

**DETECÇÃO DE VIÉS IDEOLÓGICO DE PORTAIS
DE NOTÍCIAS NA WEB**

VICTORIA AIRES

**DETECÇÃO DE VIÉS IDEOLÓGICO DE PORTAIS
DE NOTÍCIAS NA WEB**

Dissertação apresentada ao Programa de Pós-Graduação em Informática do Instituto de Computação da Universidade Federal do Amazonas como requisito parcial para a obtenção do grau de Mestre em Informática.

ORIENTADOR: FABÍOLA GUERRA NAKAMURA

Manaus

Agosto de 2020

Ficha Catalográfica

Ficha catalográfica elaborada automaticamente de acordo com os dados fornecidos pelo(a) autor(a).

A298d Aires, Victoria Patricia Silva
Detecção de Viés Ideológico em Portais de Notícias na Web /
Victoria Patricia Silva Aires . 2020
87 f.: il. color; 31 cm.

Orientadora: Fabiola Guerra Nakamura
Dissertação (Mestrado em Informática) - Universidade Federal do
Amazonas.

1. Viés na mídia. 2. Análise de notícias. 3. Classificação. 4. Web.
I. Nakamura, Fabiola Guerra. II. Universidade Federal do
Amazonas III. Título



PODER EXECUTIVO
MINISTÉRIO DA EDUCAÇÃO
INSTITUTO DE COMPUTAÇÃO

PROGRAMA DE PÓS-GRADUAÇÃO EM INFORMÁTICA



FOLHA DE APROVAÇÃO

"Detecção de viés ideológico de portais de notícias na Web"

VICTORIA PATRICIA SILVA AIRES

Dissertação de Mestrado defendida e aprovada pela banca examinadora constituída pelos Professores:

Fabiola Guerra Nakamura

Profa. Fabiola Guerra Nakamura - PRESIDENTE

Aluísio Soares da Silva

Prof. Aluísio Soares da Silva - MEMBRO INTERNO

Juliana Freire

Profa. Juliana Freire - MEMBRO EXTERNO

Manaus, 10 de Agosto de 2020

Agradecimentos

Agradecer a todos que contribuíram de alguma forma para o desenvolvimento deste trabalho não é uma tarefa fácil. Primeiro, à minha família, especialmente meus pais, por apoiarem minhas decisões em todos os momentos. Ao meu querido Juan Seabra pelo companheirismo, cuidado, preocupação e apoio que eu precisei em cada etapa e que foram muito importantes para eu seguir em busca dos meus sonhos.

A todos os meus amigos do PPGI, em especial do DNS Lab, pelos almoços, lanches, conversas, discussões científicas e todo o suporte que recebo desde a época da graduação. Todas as contribuições que recebi e a troca de conhecimentos foram muito importantes durante minha caminhada.

Ao segundo grupo de pesquisa que ganhei durante o caminho, VIDA Lab, da New York University, onde pude conhecer pessoas brilhantes que com certeza me inspiraram muito. Agradeço todos os eventos, almoços, reuniões, conversas, cafés na copa, e todo o companheirismo que recebi.

Aos meus demais amigos, incluindo Thais Gomes, Jhonathan Miranda, Ingrid Santos, Maria Clara Melo, Samantha Correa e Ana Vitória Vitoriano. Mesmo não convivendo comigo diariamente, essas pessoas constantemente me inspiram a seguir em frente e ser uma profissional melhor.

Por fim, gostaria de agradecer aos professores com quem trabalhei neste período. Ao professor Altigran da Silva pelo suporte e apoio que me disponibilizou, contribuindo para o desenvolvimento desta pesquisa. À professora Juliana Freire por me receber em seu grupo de pesquisa, pela disponibilidade e contribuições valiosas que direcionaram este trabalho. E aos meus orientadores Eduardo e Fabíola Nakamura pelos anos de orientação, dedicação, oportunidades e conselhos que pude receber neste período. Agradeço também à CAPES pelo suporte através de bolsa de mestrado, e à NYU pelo apoio durante meu mestrado sanduíche.

Esta pesquisa, conforme previsto no Art. 48 do decreto nº 6.008/2006, foi parcialmente financiada pela Samsung Eletrônica da Amazônia Ltda, nos termos da Lei Federal nº 8.387/1991, através de convênio nº 003/2019, firmado com o ICOMP/UFAM.

*“So comes snow after fire,
and even dragons have their endings.”*
(J. R. R. Tolkien)

Resumo

Nos dias atuais, websites ou portais de notícias são os principais meios pelos quais as pessoas consomem informação. Entretanto, assim como mídias tradicionais, esses veículos podem ter um viés na maneira como reportam notícias, favorecendo uma ideologia de interesse. Combinado às mídias sociais e à facilidade de compartilhamento e alcance desse tipo de conteúdo, esse fato contribui fortemente para a polarização, crimes de ódio e outras consequências na opinião pública. Para tornar as informações mais transparentes ao público, é necessário desenvolver métodos para caracterizar a orientação ideológica destes portais automaticamente. Abordagens propostas recentemente não são exatamente adequadas para este problema, pois dependem, em sua maioria, de fontes externas, gerando resultados imprecisos caso contrário. Diante disso, neste trabalho apresentamos métodos para detectar viés ideológico em portais de notícias baseado apenas nos artigos de notícias oriundos destes portais, sem nenhuma fonte externa. Exploramos duas abordagens: análise de hiperlinks e conteúdo textual. O objetivo é demonstrar a eficiência e eficácia dessa estratégia comparada à literatura atual. Como resultados, mostramos que uma abordagem baseada em hiperlinks é capaz de detectar viés ideológico em um cenário polarizado através de um método baseada em padrões de citações. Além disso, apresentamos uma abordagem baseada em conteúdo textual associada a conceitos de Teoria da Informação e mostramos que o método é capaz de superar um baseline mais tradicional, obtendo quase o dobro de acurácia/F1 em três bases de dados e três tarefas de classificação diferentes (bi-classe e multi-classe), enquanto emprega um conjunto de apenas quatro atributos (contra 282 do baseline) na detecção de diferentes níveis de viés ideológico em portais de notícias.

Palavras-chave: Viés na mídia, Análise de notícias, Classificação, Web.

Abstract

Nowadays, websites or news portals are the main sources of information to most people. However, like traditional media, these vehicles can have a bias in the way they report news, favoring an ideology of interest. Combined with social media and the ease of spreading this type of content, this fact strongly contributes to polarization, hate crimes and other consequences in public opinion. To make the information more transparent to the public, it is necessary to develop methods to characterize the ideological orientation/leaning of these portals automatically. Recent approaches are not exactly suitable for this problem, as they mostly depend on external sources, generating inaccurate results otherwise. Therefore, in this work, we present methods to detect ideological/political bias in news portals based only on news articles from these portals, without any external sources. We developed two approaches: exploring hyperlinks and textual content. The objective is to demonstrate the efficiency and effectiveness of this strategy compared to the current literature. As a result, we show that an approach based on hyperlinks is capable of detecting ideological biases in a polarized scenario through a method based on citation patterns. In addition, we present an approach based on textual content associated with Information Theory concepts and show that the method is able to overcome a more traditional baseline, obtaining almost twice the accuracy/F1 in three datasets and three distinct classification tasks (bi-class and multi-class), while employing a set of only four features (against 282 employed by the baseline) when detecting different levels of ideological bias in news portals.

Keywords: Media bias, News analysis, Classification, Web.

Lista de Figuras

2.1	Exemplos de distribuições de grau. À esquerda, em redes aleatórias e ao lado, em redes livres de escala. Na primeira, a distribuição dos graus segue uma distribuição de Poisson, enquanto na segunda, segue a lei de potência (Scholz, 2012).	10
2.2	Exemplos de redes aleatórias, <i>small-world</i> e livres de escala.	11
4.1	Visão geral da abordagem baseada em hiperlinks.	29
4.2	Um exemplo de um website classificado por Media Bias Fact Check (2019).	35
4.3	Distribuições de grau simples e ponderado.	37
4.4	Distribuições de centralidades de autovetor e intermediação das redes 1, 2 e 3.	39
4.5	Relações de citações inter e intra-classes de viés na rede 1.	43
4.6	Relações de citações inter e intra-classes de viés na rede 2.	45
4.7	Relações de citações inter e intra-classes de viés na rede 3.	46
4.8	Matrizes de confusão obtidas pelo método ao classificar as redes 1, 2 e 3. Os valores estão normalizados por linhas.	47
5.1	Visão geral da abordagem baseada em conteúdo.	51
5.2	Exemplo da citação de dois termos por fontes de quatro classes de viés na base de dados Notícias-Julho (descrita no Capítulo 5). Os valores estão normalizados.	53
5.3	Matrizes de confusão obtidas pelo baseline e a melhor estratégia de seleção de termos na base <i>Noticias-Julho</i> . Os valores estão normalizados por linhas.	61
5.4	Matrizes de confusão obtidas pelo baseline e a melhor estratégia de seleção de termos na base <i>Noticias-Fevereiro</i> . Os valores estão normalizados por linhas.	62
5.5	Matrizes de confusão obtidas pelo baseline e a melhor estratégia de seleção de termos na base <i>News-Aggregator</i> . Os valores estão normalizados por linhas.	63

5.6	Matrizes de confusão obtidas pelo baseline e a melhor estratégia de seleção de termos na base <i>Noticias-Julho</i> . Os valores estão normalizados por linhas.	64
5.7	Matrizes de confusão obtidas pelo baseline e a melhor estratégia de seleção de termos na base <i>Noticias-Fevereiro</i> . Os valores estão normalizados por linhas.	65
5.8	Matrizes de confusão obtidas pelo baseline e a melhor estratégia de seleção de termos na base <i>News-Aggregator</i> . Os valores estão normalizados por linhas.	66
5.9	Matrizes de confusão obtidas pelo baseline e a melhor estratégia de seleção de termos na base <i>Noticias-Julho</i> . Os valores estão normalizados por linhas.	67
5.10	Matrizes de confusão obtidas pelo baseline e a melhor estratégia de seleção de termos na base <i>Noticias-Fevereiro</i> . Os valores estão normalizados por linhas.	68
5.11	Matrizes de confusão obtidas pelo baseline e a melhor estratégia de seleção de termos na base <i>News-Aggregator</i> . Os valores estão normalizados por linhas.	69
A.1	Matrizes de confusão obtidas pelo método ao classificar as redes 1, 2 e 3, quando a dissimilaridade foi calculada através da distância de Jaccard. Os valores estão normalizados por linhas.	83

Lista de Tabelas

2.1	Matriz de confusão.	20
3.1	Sumarização dos trabalhos relacionados.	26
4.1	Distribuição de websites para cada classe de viés ideológico.	35
4.2	Redes construídas a partir dos dados coletados, representando diferentes classes de viés.	36
4.3	Nós com maiores centralidades de grau ponderado nas rede 1, 2 e 3.	38
4.4	Nós com maiores centralidades de autovetor e intermediação na rede 1, 2 e 3, respectivamente.	40
4.5	Outras propriedades das redes 1, 2 e 3.	40
4.6	Resultado da detecção de comunidades utilizando cinco algoritmos na rede 1.	42
4.7	Resultado da detecção de comunidades utilizando cinco algoritmos na rede 2.	44
4.8	Resultado da detecção de comunidades utilizando cinco algoritmos na rede 3.	46
4.9	Valores de acurácia e precisão, revocação e F1 por classe para cada rede ao classificar o viés ideológico usando padrões de citações.	48
5.1	Bases de dados utilizadas no método baseado em texto.	57
5.2	Número de bigramas selecionados para a construção dos histogramas de frequência em cada estratégia para cada base de dados.	60
5.3	Resultados do método classificando extremos (Ex) e moderados (M) na base de dados <i>Noticias-Julho</i> . Em negrito, o melhor resultado em cada métrica.	61
5.4	Resultados do método classificando extremos (Ex) e moderados (M) na base de dados <i>Noticias-Fevereiro</i> . Em negrito, o melhor resultado em cada métrica.	61
5.5	Resultados do método classificando extremos (Ex) e moderados (M) na base de dados <i>News-Aggregator</i> . Em negrito, o melhor resultado em cada métrica.	62
5.6	Resultados do método classificando websites de esquerda (E) e direita (D) na base de dados <i>Noticias-Julho</i> . Em negrito, o melhor resultado em cada métrica.	64

5.7	Resultados do método classificando websites de esquerda (E) e direita (D) na base de dados <i>Noticias-Fevereiro</i> . Em negrito, o melhor resultado em cada métrica.	65
5.8	Resultados do método classificando websites de esquerda (E) e direita (D) na base de dados <i>News-Aggregator</i> . Em negrito, o melhor resultado em cada métrica.	65
5.9	Resultados do método classificando websites de esquerda (E), centro-esquerda (CE), centro-direita (CD) e direita (D) na base de dados <i>Noticias-Julho</i> . Em negrito, o melhor resultado em cada métrica.	66
5.10	Resultados do método classificando websites de esquerda (E), centro-esquerda (CE), centro-direita (CD) e direita (D) na base de dados <i>Noticias-Fevereiro</i> . Em negrito, o melhor resultado em cada métrica.	67
5.11	Resultados do método classificando websites de esquerda (E), centro-esquerda (CE), centro-direita (CD) e direita (D) na base de dados <i>News-Aggregator</i> . Em negrito, o melhor resultado em cada métrica.	68
A.1	Valores de acurácia e precisão, revocação e F1 por classe para cada rede ao classificar o viés ideológico usando padrões de citações e distância de Jaccard.	82
B.1	Resultados do método classificando websites de vieses extremos (Ex) e moderados (M) na base de dados <i>Noticias-Julho</i> usando unigramas. Em negrito, o melhor resultado em cada métrica.	84
B.2	Resultados do método classificando websites de vieses extremos (Ex) e moderados (M) na base de dados <i>Noticias-Fevereiro</i> usando unigramas. Em negrito, o melhor resultado em cada métrica.	85
B.3	Resultados do método classificando websites de vieses extremos (Ex) e moderados (M) na base de dados <i>News-Aggregator</i> usando unigramas. Em negrito, o melhor resultado em cada métrica.	85
B.4	Resultados do método classificando websites de vieses de esquerda (E) e direita (D) na base de dados <i>Noticias-Julho</i> usando unigramas. Em negrito, o melhor resultado em cada métrica.	85
B.5	Resultados do método classificando websites de vieses de esquerda (E) e direita (D) na base de dados <i>Noticias-Fevereiro</i> usando unigramas. Em negrito, o melhor resultado em cada métrica.	86
B.6	Resultados do método classificando websites de vieses de esquerda (E) e direita (D) na base de dados <i>News-Aggregator</i> usando unigramas. Em negrito, o melhor resultado em cada métrica.	86

B.7	Resultados do método classificando websites de esquerda (E), centro-esquerda (CE), centro-direita (CD) e direita (D) na base de dados Notícias-Julho usando unigramas. Em negrito, o melhor resultado em cada métrica.	86
B.8	Resultados do método classificando websites de esquerda (E), centro-esquerda (CE), centro-direita (CD) e direita (D) na base de dados Notícias-Fevereiro usando unigramas. Em negrito, o melhor resultado em cada métrica.	87
B.9	Resultados do método classificando websites de esquerda (E), centro-esquerda (CE), centro-direita (CD) e direita (D) na base de dados News-Aggregator usando unigramas. Em negrito, o melhor resultado em cada métrica.	87

Sumário

Agradecimentos	iv
Resumo	vi
Abstract	vii
Lista de Figuras	viii
Lista de Tabelas	x
1 Introdução	1
1.1 Problema	2
1.2 Motivação	3
1.3 Justificativa	3
1.4 Objetivos	4
1.5 Contribuições	5
1.6 Organização da proposta	5
2 Fundamentação teórica	7
2.1 Definições	7
2.1.1 Portais e artigos de notícias	7
2.1.2 Viés e alinhamento ideológico	8
2.2 Ciência de redes	8
2.2.1 Redes complexas	9
2.2.2 Detecção de comunidades	11
2.2.3 Métricas de centralidade	14
2.2.4 Outras métricas	16
2.3 Aprendizagem de máquina	17
2.3.1 <i>Support Vector Machine</i> (SVM)	18

