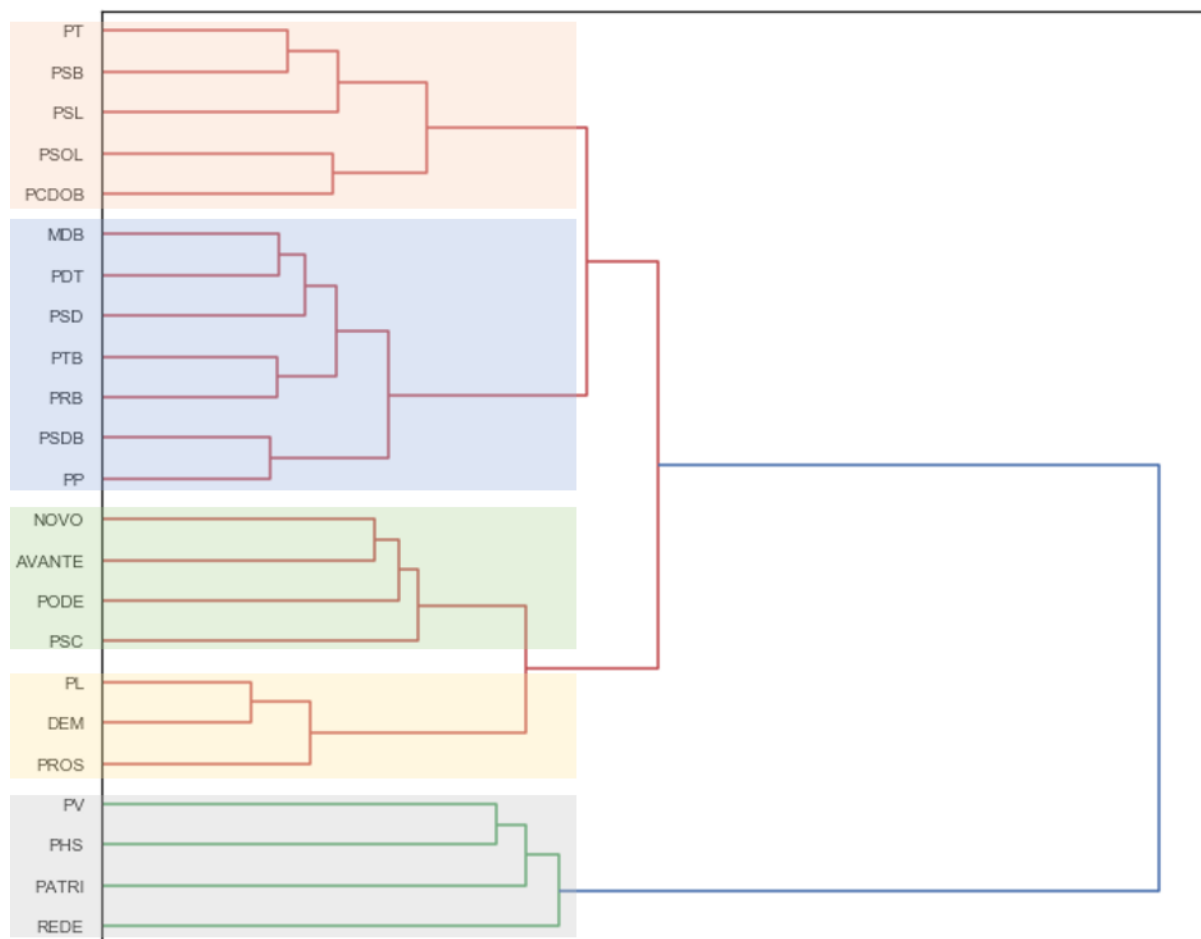
Partido	Maior Semelhança		Menor Semelhança	
	Partido	Cosseno	Partido	Cosseno
AVANTE	PSB	0,28	PV	0,68
DEM	PSL	0,24	PV	0,78
MDB	PSD	0,15	PV	0,73
NOVO	PT	0,25	PV	0,85
PATRI	PT	0,34	PV	0,99
PCDOB	PT	0,18	PATRI	0,65
PDT	PT	0,19	REDE	0,63
PHS	AVANTE	0,42	PATRI	0,87
PL	MDB	0,21	PATRI	0,80
PODE	PT	0,20	PV	0,69
PP	PRB	0,15	PV	0,79
PRB	PP	0,15	PV	0,78
PROS	PSB	0,30	PV	0,91
PSB	PT	0,06	PV	0,73
PSC	PSL	0,21	PATRI	0,74
PSD	MDB	0,15	PV	0,74
PSDB	PP	0,21	PV	0,72
PSL	PSC	0,21	PV	0,77
PSOL	PT	0,11	PV	0,79
PT	PSB	0,06	PV	0,72
PTB	PT	0,28	PV	0,72
PV	PCDOB	0,54	PATRI	0,99
REDE	PSOL	0,29	PV	0,93

Fonte: Elaborado pelos autores

No gráfico da Figura 3, é possível visualizar a similaridade dos discursos comparando os partidos políticos em pares. Todavia, para constatar os grupos de partidos que detêm uma maior similaridade nos seus discursos, utilizou-se o método de clusterização hierárquica. O objetivo desse método é agrupar cada variável em classes ou grupos sucessivamente maiores através da sua similaridade ou distância. O resultado é uma árvore hierárquica (dendograma) que represente os níveis de agrupamento para as variáveis, neste caso os partidos políticos. Dessa forma, foi realizado

o agrupamento da frequência dos principais termos presentes nos discursos dos partidos, obtendo como resultado cinco *clusters*, como apresentado na Figura 4.