2.3.2	Validação cruzada <i>leave-one-out</i> (LOOCV)	19
2.3.3	Métricas de avaliação	19
2.4	Métricas de divergência	20
2.5	Considerações do capítulo	21
3	Trabalhos relacionados	22
3.1	Métodos baseados em conteúdo textual	22
3.2	Métodos baseados em hiperlinks	23
3.3	Métodos baseados em comentários	24
3.4	Métodos baseados em dados de redes sociais	24
3.5	Discussão	25
3.6	Considerações do capítulo	27
4	Detecção de viés ideológico baseada em hiperlinks	28
4.1	Método baseado em hiperlinks	28
4.1.1	Obtenção de dados	29
4.1.2	Criando as redes	30
4.1.3	Detecção de comunidades	31
4.1.4	Padrões de citações	31
4.1.5	Avaliando a classificação	33
4.2	Resultados obtidos	34
4.2.1	Coleta de dados	34
4.2.2	Construção e caracterização topológica das redes	35
4.2.3	Aplicando detecção de comunidades	41
4.2.4	Classificação baseada em padrões de citações	46
4.3	Considerações do capítulo	49
5	Detecção de viés ideológico baseada em conteúdo textual	50
5.1	Método baseado em conteúdo textual	50
5.1.1	Obtenção de dados	51
5.1.2	Preparando os dados	52
5.1.3	Computando a importância dos termos	52
5.1.4	Calculando funções de probabilidade	53
5.1.5	Computando dissimilaridades entre portais de notícias e classes de viés	55
5.1.6	Classificando e avaliando o desempenho	55
5.2	Resultados obtidos	56
5.2.1	Bases de dados obtidas	56

5.2.2	Metodologia experimental	57
5.2.3	Desempenhando tarefas de classificação bi-classe e multi-classe	59
5.3	Considerações do capítulo	69
6	Considerações finais	71
6.1	Detecção de viés ideológico via hiperlinks	71
6.2	Detecção de viés ideológico via conteúdo textual	72
6.3	Publicações	73
6.4	Trabalhos futuros	74
	Referências Bibliográficas	75
A	Resultados adicionais da classificação baseada em padrões de citações	82
B	Resultados adicionais da classificação baseada em conteúdo textual	84
B.1	Classificando extremos e moderados	84
B.2	Classificando portais de esquerda e direita	85
B.3	Classificando portais de esquerda, centro-esquerda, centro-direita e direita	86

Capítulo 1

Introdução

Com a popularização da Web e mídias sociais, websites e portais de notícias tornaram-se as principais fontes de informação para a população em geral. Para simplificar, utilizaremos os termos websites de notícias e portais de notícias como sinônimos a partir deste ponto, dado que essa simplificação não impacta nas contribuições desta pesquisa. Comparado a veículos tradicionais como jornais, televisão e rádio, portais de notícias são rápidos, têm alcance mundial e permitem interações entre a audiência (Harper, 2010; Aires et al., 2019; Dallmann et al., 2015). No entanto, assim como em mídias tradicionais, estes portais podem ter um viés na maneira como reportam notícias. Ao fazer uso de aspectos como omissão seletiva e escolha de palavras, cada fonte transmite uma impressão diferente acerca de um fato (Gentzkow e Shapiro, 2006). Isto pode impactar a forma como os leitores percebem eventos, decisões políticas e discussões relacionadas a diferentes tópicos. Estudos mostram, por exemplo, que o viés da mídia pode influenciar no resultado de eleições (Chiang e Knight, 2011; DellaVigna e Kaplan, 2007). Contudo, identificar o viés político ou ideológico da mídia é uma tarefa difícil até mesmo para humanos, dado o alto nível de subjetividade envolvido (Yigit-Sert et al., 2016). Dessa forma, desenvolver métodos automáticos para este problema é uma interessante oportunidade de pesquisa.

Neste contexto, alguns estudos têm sido desenvolvidos para automatizar o processo de detectar viés ideológico em portais de notícias, focando em diferentes aspectos como conteúdo textual (Krestel et al., 2012; Dallmann et al., 2015), hiperlinks e citações (Efron, 2004; Lin et al., 2011), e dados de redes sociais (Elejalde et al., 2017; Ribeiro et al., 2018), por exemplo. Entretanto, atualmente esses trabalhos não fornecem um método capaz de realizar precisamente a detecção de viés ideológico, em especial num contexto geral, pois focam principalmente em caracterizações, métricas e estudos de caso. Além disso, alguns dependem de fontes externas, o que é um obstá-

culo para o desenvolvimento de uma abordagem capaz de identificar diferentes vieses de maneira automática em contextos diversos.

Neste trabalho, o objetivo é apresentar abordagens mais generalistas para detectar de forma precisa o viés ideológico em portais de notícias. Não limitamos um método a uma análise exclusivamente polarizada, considerando outros níveis de viés político/ideológico, em especial extremos e moderados em ambos os espectros políticos (esquerda/direita). Além disso, nossos métodos são independentes de fontes externas, fazendo uso de informações presentes nos próprios artigos de notícias, como hiperlinks e conteúdo textual, para realizar essa classificação. Demonstramos que essa estratégia leva à construção de um método mais eficaz e menos complexo que abordagens mais tradicionais da literatura, possibilitando assim o desenvolvimento de uma ferramenta capaz de automatizar essa tarefa.

O restante deste capítulo está organizado da seguinte forma: descrição do problema (Seção 1.1), motivação (Seção 1.2), justificativa (Seção 1.3), objetivos (Seção 1.4), e, por fim, organização do restante da proposta (Seção 1.6).

1.1 Problema

Neste trabalho, tratamos do problema de detectar viés ideológico de portais de notícias na Web. Dada uma fonte, como um portal de notícias, o objetivo é determinar se a mesma expressa ideologia de esquerda ou direita, extrema ou moderada.

Muitos trabalhos que abordam este problema atualmente tratam-se de caracterizações, métricas ou estudos de caso. Outros desenvolvem de fato um método para detectar automaticamente o viés de notícias, mas dependem de fontes externas para realizar uma classificação mais precisa, ou focam apenas em um cenário polarizado (esquerda/direita, ou democrata/republicano). Outros, ainda, são muito complexos, algo que interfere na explicabilidade e desempenho do método. Diante disso, neste trabalho nosso foco é desenvolver métodos que: (i) independem de fontes externas, usando apenas aspectos do conteúdo dos artigos de notícias; (ii) foquem em uma análise mais polarizada, mas que possa ser generalizada para qualquer caso, incorporando outros níveis de viés; (iii) sejam capazes de assinalar acuradamente o viés ideológico de portais de notícias através de uma abordagem menos complexa em comparação ao estado-da-arte.

1.2 Motivação

Portais de notícias na Web, isto é, websites de notícias, tais como New York Times, Washington Post ou Huffpost, e mídias sociais como Facebook e Twitter têm sido a principal fonte de notícias de grande parte da população. Um estudo recente mostrou que 62% dos adultos acessam notícias em mídias sociais nos Estados Unidos (Gottfried e Shearer, 2016). No entanto, do mesmo modo que veículos tradicionais, notícias provenientes destes portais podem ser enviesadas, isto é, denotar um viés político/ideológico (Morstatter et al., 2018; Knoche et al., 2019). Esse é um problema de grande importância para a nossa sociedade, parcialmente porque os leitores tendem a investir pouco tempo na leitura completa de notícias e ainda menos (ou nenhum) verificando a qualidade da fonte que as publicaram (Chen et al., 2018; Blom e Reinecke Hansen, 2015). Dessa forma, esse viés ideológico pode impactar o modo como os leitores percebem eventos, decisões políticas e discussões sobre diferentes tópicos. Estudos mostram, por exemplo, que o viés da mídia pode influenciar no resultado de eleições, manipulando sistematicamente as crenças do público e favorecendo uma dada ideologia de interesse (D'Alessio e Allen, 2000; DellaVigna e Kaplan, 2008). Combinado a mídias sociais e à facilidade de compartilhamento e alcance, esse tipo de conteúdo contribui para a polarização, crimes de ódio e outras consequências de uma retórica divisiva (The Batch, 2020).

Idealmente, cada indivíduo deveria ser capaz de identificar o viés de uma fonte de notícias. Entretanto, devido à alta subjetividade associada a este processo, é necessária a classificação de especialistas, utilizando uma metodologia detalhada e trabalhosa para realizar suas classificações (Fortuna et al., 2009). Desta forma, dificulta-se a realização desta classificação em escala global e, conseqüentemente, a disseminação desta informação para o público, permitindo que os impactos negativos do viés da mídia atinjam a sociedade (Shu et al., 2017). Assim, surge a necessidade de desenvolver métodos que realizem detecção automática do viés ideológico de portais de notícias na Web, como maneira de tornar as informações mais transparentes e acessíveis para a audiência destas fontes, permitindo que o público reflita e decida se tais fontes são de fato enviesadas e se as informações transmitidas são incorretas, incompletas, enganosas ou falsas.

1.3 Justificativa

A detecção de viés ideológico em páginas web vem sendo explorada na literatura recente. Trabalhos anteriores baseiam-se em diferentes aspectos como conteúdo tex-

tual (Krestel et al., 2012; Baly et al., 2018), hiperlinks e citações (Efron, 2004; Lin et al., 2011), e dados de redes sociais (Elejalde et al., 2017; Ribeiro et al., 2018), por exemplo. Entretanto, atualmente esses trabalhos podem ser divididos em dois grupos. No primeiro grupo, há trabalhos que fornecem caracterizações, métricas para mensurar viés ideológico, ou apenas um estudo de caso, muitas vezes geograficamente localizado, focando em caracterizar inclinações em relação a um partido, principalmente no caso estadunidense (republicano/democrata). No segundo grupo, há trabalhos que desenvolvem uma classificação automática. Porém, também possuem limitações. Dentre elas, destaca-se a dependência de fontes externas para atingir uma classificação mais acurada, gerando um método pouco preciso na ausência destas; e a alta complexidade, empregando um grande conjunto de atributos comuns em processamento textual. Esses fatores impactam a explicabilidade e a eficiência dos métodos. Além disso, geralmente baseiam-se em um cenário exclusivamente polarizado (esquerda/direita, democrata/republicano). Combinado aos fatores citados previamente, constituem obstáculos para o desenvolvimento de uma abordagem eficaz e eficiente para detectar viés ideológico em um cenário mais geral, considerando diferentes níveis de vieses (mais extremos, mais moderados) em contextos diversos. Desta maneira, neste trabalho temos como objetivo desenvolver métodos capazes de incorporar essas características, aprimorando o estado-da-arte.

1.4 Objetivos

O objetivo geral deste trabalho é desenvolver e demonstrar a eficácia de métodos para detecção de viés ideológico de portais de notícias na Web baseados em aspectos extraídos dos artigos de notícias, como hiperlinks e conteúdo textual, visando resultados mais eficientes e precisos quando comparados ao estado-da-arte.

Os objetivos específicos incluem:

- Desenvolver uma abordagem para detecção de viés ideológico baseada em hiperlinks, isto é, considerando as citações realizadas pelos portais, empregando técnicas de análise de redes.
- Propor um método baseado em conteúdo textual para caracterizar dissimilaridade de discurso entre portais de notícias em uma abordagem que empregue conceitos de Teoria da Informação.
- Demonstrar a eficácia e eficiência dos métodos propostos em relação ao estado-da-arte, isto é, uma caracterização mais precisa e um método de menor complexidade

espacial que outros métodos mais tradicionais.

1.5 Contribuições

As principais contribuições geradas durante o desenvolvimento desta pesquisa foram:

- Desenvolvimento de um método que usa os hiperlinks em páginas web para detectar viés ideológico.
- Comparação de duas características estruturais de redes, verificando qual (ou quais) são mais relacionadas ao viés ideológico dos portais modelados.
- Demonstração de que a composição dos grupos ou comunidades nas redes está diretamente relacionada ao viés ideológico apenas no caso de websites de polaridades mais extremas.
- Comprovação, através de experimentos, que os padrões de citações estão mais relacionados ao viés ideológico que as comunidades, levando a um modelo capaz de obter até 76% de acurácia, considerando tanto portais extremos quanto moderados.
- Desenvolvimento de um método para detecção de viés ideológico através da análise de conteúdo textual empregando conceitos de Teoria da Informação, como entropia e divergência estatística.
- Comparação de diferentes estratégias para seleção de termos, demonstrando que escolher termos em uma faixa de entropia (10% e 75% menores valores) é útil para selecionar os termos mais relevantes.
- Demonstração, através de experimentos, da eficácia e eficiência do nosso modelo, capaz de classificar acuradamente o viés ideológico em problemas bi-classe e multi-classe empregando um conjunto de apenas quatro atributos e atingindo até 91% de acurácia e 0,91 de F1 contra acurácia de 43% e F1 de 0,46 obtidos no mesmo caso por um baseline que utiliza 282 atributos textuais.

1.6 Organização da proposta

Esta dissertação está organizada da seguinte forma: no Capítulo 2 são apresentados os fundamentos necessários para a compreensão deste trabalho; no Capítulo 3 revisamos os trabalhos relacionados, fazendo uma breve análise da literatura existente;

no Capítulo 4, relatamos a metodologia e os resultados obtidos no desenvolvimento da abordagem baseada em hiperlinks, no Capítulo 5, descrevemos a metodologia e os resultados gerados no método baseado em conteúdo textual; por fim, no Capítulo 6 incluem-se as considerações finais desta pesquisa.

Capítulo 2

Fundamentação teórica

Neste capítulo, serão descritos os fundamentos necessários para a compreensão desta dissertação. A Seção 2.1 inclui conceitos relacionados ao problema de detecção de viés em portais de notícias. A seguir, na Seção 2.2 são descritos técnicas e conceitos de ciência de redes que serão empregados neste trabalho. Na Seção 2.3 introduzimos alguns conceitos de aprendizagem de máquina, como o classificador que utilizamos, nossa estratégia de validação e métricas de avaliação. Por sua vez, na Seção 2.4 descrevemos algumas métricas para medir divergência de funções de probabilidade. Por fim, na Seção 2.5 são apresentadas as considerações do capítulo.

2.1 Definições

No decorrer desta proposta, faremos uso de diferentes termos aos nos referirmos a tópicos relacionados ao problema de detecção de alinhamento ideológico em portais de notícias. Diante disso, faz-se necessário definir o que cada um destes termos representa. A seguir, introduziremos definições necessárias à compreensão desta pesquisa.

2.1.1 Portais e artigos de notícias

O foco desta pesquisa é a detecção de alinhamento em portais de notícias na Web. Desta forma, é necessário definirmos o que esses termos representam exatamente. Para isso, elaboramos as definições a seguir.

Definição 1 *Um portal ou fonte de notícias é um website que distribui notícias através da Web. Ele pode ou não possuir uma versão em outras mídias, como jornais tradicionais e televisão.*

Definição 2 *Um **artigo de notícia**, ou apenas **artigo** é um documento hipertexto com URL acessível publicamente, divulgado em um portal de notícias.*

2.1.2 Viés e alinhamento ideológico

Viés na mídia pode ser definido em linhas gerais como uma perspectiva parcial dos fatos (Shoemaker e Reese, 1996). Este viés pode ter diferentes formas. Algumas delas são (Saez-Trumper et al., 2013; Heinze, 2006):

- **Viés de seleção ou *gatekeeping***: preferência em selecionar certas histórias ou acontecimentos a serem cobertos.
- **Viés de cobertura**: preferência em dedicar grande cobertura (tanto em tempo quanto em espaço) a certos acontecimentos ou pessoas.
- **Viés de declaração ou de posição política**: preferência por expressar comentários ou declarações de modo mais favorável (ou desfavorável) a uma ideologia ou partido político.

Diante desses conceitos, desenvolvemos a definição a seguir para nos referirmos a viés em mídias de notícias.

Definição 3 ***Viés ideológico ou orientação ideológica** de um portal de notícias é a preferência por uma ideologia ou partido político expressa através de declarações presentes no texto.*

Por fim, em nossa pesquisa estamos interessados em detectar se um conjunto de fontes possui a mesma orientação ideológica. Este conceito é definido a seguir.

Definição 4 *O **alinhamento ideológico**, ou apenas **alinhamento** de duas ou mais fontes de notícias existe quando tais fontes possuem a mesma orientação ideológica.*

2.2 Ciência de redes

A **ciência de redes** (em inglês, *network science*) é uma área interdisciplinar, envolvendo conceitos de matemática, ciência da computação e outras disciplinas. O objetivo desta ciência é a caracterização de fenômenos que podem ser modelados como redes, tais como sistemas biológicos, sociais e tecnológicos (Tiropanis et al., 2015).