**Figura 4** – Dendograma do agrupamento de termos presentes nos discursos por partido



Fonte: Elaborado pelos autores

O primeiro *cluster* é composto pelos partidos PT, PSB, PSL, PSOL e PCdoB, onde os termos presentes nos discursos mais recorrentes foram: povo, trabalhador, previdência, direitos, educação, pessoas, polícia, mulheres, trabalho e população.

O segundo *cluster* agrupou a maior quantidade de partidos, fazendo parte o MDB, PDT, PSD, PTB, PRB, PSDB e PP. Neste grupo, as palavras com maior frequência foram: cidades, mulheres, municípios, trabalho, pessoas, previdência, trabalhador, povo, educação e reclusão.

Os partidos NOVO, AVANTE, PODE e PSC constituíram o terceiro *cluster*. As palavras mais frequentes para esse grupo por ordem são: previdência, pessoas, trabalhador, trabalho, povo, população, direitos, educação, tributária e corrupção.

No quarto *cluster* estão os partidos PL, DEM e PROS, cujos termos mais recorrentes foram: povo, saúde, pessoas, trabalhador, cidades, trabalho, população, educação, polícia e reclusão.

O último *cluster* é composto pelos partidos PV, PHS, PATRI e REDE. Os termos mais utilizados por esse grupo de partidos foram: indígena, povo, educação, direitos, trabalhador, ensino, estudante, justiça, pessoas e previdência.

## 5. CONCLUSÃO

Este artigo contribui, principalmente, propondo um método para aferição da coesão partidária entre os parlamentares que ocupam as cadeiras na Câmara dos Deputados sobre duas vertentes: seus votos e pronunciamentos proferidos em plenário.

A importância da coesão está no fato que ela afeta diretamente na atuação de um partido político e em sua representatividade. Ou seja, a coordenação do comportamento de seus parlamentares é essencial na sua atuação e reflete diretamente na aprovação de proposições, repasses de recursos federais, distribuição de cargos entre as forças políticas e mantém a organização das relações entre os poderes Executivo e Legislativo. Assim, a mensuração do nível de coesão se torna relevante dentro do cenário político nacional.

Quanto aos resultados obtidos, para o ano de 2019, a análise dos dados de votação resultou no indicador ICPV que variou entre 0,22 e 0,56. Para os dados de discursos, o indicador ICPD resultou em valores entre 0,20 e 0,58. Como estes indicadores podem obter valores entre 0 e 1, os resultados mostram que os partidos têm muito a evoluir quanto a sua coesão, pois mesmo os partidos com os melhores resultados não se aproximam do valor máximo.

Já o comparativo entre os resultados dos indicadores no decorrer dos anos, nota-se que apesar de simbolizarem atos políticos diferentes – votações e pronunciamentos – e terem métodos distintos para seus cálculos, no período de 2010 a 2015, grande parte dos partidos obtiveram um comportamento correlato, ou seja, ambos indicadores aumentam ou diminuem de um ano para o outro de forma semelhante. Para os anos de 2016, 2017 e 2018, os resultados não mostraram uma clara correlação entre os indicadores devido, principalmente, a fatos como a crise econômica e política, forte pressão popular e sucessivos escândalos envolvendo parlamentares e importantes representantes partidários. Esses fatos influenciaram a coesão em cada indicador e partido de forma diferente. Diante dos acontecimentos, alguns partidos conquistaram um aumento e outros reduziram drasticamente a sua coesão no mesmo período.

Essa incoerência entre os resultados obtidos pelos dois indicadores nos últimos anos avaliados faz com que o modelo utilizado não seja conclusivo. Um modelo estatístico em sua essência, consiste na construção de hipóteses a partir da análise de dados para entender, prever ou explicar o comportamento de um fenômeno. Como tal, um modelo é uma simplificação da realidade e para seu uso é adicionada uma certa camada de abstração para facilitar seu processo de inferência. Ou seja, é muito difícil para um único modelo descrever perfeitamente um fenômeno do mundo real, dado que temos uma visão subjetiva do mundo, focando apenas em alguns aspectos mensuráveis do problema. Não obstante, a inferência estatística bem-sucedida acontece porque mesmo os problemas complexos têm um certo grau de consistência que exploramos. Portanto, como citado anteriormente, os modelos estatísticos podem nos ajudar a explicar e prever e são úteis para tirar conclusões sobre determinado fenômeno. Assim, os resultados, mesmo que inconclusivos, são promissores pois podem fornecer *insights*, informações e novos dados que servem de apoio à tomada de decisão para o meio político.

Os modelos propostos para cálculo dos indicadores podem ser utilizados por partidos políticos, bancadas ou mesmo pela população em geral como forma de acompanhamento da atuação dos parlamentares durante o mandato, avaliando a atuação individual ou mesmo a capacidade do partido em induzir as ações de seus parlamentares.

Ademais, os indicadores constituem um aparato objetivo de análise dos atos parlamentares, podendo ser utilizada também como fator de comparação entre estudos sobre a coesão de grupos políticos.

A heterogeneidade na quantidade de discursos proferidos pelos membros dos partidos é observada como uma limitação da pesquisa. Observa-se que partidos com poucos representantes via de regra obtêm maiores valores para o ICPD. Enquanto o algoritmo de mineração de dados compara algumas dezenas de palavras para partidos que detêm um grande número de cadeiras, para os pequenos partidos, os conjuntos de termos ficam limitados. Assim, neste caso, o indicador de coesão sobre os votos deve ser priorizado sobre o indicador sobre os discursos.

Ressalta-se que o método utilizado para análise dos discursos é uma proposta, onde alinhada com o método de Rice na votação de proposições, traz uma maior abrangência na análise da coesão partidária.

Espera-se, como fruto deste trabalho, ter-se colaborado para o melhor entendimento de como se desenvolver um processo de descoberta de conhecimento sobre a coesão partidária comparando a atuação do parlamentar frente a seu partido.



HEIMERL, et al. 2014. Word Cloud Explorer: Text Analytics Based on Word Clouds. *47th Hawaii International Conference on System Sciences*, Waikoloa, HI, p. 6-9.

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. s.d. *Série histórica do PIB*. Rio de Janeiro: IBGE. Disponível em: <[https://agenciadenoticias.ibge.gov.br/media/com\\_mediaibge/arquivos/7531a821326941965f1483c85caca11f.xls](https://agenciadenoticias.ibge.gov.br/media/com_mediaibge/arquivos/7531a821326941965f1483c85caca11f.xls)>. 18 dez 2020.

MELO, C. R. 2004. *Retirando as cadeiras do lugar: migração partidária na Câmara dos Deputados (1985-2002)*. Belo Horizonte : Editora UFMG.

MOREIRA, D. 2019. Karaokê da Câmara dos Deputados: o uso do microfone na atividade parlamentar. *Opinião Pública*, v. 25, n. 3, p. 597-629.

NASCIMENTO, E. B. 2012. Atuação Parlamentar Comparativa nas Votações Nominais da Câmara dos Deputados no Último ano dos Governos FHC e Lula. *Centro de Formação, Treinamento e Aperfeiçoamento da Câmara dos Deputados/Cefor*, Brasília.

NEIVA, R. R. P. 2011. Coesão e disciplina partidária no Senado Federal. *Dados*, v. 54, n. 2, p. 289-318.

NETO, A. ; SANTOS, O. F. 2003. O segredo ineficiente revisto: o que propõem e o que aprovam os deputados brasileiros. *Dados*. v. 46, n. 4, pp. 661-698.

O'HALLORAN, S; MASKEY, S.; MCALLISTER, G. PARK, D. K. 2016. Data Science and Political Economy: Application to Financial Regulatory Structure. SF: The Russell Sage Foundation. *Journal of the Social Sciences*, Vol. 2, No. 7, Big Data in Political Economy, pp. 87-109.

PROVOST, F.; FAWCETT, T. 2016. *Data Science para negócios: o que você precisa saber sobre mineração de dados e pensamento analítico de dados*. Rio de Janeiro: Alta Books.

RICE, S. A. 1925. The behaviour of legislative groups: a method of measurement. *Political Science Quarterly*, v. 40, n. 1, p. 60-72.

SCHWARTZ, F. P. 2018. *Análise do discurso parlamentar por meio da técnica de Processamento de Linguagem Natural: abordagem estatística e aprendizagem de máquina*. Brasília – DF. Relatório de Pesquisa (Estágio Pós-Doutoral). UnB.

SILVA, A. M. 2018. *Antes tarde do que nunca: uma análise dos fatores associados à saída dos partidos da coalizão*. Recife – PE. Dissertação (mestrado). Universidade Federal de Pernambuco.

SINOARA, R. A.; ANTUNES, J.; REZENDE, S. O. 2017. Text mining and semantics: a systematic mapping study. *Journal of the Brazilian Computer Society*, v. 23, n. 9.

TSEBELIS, G. 1995. Decision Making in Political Systems: Veto Players in Presidentialism, Parliamentarism, Multicameralism, and Multipartism. *British Journal of Political Science*, vol. 25, pp. 289-325.