Uma rede é uma representação das interações de diferentes sistemas do mundo real. A estrutura matemática utilizada para modelar redes é denominada **grafo**. Um

grafo é uma estrutura matemática e de dados composta por dois conjuntos: vértices (ou nós) e arestas (ou links), onde vértices representam indivíduos ou elementos e as arestas, os relacionamentos entre eles (Diestel, 2018). Segundo Barabási et al. (2016), em ciência de redes é comum utilizar os termos redes e grafos como sinônimos.

Outras propriedades relevantes de redes dizem respeito à direção e peso dos links. Redes podem ser **direcionadas** ou **não direcionadas**, isto é, os links podem ter uma direção, apontando de uma origem para um destino, ou nenhuma direção. Além disso, podem ser **ponderadas** ou **não ponderadas**. No primeiro caso, os links podem ter pesos que representam algo sobre o relacionamento entre os nós conectados. No segundo caso, não há pesos associados aos links.

Dessa forma, grafos podem modelar diversas redes, por exemplo, mapas, relações sociais e interações entre proteínas (West et al., 2001). Em particular, grafos podem ser utilizados para modelar a Web. A Web é uma rede direcionada, isto é, os links apontam de um documento para outro. Nesta rede, os nós são as páginas web e os links, os hiperlinks de uma página à outra. Analisar a estrutura desta rede é útil para desenvolver melhores coletores web e compreender as relações dos conteúdos presentes na Web (Broder et al., 2000). Além disso, a Web é um exemplo de **rede complexa**, assim como grande parte das redes que modelam sistemas reais.

Nesta pesquisa, possuímos interesse em identificar grupos ou comunidades formadas pelas páginas web, além de analisar quais os indivíduos mais relevantes ou influentes. A seguir, descreveremos em mais detalhes conceitos de ciência de redes que serão necessários para a compreensão do trabalho, incluindo: redes complexas, algoritmos de detecção de comunidades, métricas de centralidade e outras métricas para caracterizar redes.

2.2.1 Redes complexas

Como citado anteriormente, existe uma variedade de sistemas complexos que podem ser modelados através de redes. Alguns exemplos incluem redes de telecomunicações, relações sociais, a *World Wide Web* e a Internet, redes de citação e colaboração de cientistas, interação de proteínas e outras (Newman, 2010).

Uma característica comum a estas redes é sua grande escala, tornando difícil representá-las globalmente. Uma abordagem inicial para realizar esta representação são as **redes aleatórias**. Neste tipo de rede, um conjunto com n nós é conectado por um número fixo de links de maneira completamente aleatória. Em redes aleatórias, a distribuição do grau dos vértices segue uma **distribuição de Poisson** (Barabási et al., 2016).

Porém, redes do mundo real não possuem esta característica regular, onde todos os nós têm igual probabilidade de estarem conectados: estas redes apresentam regras mais complicadas para estabelecer conexões entre os vértices ou mesmo para sua evolução. Em muitos casos, a distribuição de grau segue a **lei da potência**, sem nenhuma característica de escala (Dehmer e Emmert-Streib, 2009; Barabási, 2003). Este fato levou à criação do conceito de **redes livres de escala**, também chamadas de **redes complexas**.

A Figura 2.1 apresenta exemplos de distribuições de grau em redes aleatórias e redes livres de escala. Observa-se que, no primeiro caso, o grau da maioria dos nós concentra-se em torno da média, enquanto no segundo caso, há muitos nós com poucos links e poucos nós com muitos links.

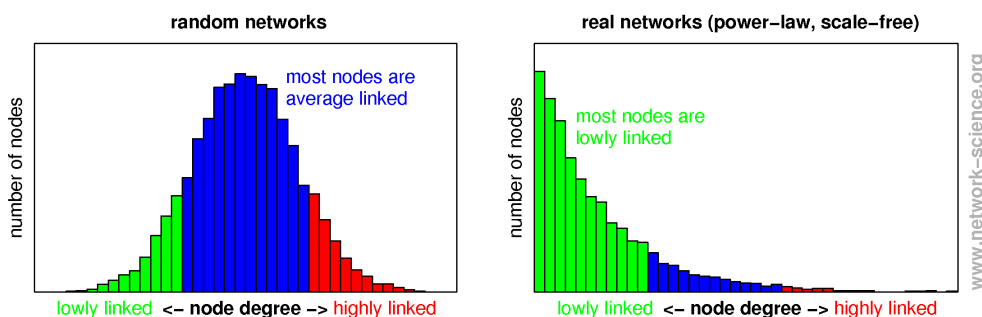


Figura 2.1: Exemplos de distribuições de grau. À esquerda, em redes aleatórias e ao lado, em redes livres de escala. Na primeira, a distribuição dos graus segue uma distribuição de Poisson, enquanto na segunda, segue a lei de potência (Scholz, 2012).

Neste contexto, há ainda as redes *small-world* (ou **mundo pequeno**), cujo nome deriva do fenômeno de mundo pequeno ou seis graus de separação. Este fenômeno tem sua origem nos experimentos de Stanley Milgram em 1967 (Travers e Milgram, 1967), onde é definido que duas pessoas arbitrárias estão conectadas por apenas seis graus de separação, ou seja, o diâmetro do grafo correspondente obtido das relações sociais não é maior que seis.

Redes *small-world* são redes em que a distância entre os vértices é pequena e possuem a característica de serem altamente conectados: a maior componente conexa da rede contém uma porção significativa dos vértices (Van Der Hofstad, 2016). Isto significa que este tipo de rede tende a apresentar um alto coeficiente de agrupamento, levando à formação de *clusters* ou comunidades. A Figura 2.2 apresenta exemplos de redes aleatórias, *small-world* e livres de escala.

Diante destes conceitos, define-se que redes complexas possuem as seguintes propriedades (Shen, 2013):

- distribuição de grau seguindo a lei de potência (*power-law degree*);
- características de mundo pequeno (*small-world*);
- alto coeficiente de agrupamento e estrutura de comunidades.

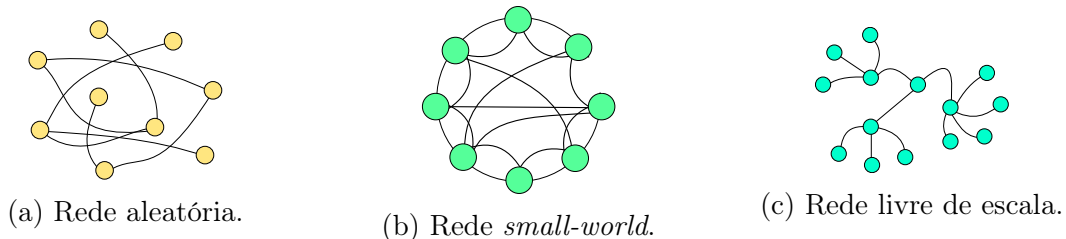


Figura 2.2: Exemplos de redes aleatórias, *small-world* e livres de escala.

2.2.2 Detecção de comunidades

Comunidades podem ser definidas como grupos de vértices mais conectados entre si que com o restante do grafo (Bedi e Sharma, 2016). Estudar comunidades pode revelar grupos de vértices que possuem alguma relação ou partilham de uma mesma característica. Na Web, por exemplo, comunidades podem denotar um conjunto de páginas acerca de um mesmo tópico ou que possuem a mesma opinião em relação a algum tema.

Algoritmo de Kernighan-Lin

O algoritmo de Kernighan-Lin, proposto por Kernighan e Lin (1970), é um método heurístico para o problema de particionamento de grafos. Dado como entrada um grafo G , ponderado e não direcionado, o objetivo é particionar os vértices de G em dois subconjuntos A e B de tamanho $n/2$, onde n é o número de vértices de G , de maneira a minimizar o custo total da soma das arestas entre os subconjuntos. Por dividir o grafo em dois subconjuntos, este algoritmo também é chamado de **bipartição**.

Inicialmente, os subconjuntos A e B são divididos de maneira aleatória. Enquanto for possível diminuir o custo das arestas entre os dois subconjuntos, o processo segue da seguinte forma:

- Através de uma heurística gulosa, são selecionados dois vértices, um de cada subconjunto, cuja troca leva à maior redução ou menor aumento do custo das arestas entre as duas partições.
- Estes vértices são trancados, ou seja, não podem mais trocar de partição nos próximos passos.

- O processo continua até que todos os vértices estejam trancados.

O custo de cada passo deste algoritmo é $O(n^3)$. Sendo r o número total de iterações, o custo total é $O(rn^3)$. Esta foi a primeira abordagem na literatura para identificar grupos em um dado grafo.

Algoritmo de Clauset, Newman e Moore

O algoritmo de Clauset et al. (2004) é um método hierárquico e aglomerativo para detecção da estrutura de comunidades de uma rede. Isto significa que, inicialmente, cada nó representa uma comunidade e, a cada etapa do algoritmo, novas comunidades vão sendo construídas a partir da aglomeração de nós. Além disso, ao final é obtido um dendrograma que representa a hierarquia das comunidades detectadas na rede.

Este algoritmo é uma melhoria do algoritmo de Girvan-Newman (Girvan e Newman, 2002), o primeiro algoritmo específico para o problema de detecção de comunidades. Diferente do algoritmo para particionamento de grafos, estes algoritmos (e seus derivados) não necessitam receber o número de grupos como entrada. A heurística de ambos os algoritmos é baseada na otimização gulosa de uma propriedade denominada **modularidade**.

A modularidade Q é uma propriedade de uma rede e um particionamento desta rede em comunidades (Girvan e Newman, 2002) cujo objetivo é medir a qualidade desta divisão. Em linhas gerais, a modularidade é a razão entre o número de links dentro de cada comunidade e o número de links entre as comunidades, subtraindo a razão esperada de um grafo completamente aleatório (Clauset et al., 2004). Ou seja, uma boa divisão é aquela onde existem muitos links entre os nós de uma mesma comunidade, e poucos links entre nós de comunidades diferentes. O valor de Q varia entre -1 e 1 . Quanto mais próximo de 1 , melhor a qualidade da estrutura de comunidades da rede.

O algoritmo recebe como entrada uma rede não-direcionada, ponderada ou não, e inicia com cada nó sendo o único membro de uma comunidade. O processo segue da seguinte forma:

- Para cada par de comunidades i, j , calcular ΔQ_{ij} , valor que indica o quanto a modularidade mudaria ao unir as comunidades i e j .
- Selecionar o maior ΔQ_{ij} e unir as comunidades i e j .
- Repetir o processo até que reste apenas uma comunidade.

Ao final do processo, a saída é um dendrograma que representa as comunidades obtidas em cada etapa. A complexidade deste algoritmo é $O(md \log n)$, onde m é o

número de links da rede, n é o número de nós e d é a profundidade do dendrograma. Segundo os autores, uma vez que redes reais são esparsas, $m \sim n$, e $d \sim \log n$. Desta forma, a complexidade pode ser expressa como $O(n \log^2 n)$.

Algoritmo Spinglass

O algoritmo Spinglass foi proposto por Reichardt e Bornholdt (2006), e possui como premissa utilizar conceitos da física para realizar detecção de comunidades em redes complexas. Dessa forma, o algoritmo é baseado no modelo de Potts Wu (1982), um modelo para fenômenos de física do estado sólido. Este modelo, por sua vez, possui base no conceito de *spins*, propriedades associadas ao movimento angular de partículas subatômicas, como prótons e elétrons (Griffiths e Schroeter, 2018).

A entrada do algoritmo é um grafo direcionado e ponderado e o número de *spins* desejados. O princípio básico é que arestas devem conectar vértices do mesmo estado de *spin* (ou comunidades), e vértices de estados diferentes devem ser desconectados. Assim, o objetivo é encontrar o estado fundamental de um modelo de Potts (Yang et al., 2016). Para isso, é utilizada a meta-heurística *simulated annealing*. A complexidade deste algoritmo em um grafo esparso é aproximadamente $O(n^{3.2})$.

Algoritmo Leading Eigenvector

O algoritmo Leading Eigenvector foi proposto por Newman (2006) e é baseado na otimização da modularidade. Porém, neste caso, a ideia central é uma otimização espectral utilizando autovalores e autovetores de uma matriz denominada matriz de modularidade. Este é um algoritmo divisivo, ou seja, o grafo completo é considerado como a primeira comunidade e a cada iteração, esta comunidade vai sendo dividida em outras.

A entrada é um grafo não-direcionado e ponderado, além do número de comunidades desejadas. Inicialmente, é calculado o autovetor para o maior autovalor positivo, e o grafo é dividido em duas partes de modo que a modularidade seja maximizada (Yang et al., 2016). Essas divisões são aplicadas enquanto for possível realizar melhoria no valor da modularidade, ou até atingir o número de comunidades dadas como entrada. A complexidade de cada divisão é $O(n(m + n))$ ou $O(n^2)$ em grafos esparsos.

Algoritmo de Louvain

O algoritmo de Louvain, assim como alguns dos algoritmos anteriores, também é uma heurística gulosa para otimização da modularidade e foi proposto por Blondel et al. (2008). Trata-se de um algoritmo aglomerativo, ou seja, as comunidades vão sendo

construídas a partir da união de outras comunidades, até que reste apenas uma, correspondente ao grafo dado como entrada.

A entrada deste algoritmo é um grafo não-direcionado e ponderado. O processo é dividido em duas etapas (Aires e Nakamura, 2017). Na primeira, cada vértice é considerado uma comunidade. Após isso, para cada vértice i é avaliado o ganho de modularidade que ocorreria se tal vértice fosse removido de sua comunidade e colocado na comunidade de seus adjacentes. Após avaliar todos os vizinhos, o vértice i é colocado na comunidade onde o ganho é máximo, mas apenas se o ganho é positivo. Se não for, o vértice permanece em sua comunidade. Este processo se repete até que nenhum indivíduo possa melhorar a modularidade da rede. Na segunda etapa, o algoritmo considera cada comunidade construída na primeira fase como vértices. Para isto, os pesos das conexões entre os novos vértices são dados pela soma dos pesos das arestas entre vértices nas duas comunidades correspondentes. Conexões entre vértices na mesma comunidade tornam-se laços no novo grafo. Uma vez que a segunda etapa esteja concluída, a primeira etapa é reaplicada no novo grafo. O processo se repete até que não haja mais alterações positivas na modularidade. Cada passo deste algoritmo, isto é, repetições da primeira e segunda etapas, possui complexidade $O(n \log n)$.

2.2.3 Métricas de centralidade

Métricas de centralidade são medidas que respondem quais são os nós centrais ou mais influentes do grafo, seguindo um critério específico (Golbeck, 2015). Neste trabalho, utilizamos as seguintes medidas de centralidade.

Centralidade de grau e grau ponderado

A centralidade de grau diz respeito ao grau do vértice, isto é, o número de arestas conectadas a ele (Perra e Fortunato, 2008). Em ciência de redes, o grau indica que um indivíduo com maior número de conexões é mais central, isto é, possui maior influência na rede.

O grau **simples** contabiliza apenas o número de links. Se a rede for ponderada, pode ser computado também o grau **ponderado**, considerando o peso dos links. Sendo w o peso associado aos links de uma rede, o grau ponderado GP (ou simples) de um nó v é dado pela Equação 2.1.

$$c_G(v) = \sum_{u \neq v} w_{u,v}, \quad (2.1)$$

onde $w_{u,v}$ é o peso do link entre u e v . Caso deseje-se calcular o grau simples, este peso corresponde a 1.

No caso de redes direcionadas, é possível dividir a análise ao considerar os links incidentes, ou seja, que apontam para v ; e links adjacentes, saindo de v . Neste caso, o grau é de entrada (dado pela Equação 2.2) e de saída (dado pela Equação 2.3). O grau total seria a soma do grau de entrada e do grau de saída.

$$c_{G_{\text{entrada}}}(v) = \sum_{u \neq v} w_{u,v}. \quad (2.2)$$

$$c_{G_{\text{saída}}}(v) = \sum_{u \neq v} w_{v,u}. \quad (2.3)$$

Centralidade de autovetor

A **centralidade de autovetor** (do inglês, *eigenvector centrality*) é baseada no princípio de que a importância de um nó depende da importância de seus vizinhos. Isto significa que nós mais centrais possuem mais conexões de qualidade (Newman, 2010).

Seja A a matriz de adjacência da rede, isto é, a matriz que representa os relacionamentos entre os nós, onde A_{uv} indica o peso do link entre u e v . Considerando que um nó v possui um prestígio x_v , tal prestígio é proporcional à soma dos prestígios dos vizinhos que apontam para v , isto é, links incidentes. Tal prestígio é o componente v do autovetor da matriz A transposta com autovalor λ , como expresso na Equação 2.4 (Perra e Fortunato, 2008).

$$\lambda x_v = \sum_{u:u \rightarrow v} x_u = \sum_u A_{uv} x_u = (A^t x)_v. \quad (2.4)$$

Centralidade de intermediação

A **centralidade de intermediação** (do inglês, *betweenness centrality*) mede o quanto um nó faz parte dos caminhos entre dois vértices (Golbeck, 2015). Dessa forma, considera que um nó mais central funciona como ponte entre mais nós na rede.

Seja V o conjunto de vértices, $\sigma(s, t)$ é o número de caminhos mínimos entre s e t , e $\sigma(s, t|v)$ é o número de caminhos mínimos entre s e t que passam por v . A intermediação de um nó v é dada pela Equação 2.5 (Brandes, 2008).

$$c_B(v) = \sum_{s,t \in V} \frac{\sigma(s, t|v)}{\sigma(s, t)} \quad (2.5)$$

2.2.4 Outras métricas

Além de métricas de centralidade, é possível mensurar outras propriedades de redes. Tais propriedades podem ser úteis ao classificar uma rede como livre de escala, mundo pequeno ou aleatória. Neste trabalho, faremos uso das métricas a seguir para realizar caracterizações de diferentes redes.

Coeficiente de agrupamento

O **coeficiente de agrupamento** se trata de uma métrica para indicar a tendência dos nós de uma rede a formarem grupos (Wasserman et al., 1994). Em uma rede, o coeficiente de agrupamento aponta a coesão dos grupos em uma única medida.

Existem duas versões desta métrica: global e local (Barabási et al., 2016). O coeficiente de agrupamento global, também denominado **transitividade**, indica o total de triângulos fechados em uma rede. Esta métrica oferece uma indicação do agrupamento da rede inteira e pode ser aplicado a redes direcionadas ou não-direcionadas, ponderadas ou não. É calculado segundo a Equação 2.6.

$$C = \frac{3 \times \text{número de triângulos}}{\text{Número de triplas conectadas}} \quad (2.6)$$

onde uma tripla conectada é um conjunto de três nós ABC de modo que A esteja conectado com B e B esteja conectado com C .

Já o coeficiente de agrupamento local captura a tendência de os vizinhos de um nó estabelecerem links entre si. Assim, para um nó i com grau k_i , o coeficiente de agrupamento local é dado pela Equação 2.7, abaixo.

$$C_i = \frac{2L_i}{k_i(k_i - 1)} \quad (2.7)$$

onde L_i representa o número de links entre os k_i vizinhos do nó i .

Dessa forma, C_i quantifica a densidade local de links em uma rede, ou seja, o quão densamente interconectados são os vizinhos de um nó nesta rede. O grau de agrupamento para a rede completa pode ser indicado pelo **coeficiente de agrupamento médio**, ao fazer a média dos coeficientes de agrupamento locais dos nós de uma rede, segundo a Equação 2.8, a seguir.

$$\langle C \rangle = \frac{1}{n} \sum_{i=1}^n C_i \quad (2.8)$$

O coeficiente de agrupamento global e médio, embora capturem o agrupamento dos nós em uma rede, utilizam estratégias diferentes e, portanto, não são iguais (Bol-

lobás e Riordan, 2003). Neste trabalho, utilizaremos ambas as métricas para observar o agrupamento de uma rede.

Conectividade, diâmetro e densidade

A conectividade de uma rede diz respeito à existência de caminhos entre todos os nós. Uma rede é dita **conexa** se há caminhos entre todos os pares de nós. Caso não exista caminho entre pelo menos um par de nós, a rede é **desconexa** (Barabási et al., 2016).

Uma rede desconexa é dividida em sub-redes denominadas componentes, que podem ser definidos como um subconjunto dos nós de modo que existe um caminho entre quaisquer pares de nós pertencentes ao componente, mas não caminhos para nós que não pertencem a este componente. A identificação destas componentes conexas é útil para caracterizar redes complexas, uma vez que uma de suas propriedades é a presença de uma componente conexa contendo uma quantidade significativa de nós (Van Der Hofstad, 2016).

Por sua vez, o **diâmetro** é o comprimento do maior caminho entre dois vértices em uma rede (Newman, 2010). Em redes complexas que apresentam propriedades de mundo pequeno, o valor do diâmetro é relativamente pequeno, como no experimento de Travers e Milgram (1967), onde o diâmetro da rede social é 6. Dessa forma, essa métrica pode ser útil para indicar se uma rede possui propriedades de redes livres de escala.

Por fim, a **densidade** de um grafo indica o quão próximo o número de arestas está da quantidade de links existentes em um grafo completo, onde todos os vértices estão conectados a todos os outros. A densidade em um grafo direcionado é calculada segundo a Equação a seguir, onde m é o número de arestas e n é o número de vértices (Diestel, 2018).

$$d = \frac{m}{n(n-1)} \quad (2.9)$$

Redes complexas possuem como propriedade serem altamente esparsas (Barabási, 2003). Assim, utilizaremos a densidade das redes como mais uma ferramenta para avaliar o quanto o relacionamento modelado se aproxima de uma rede livre de escala.

2.3 Aprendizagem de máquina

A **aprendizagem de máquina** preocupa-se com reconhecimento de padrões (Bishop, 2006) e aprendizado computacional no contexto de inteligência artificial, ou seja, como

construir programas capazes de melhorar automaticamente através de experiência (Mitchell, 1997). Técnicas de aprendizagem de máquina têm sido utilizadas em aplicações nas áreas de administração, comércio, indústria, dentre outras (Paliouras et al., 2001).

Modelos de aprendizagem de máquina podem ser supervisionados ou não supervisionados. No primeiro caso, os vetores de entrada são dados juntamente com seus vetores-alvo, ou seja, valores reais que correspondem às instâncias, como classes ou um valor a ser calculado. No segundo caso, os dados não são acompanhados dos seus vetores-alvo (Bishop, 2006). Além disso, os modelos também podem ser de classificação ou regressão, onde a diferença é o tipo de predição que se deseja realizar: um valor discreto, categórico; ou um valor real, respectivamente (Mitchell, 1997).

No caso de problemas de classificação, os modelos são denominados **classificadores**, cuja principal função é analisar um conjunto de dados buscando reconhecer padrões que permitam categorizá-los, ou seja, separá-los em categorias (Dietterich, 2000). Dentre os mais utilizados, podemos citar: SVM (*Support Vector Machine*, ou Máquina de Vetores de Suporte), Naive Bayes, kNN (*k-Nearest Neighbor*, ou k-Vizinhos mais Próximos) e RandomForest (Floresta Aleatória) (Vinodhini e Chandrasekaran, 2012). A seguir, descreveremos os conceitos empregados neste trabalho em mais detalhes.

2.3.1 *Support Vector Machine (SVM)*

Neste trabalho, o classificador escolhido em nosso modelo é algoritmo *Support Vector Machine* (SVM). Este método projeta cada instância em um espaço vetorial, onde vetores marginais são usados para definir a separação entre as classes. A ideia é encontrar um hiperplano, representado pelo vetor \vec{w} , que além de separar os vetores das instâncias de cada classe, maximize a margem de separação entre as classes. Dessa forma, constitui-se um problema de otimização. Seja c_j a classe correta da instância s_j , podemos descrever a solução através da equação (Pang et al., 2002)

$$\vec{w} = \sum_j \alpha_j c_j \vec{s}_j \quad \alpha_j \geq 0, \quad (2.10)$$

onde α_j são obtidos através da resolução de um problema dual. Os vetores \vec{s}_j cujos α_j são maiores que zero são chamados vetores de suporte, uma vez que são os únicos vetores que contribuem para \vec{w} . Assim, constrói-se um hiperplano com base nas instâncias mais próximas ao vetor de separação. A classificação das instâncias de teste consiste em verificar o lado do hiperplano \vec{w} onde elas se posicionam.

2.3.2 Validação cruzada *leave-one-out* (LOOCV)

Em aprendizagem de máquina, há diferentes formas de realizar o treino, isto é, a etapa em que o modelo usa as instâncias para aprender padrões úteis. Para evitar problemas como *overfitting* (superajuste) ou *underfitting* (subajuste), utilizam-se técnicas de validação cruzada. Neste trabalho, empregamos validação cruzada *leave-one-out* (LOOCV).

Digamos que temos um conjunto de S de n instâncias. Neste tipo de validação cruzada, o modelo é treinado e testado n vezes. A cada iteração $t \in 1, 2, \dots, n$, o modelo é treinado em $S - S_t$ e testado (ou validado) em S_t . Assim, é feita uma validação cruzada de forma exaustiva, onde todas as instâncias fazem parte dos conjuntos de treino e teste a cada vez (Wong, 2015). Esse tipo de validação cruzada é indicado especialmente em casos onde o número de instâncias não é muito grande.

2.3.3 Métricas de avaliação

Para avaliar o desempenho de métodos de aprendizagem de máquina, são necessárias métricas de avaliação. Neste trabalho, estamos interessados especificamente em métricas para aprendizagem supervisionada, isto é, problemas de classificação. Dentre elas, empregamos as métricas descritas a seguir. Considerando que em um problema de classificação, temos uma classe negativa N e uma classe positiva P , podemos definir as métricas de avaliação do seguinte modo (Baeza-Yates e Davis, 2004). A **precisão** é dada pela equação

$$\text{prec} = \frac{tp}{tp + fp}, \quad (2.11)$$

onde tp é o número de instâncias da classe positiva que foram corretamente classificadas, e fp o número de instâncias que foram classificadas como positivas, mas pertenciam à classe negativa. Essa métrica indica o sucesso do classificador em assinalar o rótulo positivo a uma instância.

A **revocação** é dada pela equação

$$\text{revoc} = \frac{tp}{tp + fn}, \quad (2.12)$$

onde tp é o número de instâncias da classe positiva que foram corretamente classificadas, e fn o número de instâncias da classe positiva que foram incorretamente classificadas como negativas. Essa métrica indica quantas instâncias positivas o classificador conseguiu recuperar.

Além disso, a temos a **medida F1** é a média harmônica das métricas de precisão e revocação, e, portanto, indica um equilíbrio entre as métricas. É dada pela equação

$$F1 = 2 \cdot \frac{prec \cdot revoc}{prec + revoc}. \quad (2.13)$$

Outra métrica mais geral é a **acurácia**, dada pela equação

$$acc = \frac{tp + tn}{tp + tn + fp + fn}, \quad (2.14)$$

onde tp é o número de instâncias da classe positiva que foram corretamente classificadas, fp o número de instâncias que foram classificadas como positivas, mas pertenciam à classe negativa, tn o número de instâncias da classe negativa corretamente classificadas e fn o número de instâncias da classe positiva que foram incorretamente classificadas como negativas. Essa métrica indica o número total de instâncias que foram corretamente classificadas, considerando ambas as classes.

Por fim, uma última forma de avaliação é a matriz de confusão, dada pela Tabela 2.1. Ela permite visualizar o desempenho do classificador de uma forma mais sucinta, comparando o número de instâncias classificadas corretamente e incorretamente em cada classe.

Tabela 2.1: Matriz de confusão.

		Classe prevista	
		P	N
Classe real	P	tp	fn
	N	fp	tn

2.4 Métricas de divergência

O último conceito essencial em nosso trabalho são métricas de divergências, uma maneira de medir a diferença ou distância entre duas funções de probabilidade. Decidimos empregar algumas dessas métricas, que possuem aplicações em diferentes contextos, para compará-las. São elas a distância de Jaccard e divergência de Jensen-Shannon. Tais métricas foram selecionadas, dentre diversas outras, por serem opções mais sensíveis ao cômputo de pequenas diferenças em distribuições probabilísticas. Por exemplo, uma métrica comum em Recuperação de Informação, a distância do cosseno (Baeza-Yates et al., 1999), compara apenas o ângulo entre distribuições de probabilidade, ignorando outras propriedades. Essas particularidades justificam nossa escolha em selecionar e comparar estas duas métricas em específico.

A distância de Jaccard (Levandowsky e Winter, 1971; Lipkus, 1999) entre p e q é comumente usada em domínios da Biologia e Ciência da Computação para medir a diferença entre vetores em espaços \mathbf{R}^n e é dada por

$$\text{jac}(p, q) = 1 - \frac{\sum_{i=1}^n \min(p_i, q_i)}{\sum_{i=1}^n \max(p_i, q_i)}. \quad (2.15)$$

A divergência de Jensen-Shannon entre p e q é baseada na entropia de Shannon, que é dada por (Shannon, 1948)

$$H(p) = - \sum_{i=1}^n p_i \log p_i, \quad (2.16)$$

onde p é uma função de probabilidade. Assim, a divergência de Jensen-Shannon é definida como (Lin, 1991)

$$\text{jsd}(p, q) = \frac{H(p) + H(q)}{2} - H\left(\frac{p+q}{2}\right) \quad (2.17)$$

onde $H(\cdot)$ é a entropia de Shannon como definida na equação (2.16). Em geral, a divergência de Jensen-Shannon é uma métrica robusta para medir a diferença entre funções de probabilidade. De forma simples, a divergência de Jensen-Shannon quantifica o número de bits que diferem entre as funções comparadas (Lin, 1991) e está fortemente relacionada ao conceito de informação mútua (Endres e Schindelin, 2003).

2.5 Considerações do capítulo

Neste capítulo, apresentamos os fundamentos teóricos utilizados para o desenvolvimento deste trabalho, incluindo os conceitos de viés em portais de notícias e relacionados; além de fundamentos de ciência de redes, aprendizagem de máquina e métricas para medir divergência estatística entre funções de probabilidade. No próximo capítulo, serão apresentados os trabalhos relacionados, que também fazem uso de alguns conceitos apresentados neste capítulo.

Capítulo 3

Trabalhos relacionados

Neste capítulo, apresentaremos uma breve revisão sobre métodos para detecção de viés ou orientação ideológica em portais de notícias. Observamos que, na literatura, são considerados diferentes aspectos para realizar esta análise. Desta forma, listaremos os trabalhos segundo o aspecto focado na detecção de viés, dividindo-os em baseados no conteúdo propriamente dito (Seção 3.1) e baseados em outros elementos, como hiperlinks presentes nas páginas web (Seção 3.2), comentários deixados nas notícias (Seção 3.3) e dados provenientes de redes sociais (Seção 3.4). Na Seção 3.5 é feita uma discussão acerca dos trabalhos apresentados. Por fim, na Seção 3.6 são feitas as considerações do capítulo.

3.1 Métodos baseados em conteúdo textual

O aspecto mais natural ao ponderar sobre detecção de orientação ideológica em notícias é o próprio conteúdo, ou seja, o texto presente nos artigos de notícias. A seguir, descrevemos trabalhos que se baseiam em análise textual para desenvolver métodos para detecção de viés em artigos de notícias.

Dallmann et al. (2015) realizaram uma análise comparativa para identificar a orientação política de portais de notícias alemães. Eles construíram uma base de dados composta de notícias políticas e econômica de quatro conhecidos portais alemães, cobrindo um período de quatro anos. Os autores propõem métricas que indicam viés relativo a um partido político, incluindo menções a partidos ou a seus membros, e o sentimento associado a elas. Os resultados mostraram que os portais analisados possuíam viés com relação a partidos específicos, o que coincidia com a percepção pública.

Também no contexto da Alemanha, Krestel et al. (2012) apresentaram uma abordagem para identificar viés no vocabulário de documentos usando métodos de recupe-

ração de informação. Eles compararam o vocabulário usado em discursos oriundos do Parlamento Alemão com aquele usado em artigos de quatro jornais e revistas populares usando um modelo de espaço vetorial. Como resultado, observaram que o uso do vocabulário revela que os jornais possuíam viés em relação a uma ideologia política, o que também coincidiu com a percepção humana das fontes analisadas.

Em um contexto mais geral, Baly et al. (2018) desenvolveram um método para prever factualidade e viés político de mídias de notícias. Eles experimentaram com um conjunto variado de atributos, dentre eles: um conjunto de 282 atributos textuais, como léxicos, POS tags, *scores* de sentimento e polaridade, subjetividade e moralidade extraídos do título e conteúdo dos artigos de notícias, além de informações extraídas do Twitter e Wikipedia. Os autores mostram que essa abordagem foi melhor sucedida para detectar factualidade que viés ideológico, onde a acurácia foi de 41,74%.

3.2 Métodos baseados em hiperlinks

Artigos de notícias são disponibilizados em páginas web. Um aspecto deste tipo de documento são os hiperlinks, que podem fazer referência a outras partes do próprio documento ou a outros documentos. Os trabalhos descritos abaixo utilizam os hiperlinks presentes em páginas web de artigos de notícias, ou outros documentos hipertexto, na tarefa de detecção de orientação política.

Efron (2004) introduziu um método baseado em hiperlinks para estimar a orientação política de documentos hipertexto usando informações de cocitação. Dado um documento de interesse, o objetivo era estimar se ele está associado a comunidades de esquerda ou de direita. Para isso, o método faz uso de um modelo probabilístico e avalia a probabilidade de cocitação entre o documento e algumas sementes, ou seja, documentos cuja orientação é previamente conhecida. Como resultado, os autores mostram que o modelo superou classificadores baseados em léxicos como *naive Bayes* e SVM.

Lin et al. (2011), por sua vez, propuseram métricas empíricas para quantificar o alcance e as dinâmicas de viés em notícias, com o objetivo de demonstrar uma inclinação em favor de partidos políticos específicos, figuras públicas ou certas regiões geográficas. Tais métricas são aplicadas em artigos de notícias oriundos de portais convencionais e artigos de blogs. Dentre os resultados, os autores fizeram uma observação interessante ao investigar os hiperlinks embutidos nos artigos. Os padrões de hiperlinks sugeriram que fontes inclinadas ao partido democrata possuem maior chance de realizarem citações entre si do que fontes com inclinação ao partido republicano. Isto é, fontes com viés de esquerda recebem mais citações que fontes com viés de direita, e

essas citações são feitas tanto por portais de esquerda quanto de direita.

3.3 Métodos baseados em comentários

Outro aspecto que pode ser explorado em artigos de notícias são os comentários deixados nas páginas. Tais comentários podem ser fontes de informações adicionais sobre as notícias, além de, em geral, serem textos menos complexos que o conteúdo do artigo. Os trabalhos a seguir utilizam deste recurso para o desenvolvimento de suas abordagens para detecção de viés ideológico em portais de notícias.

A abordagem proposta por Yigit-Sert et al. (2016) possui como objetivo agrupar portais de notícias que tratem dos mesmos aspectos acerca de diversos tópicos. Para isso, a ideia central é o uso dos comentários deixados nas notícias em adição ao texto do próprio artigo. Os autores argumentam que comentários tendem a ser mais explícitos que o conteúdo do artigo, podendo, dessa forma, ser úteis na tarefa de inferir aspectos de um tópico de notícias. Apesar de ser um trabalho inicial, os autores indicam que a análise do conteúdo e dos comentários associados é uma direção promissora para esta tarefa, em especial quando há um grande número de comentários disponíveis.

Um trabalho mais completo foi apresentado por Park et al. (2011), que introduziram uma abordagem para detecção de orientação política de artigos de notícias focando na análise de sentimentos de comentários deixados nestas páginas. O método é baseado na intuição de que existem usuários com preferência política explícita. A partir disso, são detectados padrões de sentimentos expressos por tais usuários, de modo a permitir uma classificação do viés ideológico de artigos de notícias. Por exemplo, se um comentário é negativo, a orientação política do artigo é contrária à do usuário. Como resultados, foi obtida uma acurácia de 80% ao utilizar esta abordagem para classificar artigos como conservadores ou liberais.

3.4 Métodos baseados em dados de redes sociais

Além de aspectos presentes nas páginas web que contém artigos de notícias, dados de redes sociais podem ser recursos úteis para detectar o viés político de um portal de notícias. A seguir, descrevemos trabalhos que utilizam aspectos provenientes de redes sociais em suas abordagens.

Saez-Trumper et al. (2013) realizaram um estudo sobre a presença de três tipos de viés (seleção ou *gatekeeping*, cobertura e declaração) em portais de notícias e nas comunidades de mídias sociais relacionadas a esses portais, mais especificamente no

Twitter. Para isso, métodos não supervisionados são introduzidos, incluindo a construção de um grafo baseado na similaridade entre os textos dos artigos de notícias, onde cada componente conectado representa um grupo de artigos que abordam o mesmo tema. Como resultados, os autores observaram que os grupos no grafo estavam mais relacionados a variáveis geográficas do que orientação política. Por sua vez, o viés político foi mais evidente em *tweets* relacionados aos temas.

Por sua vez, Elejalde et al. (2017) utilizaram *tweets* para computar automaticamente a orientação política e sócio-econômica de portais de notícias no contexto da mídia chilena. Os autores propuseram categorizar os portais a partir da opinião demonstrada por eles acerca de tópicos relevantes e controversos. Para isso, são utilizados *tweets* publicados em suas mídias no Twitter, que por sua vez são mapeados para um questionário denominado O Menor Questionário Político do Mundo (ou *The World's Smallest Political Quiz*)¹. Dessa forma, os portais são posicionados em um plano cartesiano que informa a respeito de suas orientações políticas. Além disso, os autores perceberam que o viés político se reflete no vocabulário escolhido e nas entidades tratadas pelos portais.

Outra estratégia presente na literatura é o uso de *word embeddings* para detectar viés político em *tweets* (Rao e Spasojevic, 2016; Gordon et al., 2020). Em particular, Rao e Spasojevic (2016) define se um *tweet* está mais inclinado aos partidos republicano ou democrata com uma acurácia de até 87,57%. Gordon et al. (2020) não descreve o desempenho do classificador, mas utiliza *word embeddings* para demonstrar que, devido aos *tweets* de Donald Trump, a categoria de candidatos republicanos atinge um *score* de viés de 0,97, quase o maior valor de intensidade de viés, que seria igual a 1.

Por fim, Ribeiro et al. (2018) empregaram uma metodologia diferente para inferir o viés político de fontes de notícias em mídias sociais como Facebook e Twitter. A ideia principal é fazer uso das interfaces de anúncios, que oferecem informações demográficas sobre a audiência de uma dada página. Desse modo, os autores mostraram que a orientação ideológica (liberal ou conservadora) de uma fonte de notícias pode ser estimada a partir da audiência da mesma, ou seja, a partir da presença de leitores liberais e conservadores dentre o público daquela fonte.

3.5 Discussão

Os métodos apresentados até então foram classificados de acordo com a abordagem escolhida para realizar a detecção de orientação ideológica em portais de notícias e

¹<https://www.theadvocates.org/quiz/>

outras fontes, como *tweets*. A Tabela 3.1 sumariza os trabalhos segundo as estratégias desenvolvidas, as classes de viés ideológico tratadas e o alvo da classificação.

Tabela 3.1: Sumarização dos trabalhos relacionados.

Autor/ano	Estratégia	Classes	Desempenho	Alvo
Dallmann et al. (2015)	Métricas de similaridade	Partidos alemães	-	Portais de notícias alemães
Krestel et al. (2012)	Vetores TF-IDF e distância do cosseno	Partidos alemães	-	Portais de notícias alemães
Baly et al. (2018)	Um conjunto variado de atributos incluindo léxicos e polaridade	Esquerda/centro/direita	41,74% (acurácia)	Portais de notícias
Efron (2004)	Co-citações/hiperlinks	Liberal/conservador	77,50% (acurácia)	Documentos web
Lin et al. (2011)	Métricas empíricas	Democrata/republicano	-	Portais de notícias
Yigit-Sert et al. (2016)	Análise de comentários	Não define	-	Portais de notícias
Park et al. (2011)	Análise de sentimentos em comentários	Liberal/conservador	80% (acurácia)	Artigos de notícias e usuários
Saez-Trumper et al. (2013)	Análise de redes e <i>tweets</i>	Liberal/conservador	-	Portais de notícias
Elejalde et al. (2017)	Diferença <i>rank</i>	Liberal/conservador	-	Portais de notícias chilenos
Rao e Spasojevic (2016)	<i>Word embeddings</i> e LSTM	Democrata/republicano	87,57% (acurácia)	<i>Tweets</i>
Gordon et al. (2020)	<i>Word embeddings</i>	Democrata/republicano	-	<i>Tweets</i>
Ribeiro et al. (2018)	Estatísticas de audiência em mídias sociais	Liberal/moderado/conservador	-	Portais de notícias

A sumarização dos trabalhos esclarece que vários dos trabalhos oferecem apenas um estudo de caso ou uma caracterização muito específica, analisando apenas um conjunto limitado de fontes. Métodos com uma abordagem mais geral, como os de Efron (2004), Baly et al. (2018) e Park et al. (2011) também possuem limitações. No primeiro caso, o método não tem um bom desempenho ao classificar páginas web com poucos hiperlinks. No segundo, há um conjunto de muitos atributos, 282 no total. Isso pode ter impactos em tempo de processamento e explicabilidade. E o terceiro, embora tenha tido um bom desempenho, depende de uma fonte externa, os comentários, que não estão sempre disponíveis em artigos de notícias.

Para tratar estas limitações, nesta pesquisa focamos no desenvolvimento de métodos para identificar o viés ideológico de portais de notícias na Web que possuam as seguintes características: (i) independente de fontes externas, utilizando os artigos de notícias para extrair atributos úteis para a tarefa; (ii) focado na classificação de portais de notícias, não *tweets* ou comentários; (iii) focado em uma análise mais polarizada, mas que possa ser generalizada para qualquer caso. Nesse contexto, acreditamos que a abordagem descrita por Baly et al. (2018), em sua versão utilizando apenas atributos extraídos dos artigos de notícias, é a mais adequada como *baseline* comparativo, dentre os trabalhos relacionados listados acima, por ser mais próximo ao nosso cenário.

3.6 Considerações do capítulo

Neste capítulo, apresentamos uma breve revisão dos trabalhos relacionados que tratam do problema de detecção de viés ou orientação ideológica em portais de notícias na Web. Os trabalhos foram divididos segundo os aspectos/estratégias escolhidas: análise de texto, hiperlinks, comentários e dados de redes sociais. Ao considerar os trabalhos, percebemos que existe oportunidade de desenvolver métodos com melhor desempenho que o estado-da-arte, que dependam apenas do conteúdos dos próprios artigos de notícias e focados especificamente para a detecção de viés ideológico em portais de notícias. No próximo capítulo, será apresentada nossa abordagem para detecção de viés ideológico com base em hiperlinks.

Capítulo 4

Detecção de viés ideológico baseada em hiperlinks

Neste capítulo, apresentaremos nossa abordagem para detecção de viés ideológico em portais de notícias com base nos hiperlinks presentes em artigos desses websites. Ao desenvolver esse método, o objetivo foi verificar se é possível fazer uso de características estruturais para caracterizar ou identificar o viés ideológico de portais de notícias. A hipótese é que websites estabelecem hiperlinks para outras fontes cujas orientações ideológicas estejam alinhadas às suas próprias. Caso essa hipótese seja verdadeira, uma modelagem baseada em redes deveria ser capaz de capturar essas nuances e poderia ser útil para o desenvolvimento de um método de classificação.

Os resultados são apresentados do seguinte modo: na Seção 4.1 descrevemos a metodologia desenvolvida para a abordagem baseada em hiperlinks. Em seguida, na Seção 4.2 apresentamos os resultados obtidos após a implementação do método, incluindo caracterizações topológicas e análise de duas propriedades estruturais: comunidades e padrões de citações. Por fim, a Seção 4.3 contém as considerações deste capítulo.

4.1 Método baseado em hiperlinks

A primeira abordagem que exploramos para a tarefa de detectar alinhamento ideológico é baseada em hiperlinks, isto é, citações entre websites, identificadas através dos hiperlinks embutidos nas páginas web. Assim como indicado no trabalho de Efron (2004), e como observado em outros contextos na Web, citações de uma página a outra podem estar relacionadas a uma ideia de concordância entre as fontes. Diante disso, a ideia é verificar se essa relação realmente pode ser observada e, portanto, se podemos fazer uso dessas informações de hiperlinks, em particular de características estruturais

de redes construídas com base nesses hiperlinks, para caracterizar o viés ideológico de portais de notícias. A Figura 4.1 traz em detalhes as etapas dessa abordagem. A seguir, descrevemos os pormenores de cada uma.

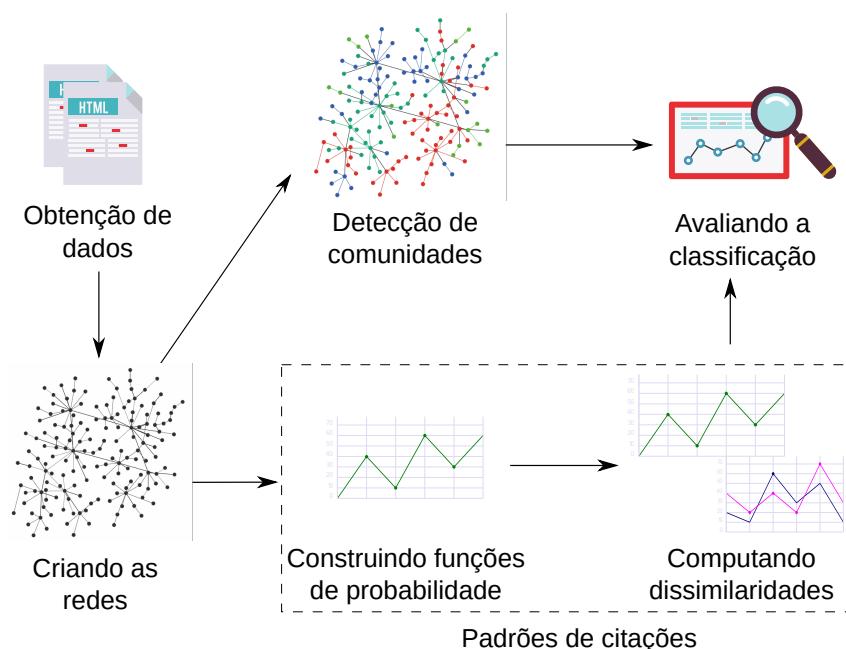


Figura 4.1: Visão geral da abordagem baseada em hiperlinks.

4.1.1 Obtenção de dados

Antes de iniciar a abordagem de detecção de viés propriamente dita, é importante descrever quais são e como são obtidos os dados necessários. Primeiramente, precisamos de uma lista, ou uma base de dados, de websites com viés ideológico previamente conhecido. Ou seja, nosso objetivo é utilizar uma classificação externa, de preferência mantida por especialistas, que realizam uma classificação de websites com relação a viés ideológico de forma subjetiva, seguindo parâmetros específicos. Portanto, destacamos que nossa intenção não é assinalar o viés ideológico de websites, mas sim desenvolver um modelo capaz de utilizar a classificação manual de especialistas para identificar padrões que permitam caracterizar esse viés.

Assim, é necessário obter uma fonte que realize essa anotação manual segundo uma metodologia específica e disponibilize esses dados publicamente. Dois exemplos de fontes são os websites Media Bias Fact Check (2019) e AllSides (2020). Ambos disponibilizam uma classificação de diversos websites em classes/níveis de viés ideológico. Assim, fazemos uso de fontes como essas para obter uma lista de websites a serem coletados, e o rótulo ou classificação atribuída quanto à orientação ideológica

dos mesmos. Nesta etapa, damos prioridade a portais de notícias, excluindo fóruns e outros tipos de websites que não façam publicação de artigos de notícias.

Essa estratégia, isto é, a escolha por construir uma base de dados própria utilizando esses critérios, deu-se principalmente pelo fato de bases públicas já existentes possuírem pouca variedade no número de fontes coletadas e pouca representação dos diferentes vieses. Por exemplo, a base All the News¹ inclui um grande número de artigos, porém, esses artigos são oriundos de apenas 15 websites. Destes 15, a grande maioria é orientada mais à esquerda. Portanto, decidimos definir uma estratégia para construir bases de dados mais adequadas para nosso problema específico.

Dessa forma, construímos uma coleção de documentos que representam o discurso de websites de diferentes vieses (esquerda/direita, extremos/moderados), selecionados no passo anterior. Os websites são dados como entrada para um coletor web, que realizará o *download* de um número limitado de artigos de notícias de cada website. Ou seja, nossa coleção de dados possui a seguinte estrutura:

- O nome (ou domínio) do portal de notícias.
- Um conjunto de páginas web (documentos HTML) que foram coletadas desse portal de notícias específico.
- A orientação ideológica desse portal de notícias, atribuída por outra fonte, que realize essa classificação de modo manual.

Esses dados, assim, representam uma amostra do discurso de cada website e serão utilizados para o desenvolvimento dos métodos de classificação/identificação de viés ideológico.

4.1.2 Criando as redes

Iniciamos a abordagem utilizando a coleção de páginas web coletadas na etapa anterior para construir uma ou mais redes que modelam o relacionamento entre essas páginas. Nesse caso, os nós da rede são os websites e os links são dados pelo número de hiperlinks de um website para outro. Ao final, obtemos uma rede ponderada e direcionada.

Nesse processo, aplicamos alguns filtros. Primeiramente, precisamos selecionar os websites que queremos considerar, ou seja, os websites que farão parte da rede. Isso significa que excluimos hiperlinks para websites que não foram selecionados. A ideia é impedir a adição de websites que não são portais de notícias e/ou não possuem viés

¹Disponível em: <https://www.kaggle.com/snapcrack/all-the-news>

previamente conhecido. Por fim, filtramos hiperlinks internos, ou seja, de um website para si mesmo. Isso impede a presença de laços na rede.

O último passo é caracterizar as redes (uma ou mais) obtidas através das distribuições de centralidades descritas no Capítulo 2. Assim, podemos verificar se as relações entre portais de notícias compõem uma rede livre de escala, e outras propriedades. Além disso, investigamos os nós mais influentes, e observamos se essa influência pode indicar alguma relação com o viés ideológico dos portais que compõem as redes.

4.1.3 Detecção de comunidades

Uma vez construídas as redes, partimos para a análise de suas características estruturais. Nossa hipótese principal é que os hiperlinks, ou seja, citações estabelecidas de um website a outro representam um alinhamento entre ambas as fontes. Portanto, consideramos que um portal possui preferência a citar fontes de mesma orientação ideológica que a sua própria.

Para investigar se essa hipótese é verdadeira, partimos primeiro para a identificação e análise de grupos formados na rede através dos links, ou seja, websites que possuem mais conexões entre si que com os demais na rede. Caso a hipótese seja verdadeira e os links estejam relacionados ao viés ideológico, esses grupos ou comunidades da rede devem ser na verdade agrupamentos de websites ideologicamente alinhados. Para identificar esses grupos, aplicamos algoritmos de detecção de comunidades. A expectativa é que sejam identificados grupos onde um certo viés ideológico seja predominante, ou seja, compostos por portais de notícias com o mesmo viés político.

4.1.4 Padrões de citações

A segunda maneira que escolhemos para avaliar a hipótese é fazer uso dos padrões de citações de cada website. Se as citações, ou hiperlinks, são estabelecidos de acordo com o viés ideológico, podemos usá-las para caracterizar esse viés. Por exemplo, uma fonte mais à direita pode preferir citar outras fontes de direita para embasar uma notícia, mantendo assim consistência com seu público e para compartilhar da credibilidade ou popularidade da fonte citada.

Para realizar essa tarefa, construiremos distribuições que modelam as citações de um website para outros de cada viés ideológico. A seguir, descrevemos de que modo essas distribuições são construídas.

Construindo funções de probabilidade

Para cada portal de notícias/website w , temos uma coleção de artigos de notícias. Digamos que temos N classes de viés ideológico². Para cada website w , computaremos uma função massa de probabilidade (fmp), ou apenas função de probabilidade, onde cada valor representa a probabilidade (frequência normalizada) de w citar websites de cada uma das N classes de viés ideológico.

Formalmente, dado um website w , uma coleção D de documentos coletados de w e um conjunto $B = \{b_1, b_2, \dots, b_N\}$ onde cada b é uma classe de viés ideológico. Para cada $d \in D, b \in B$,

$$F_b^{(w)} = \sum_{d \in D, b \in B} f_{b,d}, \quad (4.1)$$

onde cada $f_{b,d}$ representa o número de citações/hiperlinks de w para websites da classe b no documento d . Portanto, $F_b^{(w)}$ representa o número de citações estabelecidas por w para websites da classe b . Assim, cada website w é representado por uma função de probabilidade (fmp) na forma $p^{(w)} = (p_1^{(w)}, p_2^{(w)}, \dots, p_N^{(w)})$, onde

$$p_b^{(w)} = \frac{F_b^{(w)}}{\sum_{b' \in B} F_{b'}^{(w)}} \quad (4.2)$$

e, conseqüentemente, $\sum_{b \in B} p_b^{(w)} = 1$. Essa função, portanto, representa a probabilidade de w citar fontes de cada classe b de viés ideológico.

De forma análoga, construiremos funções de probabilidade que representam cada classe cada $b \in B$ de viés ideológico, mas consideraremos agora documentos pertencentes a uma classe, e não mais a um único portal.

Dada uma classe b com corpus D_b de documentos coletados de websites da classe b , para cada $i \in B$ e $d \in D_b$,

$$F_i^{(b)} = \sum_{d \in D_b, i \in B} f_{i,d}, \quad (4.3)$$

onde $f_{i,d}$ é o número de citações a websites da classe i em d . Assim, a classe b é

²Por exemplo, em nossos experimentos, as classes de viés foram esquerda, centro-esquerda, centro-direita e direita. Portanto, nesse caso, $N = 4$.

representada pela função de probabilidade $p^{(b)} = (p_1^{(b)}, p_2^{(b)}, \dots, p_N^{(b)})$ onde

$$p_i^{(b)} = \frac{F_i^{(b)}}{\sum_{i' \in B} F_{i'}^{(b)}} \quad (4.4)$$

e, conseqüentemente, $\sum_{i \in B} p_i^{(b)} = 1$. Portanto, essa função modela a probabilidade de cada classe b citar fontes de cada uma das demais ideologias, incluindo a própria b .

É importante destacar que os documentos modelados pelas equações (4.1) e (4.2) não são considerados ao computar as funções dadas pelas equações (4.3) e (4.4). Portanto, para cada website, computamos funções de probabilidade para as classes excluindo esse website, ou seja, baseadas nas citações de todos os demais portais coletados.

Computando dissimilaridades

Após computar as funções de probabilidade para cada portal de notícias, calculamos uma matriz de dissimilaridade, que modela o quão diferentes são os padrões de citações de cada website e cada classe de viés ideológico.

Para cada website w e classe de viés b , computamos $D(w||b)$, onde $D(\cdot||\cdot)$ é uma divergência, ou seja, dada um espaço de distribuições de probabilidade S , $D(\cdot||\cdot) : S \times S \rightarrow \mathbf{R}$ é uma função onde:

- $D(p||q) \geq 0$, para cada $p, q \in S$ e
- $D(p||q) = 0$ se, e somente se, $p = q$.

O objetivo de $D(\cdot||\cdot)$ é medir a diferença entre duas funções de probabilidade. Neste trabalho, decidimos empregar as divergências descritas no Capítulo 2.

Dessa forma, a matriz de dissimilaridade é computada e dada como entrada a um classificador, que fará uso desses atributos para assinalar o viés ideológico de cada website.

4.1.5 Avaliando a classificação

Após aplicar as duas estratégias, avaliaremos qual a influência que cada característica estrutural possui para identificar o viés ideológico de portais de notícias, além de verificar se nossa hipótese é verdadeira. Isto é, se existe associação entre características estruturais das redes e a ideologia dos websites.

No caso da detecção de comunidades, calculamos a quantidade de websites de cada ideologia em cada comunidade. A expectativa é que exista uma ideologia predominante, caso as comunidades reflitam esse aspecto. No caso dos padrões de citações, utilizaremos métricas de avaliação como acurácia, precisão, revocação e F1 por classe, além da matriz de confusão, para avaliar o desempenho obtido pelo classificador. Aqui, o esperado é que, caso os links estejam associados ao viés ideológico, os websites tenham padrões de citações mais parecidos com os de fontes da mesma ideologia.

4.2 Resultados obtidos

Nesta seção, descrevemos os resultados obtidos na implementação do método descrito na Seção 4.2. Esses resultados incluem: descrição do processo de coleta de dados realizado; apresentação e caracterização topológica das redes resultantes; listagem dos resultados obtidos ao aplicar detecção de comunidades às redes; por fim, descrição de uma classificação baseada em padrões de citações.

4.2.1 Coleta de dados

Como detalhado na Seção 4.1, para modelar os relacionamentos entre websites, precisamos de dois tipos de dados: o viés ideológico de websites previamente determinado por uma fonte externa, isto é, manualmente classificados segundo a opinião de especialistas; e páginas web/artigos de notícias publicados por esses websites. Para obter esses dados, utilizamos o Media Bias Fact Check (2019), um website de checagem de fatos (*fact checking*) que classifica portais de notícias com relação a viés ideológico e credibilidade. MBFC define cinco níveis, ou classes, de viés ideológico: esquerda, centro-esquerda, centro, centro-direita e direita. Essa fonte foi escolhida em lugar de outras, como o AllSides (2020), por sua classificação ser mais generalista. AllSides possui classificações diferentes para as colunas de notícias e opinião de certos veículos, por exemplo, enquanto a abordagem de MBFC é mais simples. Além disso, MBFC disponibiliza a classificação de um maior número de websites. A Figura 4.2 mostra um exemplo de um website classificado pelo MBFC.

O processo de obtenção de dados foi realizado como descrito na Seção 4.1. Nesse momento, decidimos não coletar websites do viés de centro. Alguns fatores influenciaram essa decisão. Primeiro, há um número menor de websites considerados não enviesados que são portais de notícias. A grande maioria dos portais de notícias possui viés. Um exemplo, a Wikipédia³ é um dos websites classificados por MBFC como

³Wikipédia - A enciclopédia livre. Disponível em: wikipedia.org

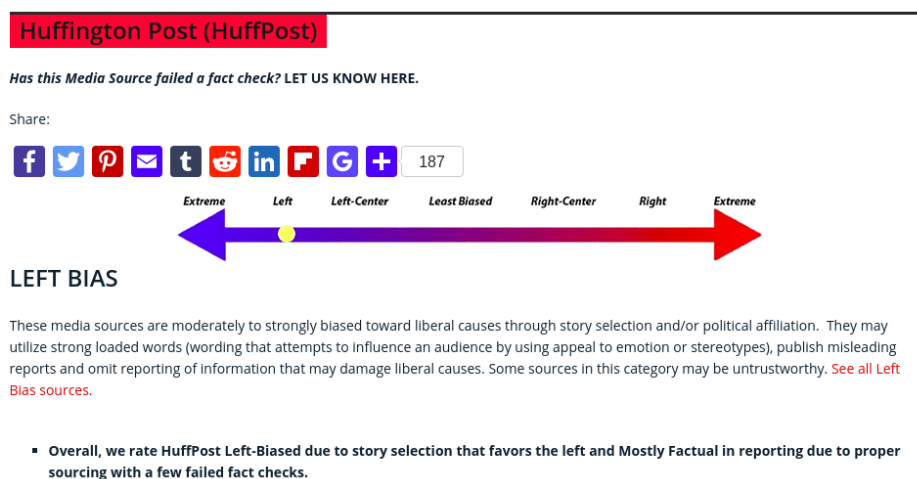


Figura 4.2: Um exemplo de um website classificado por Media Bias Fact Check (2019).

sendo da classe de centro. Segundo, uma vez que pretendemos construir redes e aplicar detecção de comunidades, e alguns dos algoritmos selecionados detectam uma partição de dois grupos, decidimos utilizá-los para encontrar um cenário polarizado, especificamente com vieses de esquerda e direita.

Dessa forma, a coleta foi realizada entre 17 e 19 de junho de 2019, retornando dados de um total de 628 websites. A Tabela 4.1 descreve o número de websites de cada classe. É importante ressaltar que os artigos coletados constituem uma coleção de documentos de tema variado, isto é, eles não são restritos a um tópico específico em comum.

Tabela 4.1: Distribuição de websites para cada classe de viés ideológico.

Viés ideológico	Nº de websites
Esquerda	133
Centro-esquerda	250
Centro-direita	127
Direita	118

4.2.2 Construção e caracterização topológica das redes

Uma vez coletadas as páginas web, partimos para a construção e caracterização das redes. As redes construídas representam diferentes conjuntos de portais, como listado na Tabela 4.2. A rede 1 modela o relacionamento de todos os portais coletados, incluindo assim portais das quatro classes de viés. Na rede 2 mantivemos nós de todas as classes de viés, no entanto, selecionamos o mesmo número de nós de cada classe, isto é, a rede modela o relacionamento de 472 portais, 118 de cada classe. Por fim,

a rede 3 modela apenas o relacionamento de portais de esquerda e direita, sendo 25 pertencentes a cada classe. É importante especificar de que modo as redes 2 e 3 foram construídas. Uma vez que os algoritmos de detecção de comunidades escolhidos exigem redes conexas, ou seja, compostas por apenas um componente conectado, geramos, de modo aleatório, diferentes redes que satisfizessem esse critério e tivessem as propriedades desejadas, como serem compostas apenas por portais de vieses mais extremos. Após, selecionamos as redes que tivessem o maior número de nós. Assim, obtivemos a composição das redes 2 e 3. Portanto, não houve escolha pré-determinada das fontes/portais de notícias que compõem as redes menores, e dessa forma, consideramos que elas são amostras da coleção de dados original.

Tabela 4.2: Redes construídas a partir dos dados coletados, representando diferentes classes de viés.

Rede	Classes	Nós	Links
1	Esquerda, centro-esquerda, centro-direita e direita	628	12001
2	Esquerda, centro-esquerda, centro-direita e direita	472	7217
3	Esquerda e direita	50	222

Após a construção destas redes, calculamos métricas de análise de redes para caracterizá-las. Para isso, selecionamos as métricas centralidade de grau ponderado, autovetor e intermediação, calculando as distribuições de cada uma destas métricas. Além disso, verificamos quais os nós mais influentes em cada rede, de forma a caracterizar os relacionamentos entre nós de diferentes ideologias. Para isso, faremos uso dos relatórios disponibilizados pelo website Media Bias Fact Check (2019) ao justificar as classificações atribuídas aos portais.

Além de métricas de centralidade, decidimos computar as seguintes métricas globais para completar a caracterização: transitividade, densidade, diâmetro e número de comunidades. A seguir, descreveremos as propriedades observadas em cada uma das redes, segundo as métricas selecionadas.

Centralidade de grau ponderado

A centralidade de grau ponderado indica os nós de maior influência local na rede com base no número de links estabelecidos entre os nós (Perra e Fortunato, 2008). Em nosso contexto, um portal possui alta influência se for muito citado por outros na rede e/ou for um grande citador. Comparando as distribuições de grau ponderado das três redes, vemos que todas parecem seguir a lei de potência, demonstrando que se tratam de redes livres de escala. Isto significa que o relacionamento entre portais de notícias

na Web segue um padrão similar a outras situações do mundo real, como redes sociais, redes que modelam processos biológicos e outras redes complexas (Barabási, 2003).

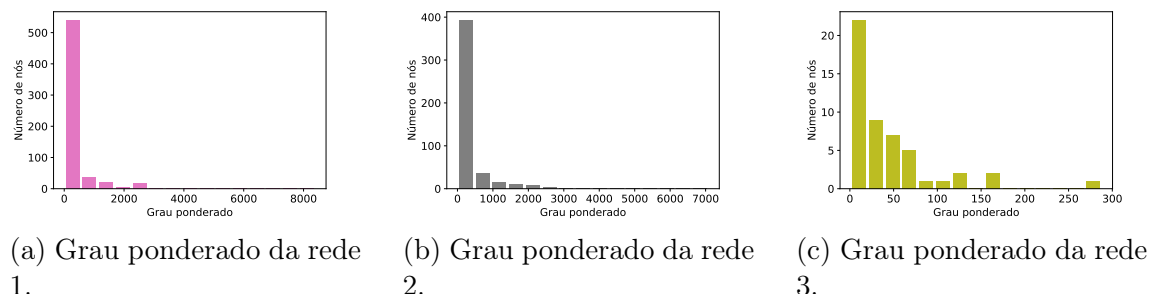


Figura 4.3: Distribuições de grau simples e ponderado.

Além de verificar as distribuições, observamos também os nós de maior destaque pelas centralidades de grau ponderado das três redes. Nas redes 1 e 2 os nós de maiores centralidades foram os mesmos, como mostra a Tabela 4.3 (Tabelas 4.3a e 4.3b). Tais destaques incluem websites de centro-esquerda (New York Times e New Moderate) e de direita (Guido Fawkes). O alto grau de entrada, isto é, citações feitas ao New York Times, provavelmente deve-se à alta credibilidade atribuída a este veículo. Por meio destas citações, outros websites podem adquirir credibilidade. Os outros dois destaques, ao contrário, são portais menores e grandes citadores na rede, com alto grau de saída. São eles New Moderate (centro-esquerda) e Guido Fawkes (direita). O primeiro referencia portais como New York Times, Washington Post, BBC e The Guardian (todos também de centro-esquerda); o segundo é um grande citador de portais de direita.

Já a rede 3 modela uma situação totalmente polarizada, incluindo apenas portais de esquerda e direita. Dentre os destaques da centralidade de grau ponderado (Tabela 4.3c) observamos a predominância dos portais CNN (esquerda) e Fox News (direita). Ambos são duas das maiores redes de notícias dos Estados Unidos, conhecidos por apresentarem uma abordagem oposta no que diz respeito ao presidente Donald Trump. Além deles, há o destaque de Daily Caller (direita). Esse website, segundo MBFC Media Bias Fact Check (2019), possui um forte viés de direita em vocabulário e seleção de histórias, com tendência a atacar portais de esquerda. Desta forma, o resultado pode indicar uma possível influência deste portal, sendo citado de modo negativo por websites de esquerda ou como uma fonte para outros portais de direita.

Tabela 4.3: Nós com maiores centralidades de grau ponderado nas rede 1, 2 e 3.

(a) Maiores centralidades de grau ponderado da rede 1.

Nó	Grau ponderado	
	Entrada	Saída
New Moderate	0	8394
Guido Fawkes	5	6514
New York Times	6226	108

(b) Maiores centralidades de grau ponderado da rede 2.

Nó	Grau ponderado	
	Entrada	Saída
New Moderate	0	7084
Guido Fawkes	5	6030
New York Times	5035	86

(c) Maiores centralidades de grau ponderado da rede 3.

Nó	Grau ponderado	
	Entrada	Saída
CNN	278	10
Fox News	162	3
Daily Caller	142	13

Centralidades de autovetor e intermediação

A centralidade de autovetor é baseada no princípio de que a importância de um nó depende da importância de seus vizinhos. Isto significa que nós mais centrais possuem mais conexões de qualidade (Newman, 2010). No contexto de nosso trabalho, um portal é mais influente se estabelece muitas citações a outros portais de alta influência. Por sua vez, a centralidade de intermediação indica o quanto um nó está incluso nos caminhos mínimos entre dois nós (Golbeck, 2015). Dessa forma, considera que um nó mais central funciona como ponte entre mais nós na rede. Assim, um portal é mais central à medida que conecta diferentes portais por meio de suas citações. As distribuições das centralidades de autovetor e intermediação, exibidas na Figura 4.4, também assemelham-se a distribuições livres de escala, mais uma vez reforçando que estas redes são de fato complexas.

Analisando os destaques segundo a centralidade de autovetor das redes 1 e 2 (Tabelas 4.4a e 4.4c), estão três portais de centro-direita. Tratam-se de portais canadenses, possuem boa credibilidade e uma tendência a favorecer Donald Trump e criticar o primeiro ministro canadense Justin Trudeau Media Bias Fact Check (2019). Esses websites destacaram-se devido a suas conexões: Financial Post e National Post também possuem alto grau ponderado de entrada na rede, explicitando que são referenciados como fontes por outros websites, em especial Regina Leader-Post. Tal preferência pode estar relacionada, mais uma vez, à credibilidade destes portais Media Bias Fact Check (2019). Do ponto de vista da centralidade de autovetor, Regina Leader-Post adquire parte da influência dos outros websites na rede através dessas citações. No caso da rede 3, os destaques, exibidos na Tabela 4.4e, além de CNN e Fox News, houve a inclusão de um portal de esquerda (Vox). Este resultado reforça o relacionamento entre Fox

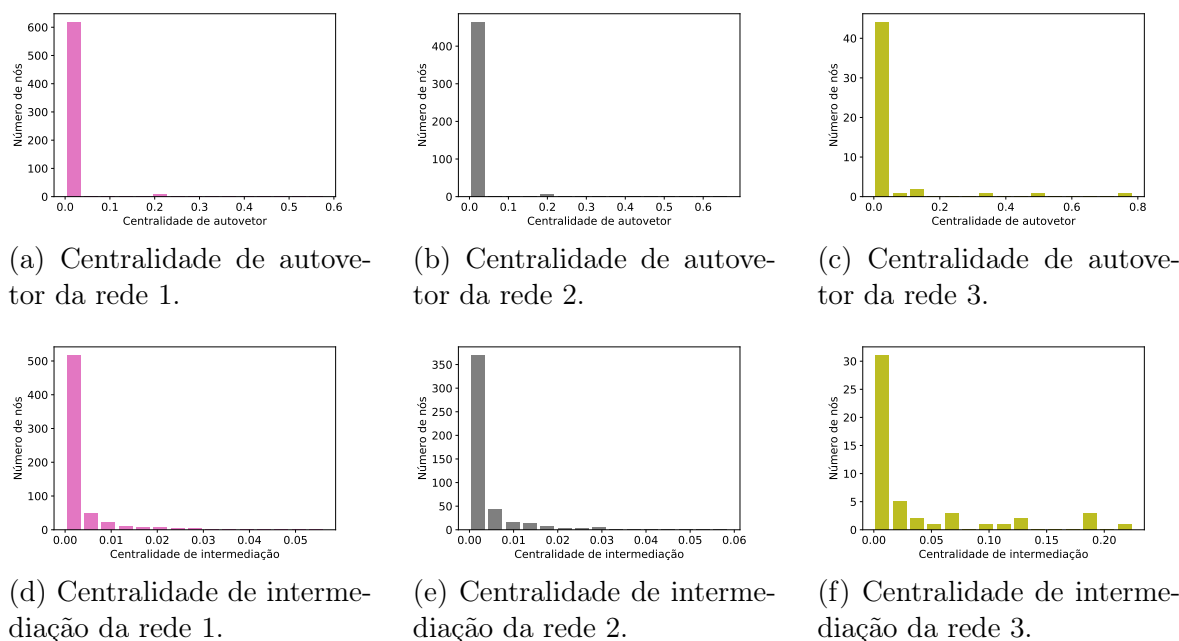


Figura 4.4: Distribuições de centralidades de autovetor e intermediação das redes 1, 2 e 3.

News e CNN, cujos links podem indicar uma postura de crítica mútua.

No caso da centralidade de intermediação, pela primeira vez as redes 1 e 2 tiveram destaques ligeiramente diferentes, com a presença de websites de esquerda e centro-esquerda, como mostram as Tabelas 4.4b e 4.4d. Em ambas os portais New York Times e Vox tiveram alta influência. Completando o ranking estão The Atlantic na rede 1 e New York Magazine na rede 2. No caso de New York Times, este resultado demonstra que o alto número de citações faz com que este website adquira alta influência local e global. Os demais, por sua vez, são portais de menor alcance, mas que estão conectados a nós de alta centralidade local, como Washington Post e o próprio New York Times, adquirindo assim destaque em um contexto global na rede. Na rede 3, os destaques (Tabela 4.4f) incluem dois portais de direita (National Review e Daily Caller) e um de esquerda (Media Matters). Observamos que estes três portais estabelecem conexões variadas na rede, tanto com portais de direita quanto de esquerda, o que explica suas influências em um contexto global.

Outras propriedades

Concluindo a caracterização, além das centralidades calculamos algumas métricas globais da rede, incluindo a transitividade, densidade, diâmetro e número de comunidades segundo o algoritmo de Louvain. A transitividade indica a tendência dos nós na rede a

Tabela 4.4: Nós com maiores centralidades de autovetor e intermediação na rede 1, 2 e 3, respectivamente.

(a) Maiores centralidades de autovetor da rede 1.

Nó	Cent. de autovetor
National Post	0,579
Financial Post	0,521
Regina Leader-Post	0,218

(b) Maiores centralidades de intermediação da rede 1.

Nó	Cent. de intermediação
The Atlantic	0,056
New York Times	0,053
Vox	0,047

(c) Maiores centralidades de autovetor da rede 2.

Nó	Cent. de autovetor
National Post	0,666
Financial Post	0,538
Regina Leader-Post	0,207

(d) Maiores centralidades de intermediação da rede 2.

Nó	Cent. de intermediação
New York Times	0,059
New York Magazine	0,056
Vox	0,05

(e) Maiores centralidades de autovetor da rede 3.

Nó	Cent. de autovetor
CNN	0,786
Fox News	0,482
Vox	0,317

(f) Maiores centralidades de intermediação da rede 3.

Nó	Cent. de intermediação
National Review	0,225
Daily Caller	0,19
Media Matters	0,184

formarem grupos (Wasserman et al., 1994). Já a densidade mostra o quão densamente conectados são os nós (Barabási, 2003). O diâmetro, por sua vez, é o comprimento do maior caminho entre dois nós em uma rede (Newman, 2010). Por fim, comunidades são grupos mais conectados entre si que aos demais na rede (Blondel et al., 2008). Redes complexas, em geral, possuem alta transitividade, baixa densidade, baixo diâmetro e estrutura de comunidades bem definida (Shen, 2013).

Tabela 4.5: Outras propriedades das redes 1, 2 e 3.

Rede	Transitividade	Densidade	Diâmetro	Comunidades
1	0,22	0,03	5	17
2	0,241	0,032	6	12
3	0,349	0,091	4	4

Alguns padrões interessantes são evidenciados pelas propriedades, listadas na Tabela 4.5. Primeiramente, à medida que selecionamos menos nós para compor as redes, aumenta a tendência destes nós se agruparem. Isso é evidenciado pela transitividade, que aumenta, e pelo número de comunidades, que diminui, indicando a formação de grupos mais conectados entre si. Além disso, observa-se que a densidade aumenta, mostrando que os conjuntos de nós que compõem cada rede possuem um forte re-

lacionamento, estabelecendo assim um número expressivo de links entre si. Isso é especialmente notável na rede formada apenas por nós de esquerda e direita. Por fim, o diâmetro da rede mostra que tratam-se de redes com a característica de *small-world* (mundo pequeno), algo próprio de redes complexas.

4.2.3 Aplicando detecção de comunidades

Uma vez coletadas as páginas web e construídas as redes, o próximo passo é aplicar algoritmos de detecção de comunidades para verificar as características dos grupos identificados em cada rede. Assim, em cada uma das redes, aplicamos os algoritmos de detecção de comunidades descritos no Capítulo 2. Em seguida, verificaremos quantos nós de cada polaridade estão presentes em cada grupo, buscando averiguar se a classificação possui alguma relação com os vieses dos indivíduos em cada comunidade. A seguir, descreveremos o resultado da execução dos cinco algoritmos escolhidos, em cada uma das redes.

Detecção de comunidades na rede completa

A primeira rede a ser analisada através dos algoritmos de detecção de comunidades é a rede 1, composta pelos 628 nós de todas as classes de viés. A Tabela 4.6 sumariza os resultados obtidos por cada um dos algoritmos. Nesta tabela, as colunas Esquerda e Direita representam o número de nós, respectivamente, das classes de esquerda e centro-esquerda; e de centro-direita e direita, em cada comunidade. Como relatado no Capítulo 2, a maioria dos algoritmos de detecção de comunidades selecionados permite a escolha do número de grupos desejados. Assim, escolhemos dois grupos, esperando que fossem identificados grupos com viés de esquerda e de direita. No caso do algoritmo de Louvain, esta propriedade não é garantida: desejamos, assim, observar de uma forma mais ampla a estrutura de comunidades da rede correspondente.

Os resultados mostram que não houve uma predominância de cada uma das classes nos dois grupos. No caso dos algoritmos de bipartição (Tabela 4.6a) e Spinglass (Tabela 4.6c), o número de nós em cada comunidade foi mais equilibrado; no entanto, em ambas as comunidades há um número alto de nós de esquerda. Já os algoritmos de Clauset, Newman e Moore (Tabela 4.6b) e Leading Eigenvector (Tabela 4.6d) tiveram uma comunidade que concentrou mais de 80% dos nós, e outra com poucos indivíduos. Ainda assim, houve predominância de nós de esquerda em ambos os grupos. No caso do algoritmo de Louvain (Tabela 4.6e), as comunidades 9 e 17, que concentram grande parte dos nós, também apresentam-se com número próximo de nós de esquerda e direita.

Tabela 4.6: Resultado da detecção de comunidades utilizando cinco algoritmos na rede 1.

(a) Algoritmo de bipartição.

	Esquerda	Direita
Comunidade 1	172	142
Comunidade 2	211	103

(c) Algoritmo Spinglass.

	Esquerda	Direita
Comunidade 1	143	57
Comunidade 2	240	188

(b) Algoritmo de Clauset, Newman e Moore.

	Esquerda	Direita
Comunidade 1	310	195
Comunidade 2	73	50

(d) Algoritmo Leading Eigenvector.

	Esquerda	Direita
Comunidade 1	361	240
Comunidade 2	22	5

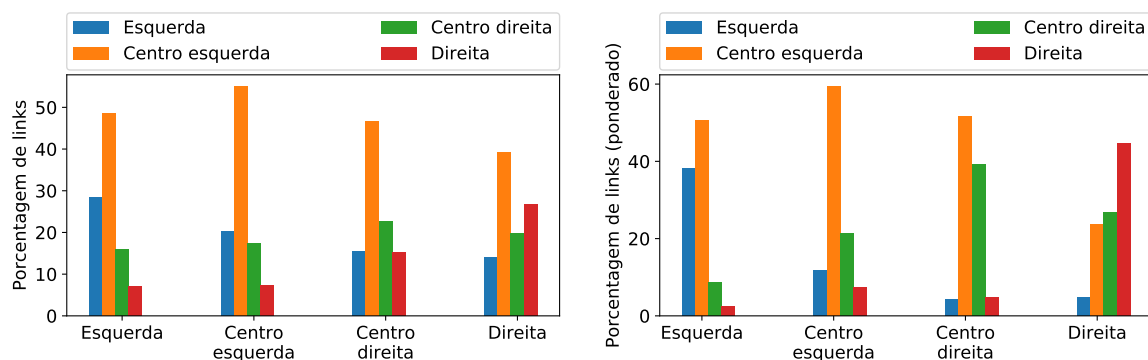
(e) Algoritmo de Louvain.

	Esquerda	Direita
Comunidade 1	2	45
Comunidade 2	3	0
Comunidade 3	1	2
Comunidade 4	2	0
Comunidade 5	0	4
Comunidade 6	6	9
Comunidade 7	15	8
Comunidade 8	5	5
Comunidade 9	67	44
Comunidade 10	4	0
Comunidade 11	45	2
Comunidade 12	2	0
Comunidade 13	2	0
Comunidade 14	20	7
Comunidade 15	11	9
Comunidade 16	0	4
Comunidade 17	198	106

Porém, as comunidades menores, como 1, 11 e 14, parecem concentrar nós de mesmo alinhamento ideológico que possuem um relacionamento mais próximo entre si.

Para entender melhor este resultado, decidimos investigar de que maneira os nós de diferentes classes se relacionam, isto é, quais os padrões de citações inter e intra-classe na rede. Na Figura 4.5, cada barra representa o número de citações, em porcentagem, que aquela classe fez à outra classe correspondente, considerando apenas o número de links na Figura 4.5a e o peso dos links na Figura 4.5b. As figuras evidenciam que os portais de centro-esquerda são mais citados que os demais, e tais citações ocorrem tanto das classes de esquerda quanto das classes de direita. Este fato explica o destaque de nós de centro-esquerda expressos pelas métricas de centralidade na caracterização apresentada anteriormente.

No caso onde o peso é considerado, as classes mais extremas tendem a citar de maneira mais alinhada. Por exemplo, a classe de esquerda cita mais portais de esquerda e centro-esquerda, enquanto a classe de direita faz mais citações a portais de direita



(a) Citações segundo o número de links.

(b) Citações segundo o peso dos links.

Figura 4.5: Relações de citações inter e intra-classes de viés na rede 1.

e centro-direita. No entanto, as classes mais centrais estabelecem mais conexões com outras classes centrais. Uma vez que os cinco algoritmos de detecção de comunidades utilizam os pesos dos links em suas abordagens, este forte relacionamento entre as classes de centro-esquerda e centro-direita parece ser a causa da composição de grupos mistos na rede.

Detecção de comunidades na rede balanceada

Após analisar uma rede composta por todos os nós de todas as classes, a segunda rede é a rede 2, composta por 472 nós, sendo 118 de cada classe. Novamente, as colunas da Tabela 4.7 agrupam as classes de esquerda e centro-esquerda; e centro-direita e direita.

Os resultados mostram algumas diferenças em relação aos obtidos na rede anterior. Os grupos descobertos pelo algoritmo de bipartição (Tabela 4.7a) possuem predominância de uma classe em cada grupo, sendo que a quantidade de nós da classe menos predominante em cada grupo foi exatamente a mesma. Este algoritmo obtém sempre dois grupos de mesmo tamanho; no entanto, é interessante que também haja a mesma quantidade de nós de cada classe, de modo invertido, nas duas comunidades. Este resultado é um primeiro indício de que há associações na rede que são relacionadas ao alinhamento ideológico entre os portais.

Algo semelhante ocorreu no algoritmo Spinglass (Tabela 4.7c). Cada um dos grupos detectados teve predominância de uma classe. Porém, neste caso a primeira comunidade reúne mais nós que a segunda. Ainda assim, este algoritmo também parece identificar associações mais próximas entre nós alinhados ideologicamente.

O mesmo não aconteceu nos algoritmos de Clauset, Newman e Moore (Tabela 4.7b) e Leading Eigenvector (Tabela 4.7d). Não houve predominância de classes nas co-

Tabela 4.7: Resultado da detecção de comunidades utilizando cinco algoritmos na rede 2.

(a) Algoritmo de bipartição.

	Esquerda	Direita
Comunidade 1	97	139
Comunidade 2	139	97

(c) Algoritmo Spinglass.

	Esquerda	Direita
Comunidade 1	208	126
Comunidade 2	28	110

(b) Algoritmo de Clauset, Newman e Moore.

	Esquerda	Direita
Comunidade 1	171	172
Comunidade 2	65	64

(d) Algoritmo Leading Eigenvector.

	Esquerda	Direita
Comunidade 1	213	213
Comunidade 2	23	23

(e) Algoritmo de Louvain.

	Esquerda	Direita
Comunidade 1	0	13
Comunidade 2	2	0
Comunidade 3	0	5
Comunidade 4	3	9
Comunidade 5	140	79
Comunidade 6	20	34
Comunidade 7	3	0
Comunidade 8	32	10
Comunidade 9	2	0
Comunidade 10	11	6
Comunidade 11	0	4
Comunidade 12	23	76

munidades detectadas por estes algoritmos. Já no caso do algoritmo de Louvain (Tabela 4.7e), dentre as 12 comunidades, as comunidades 5 e 12 apresentam predominância de cada uma das classes. Além disso, em outras comunidades menores observamos novamente nós mais conectados entre si e alinhados ideologicamente, de modo a isolarem-se em pequenos grupos.

Analisando as relações entre as diferentes classes nesta rede, apresentadas na Figura 4.6, observamos um padrão diferente do demonstrado na rede 7. Aqui, as classes mais extremas (esquerda e direita) apresentam preferência por fazer citações a elas mesmas, e em segundo lugar, à classe de centro correspondente. No entanto, as classes de centro realizam citações entre si, deixando as classes mais polarizadas em segundo plano. Isto pode ter relação com a credibilidade das fontes: em nossa caracterização, vimos que portais de centro, seja esquerda ou direita, possuem boa confiabilidade. Alguns dos portais mais extremos, por sua vez, tiveram confiabilidade mista, seja por citações, seja por falharem em *fact-checks*. Dessa forma, os portais menos enviesados parecem ser os mais citados, tanto por serem mais confiáveis quanto para transferir essa credibilidade para o portal que o cita. Estas relações influenciam os grupos formados na rede, e podem ter sido a causa do desempenho confuso obtido

pelos algoritmos de Clauset, Newman e Moore e Leading Eigenvector.

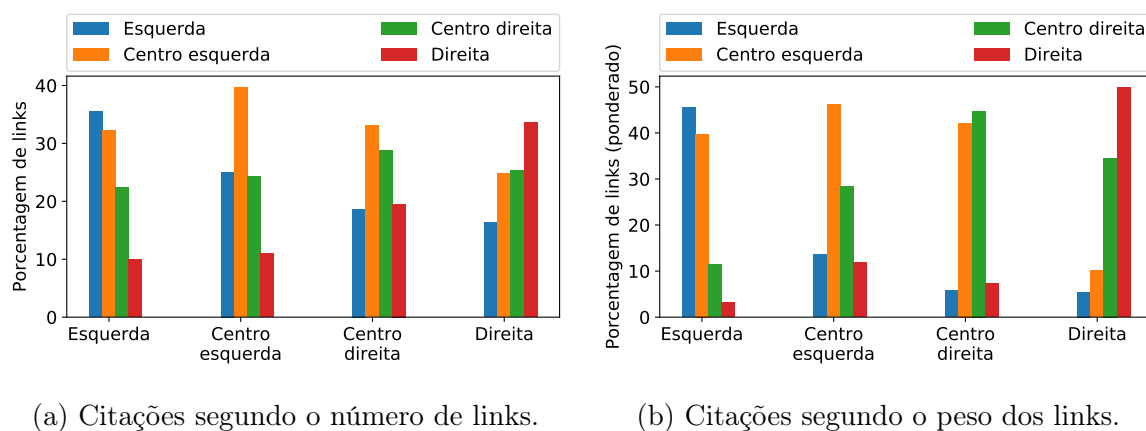


Figura 4.6: Relações de citações inter e intra-classes de viés na rede 2.

Detecção de comunidades na rede de esquerda e direita

A última rede a ser analisada é a rede 3, composta apenas por nós das classes de esquerda e direita. Neste caso, o objetivo é analisar o desempenho dos algoritmos em uma rede mais polarizada, sem portais de centro-esquerda e centro-direita. A Tabela 4.8 sumariza os resultados obtidos por cada um dos cinco algoritmos de detecção de comunidades selecionados.

Nesta rede, todos os algoritmos tiveram predominância de uma das classes em cada uma das comunidades. Assim como na rede anterior, os grupos detectados pelo algoritmo de bipartição (Tabela 4.8a) teve exatamente o mesmo número de nós predominantes de cada classe em cada comunidade. Porém, aqui o número de nós da outra classe foi muito menor, apresentando grupos que reúnem nós alinhados ideologicamente. No algoritmo Spinglass (Tabela 4.8c), houve apenas três nós da classe oposta à majoritária em cada comunidade, também com grupos onde o alinhamento ideológico parece predominar.

Diferente do que houve na rede anterior, os algoritmos de Clauset, Newman e Moore (Tabela 4.8b) e Leading Eigenvector (Tabela 4.8d) também tiveram desempenhos similares aos demais, com cada classe predominando em uma comunidade. Este último algoritmo, inclusive, teve desempenho igual ao do algoritmo de bipartição. Por fim, o algoritmo de Louvain detectou quatro comunidades, e em apenas uma delas ocorreram nós de viés oposto ao da maioria. Estes resultados demonstram que as comunidades nesta rede são mais relacionadas ao alinhamento ideológico entre os portais modelados.

Tabela 4.8: Resultado da detecção de comunidades utilizando cinco algoritmos na rede 3.

	Esquerda	Direita
Comunidade 1	2	23
Comunidade 2	23	2

(a) Algoritmo de bipartição.

	Esquerda	Direita
Comunidade 1	1	20
Comunidade 2	24	5

(b) Algoritmo de Clauset, Newman e Moore.

	Esquerda	Direita
Comunidade 1	24	2
Comunidade 2	1	23

(c) Algoritmo Spinglass.

	Esquerda	Direita
Comunidade 1	3	22
Comunidade 2	22	3

(d) Algoritmo Leading Eigenvector.

	Esquerda	Direita
Comunidade 1	12	0
Comunidade 2	0	14
Comunidade 3	13	2
Comunidade 4	0	9

(e) Algoritmo de Louvain.

Este resultado é explicado pelos padrões de citações entre as classes, apresentadas na Figura 4.7. Observamos que, nesta rede, cada classe realiza um número expressivo de citações a portais de mesmo viés. Isto significa que, neste caso, as conexões entre os nós parecem estar relacionadas ao alinhamento ideológico entre as fontes. Desta forma, as comunidades identificadas pelos diferentes algoritmos refletem esta preferência, permitindo-nos caracterizar o alinhamento destes portais.

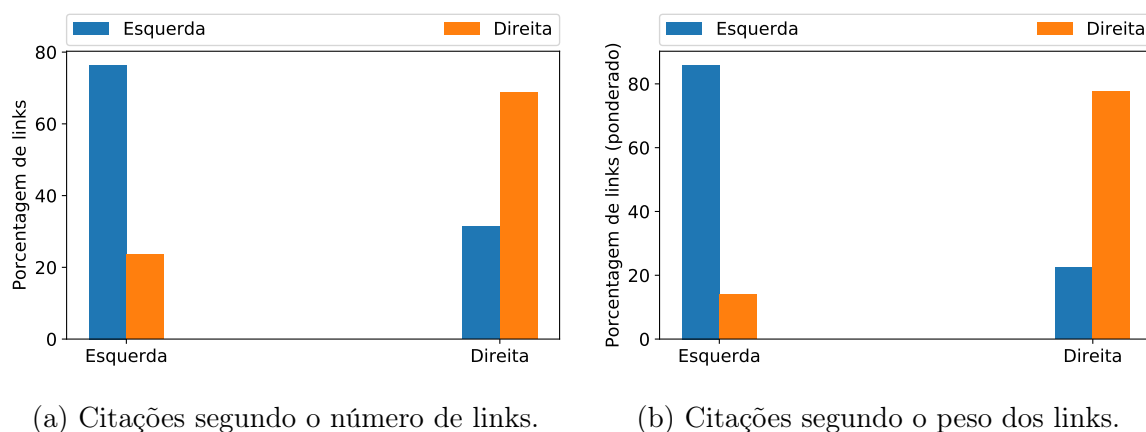


Figura 4.7: Relações de citações inter e intra-classes de viés na rede 3.

4.2.4 Classificação baseada em padrões de citações

Os resultados apresentados anteriormente mostraram a influência dos padrões de citações na composição de comunidades e a relação com viés ideológico. Diante disso,

decidimos explorar tais padrões para desenvolver atributos para um método de classificação, tal como explicado na Seção 4.1.

Assim, aplicamos o método às três redes. Bem como comentado no Capítulo 2, Seção 2.4, escolhemos duas métricas de dissimilaridade: divergência de Jensen-Shannon e distância de Jaccard, por capturarem mais nuances em distribuições de probabilidade. Realizamos experimentos com ambas as métricas, no entanto, observamos pouca diferença entre os resultados e, portanto, decidimos incluir aqui os resultados da métrica Jensen-Shannon, por ter obtido melhor desempenho. Para conferência dos valores obtidos pela distância de Jaccard, o leitor pode consultar o Apêndice A. Dessa forma, calculamos a dissimilaridade entre cada website e duas classes: esquerda (agregando as classes de esquerda e centro-esquerda) e direita (incluindo as classes de centro-direita e direita).

Com relação ao classificador, escolhemos o *Support Vector Machine* (Máquina de Vetores de Suporte) (SVM) com kernel RBF, $C = 1.0$ e demais parâmetros padrão da biblioteca scikit-learn⁴. Conduzimos os experimentos realizando validação cruzada *leave-one-out* (LOOCV) com acurácia e F1 por classe como métricas de desempenho.

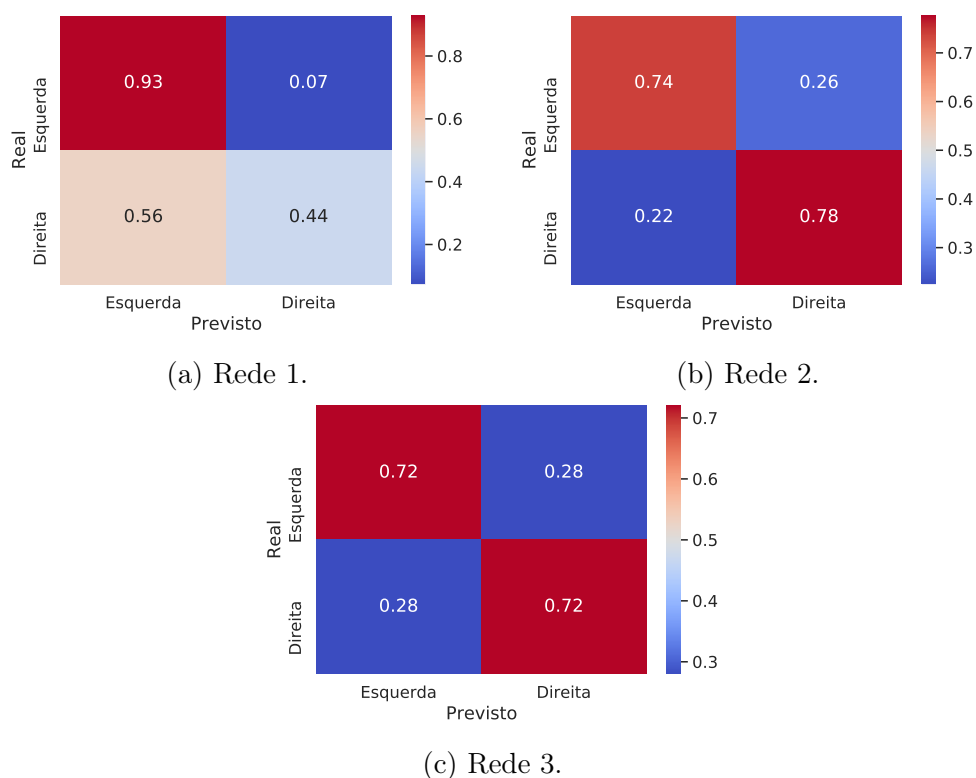


Figura 4.8: Matrizes de confusão obtidas pelo método ao classificar as redes 1, 2 e 3. Os valores estão normalizados por linhas.

⁴<https://scikit-learn.org/stable/>

Tabela 4.9: Valores de acurácia e precisão, revocação e F1 por classe para cada rede ao classificar o viés ideológico usando padrões de citações.

Redes	Precisão		Revocação		F1		Acurácia
	Esquerda	Direita	Esquerda	Direita	Esquerda	Direita	
Rede 1	0,72	0,80	0,93	0,44	0,81	0,57	74%
Rede 2	0,77	0,75	0,74	0,78	0,75	0,76	76%
Rede 3	0,72	0,72	0,72	0,72	0,72	0,72	72%

A Tabela 4.9 apresenta os resultados obtidos e a Figura 4.8 mostra as matrizes de confusão. Os resultados mostram que os valores de acurácia foram muito similares para as três redes, acima de 70%. No entanto, os valores de F1 esclarecem melhor os desempenhos nos três casos. A rede 1 é desbalanceada e possui muitos websites mais à esquerda, principalmente de centro-esquerda. Por essa razão, o modelo teve mais sucesso ao classificar portais dessa orientação, obtendo F1 de 0,81. Além disso, esse foi o caso com menor valor de F1 para a classe de direita, igual a 0,57. De fato, a Figura 4.8a mostra que o método errou a classificação de 56% dos websites de direita, ao mesmo tempo que acertou 93% dos websites de esquerda. Essa confusão provavelmente deve-se aos websites de centro-direita, que possuem um padrão de citações similar ao das fontes de centro-esquerda, como mostra a Figura 4.5b.

O método obteve melhor desempenho na rede 2, com o maior valor de acurácia (76%). Essa rede é balanceada, portanto a alta influência dos portais de esquerda foi minimizada. Nesse caso, os padrões de citações de cada classe de viés são mais distintos entre si, permitindo seu uso por um classificador para distinguir entre essas classes. Os valores de F1 agora são bem equilibrados, ou seja, ambas as classes foram igualmente classificadas. Ainda assim, como mostra a matriz de confusão (Figura 4.8b), o desempenho foi um pouco melhor ao classificar fontes de direita. Este resultado pode estar relacionado à websites de um certo viés estabelecendo links à websites da ideologia oposta para fazer uma crítica, por exemplo, criando citações que não modelam exatamente alinhamento de viés ideológico.

Este também foi o caso na rede 3. Embora esta rede possua os menores valores de acurácia e F1 por classe, percebemos que há um balanceamento entre esses valores para ambas as classes. A matriz de confusão (Figura 4.8c) também demonstra essa classificação balanceada, demonstrando que o classificador rotulou corretamente 72% de portais de cada ideologia. Isso mostra que, apesar da presença de links para websites ideologicamente opostos, o classificador foi capaz de utilizar os padrões de citações para identificar o viés político corretamente.

4.3 Considerações do capítulo

Neste capítulo, apresentamos nossa abordagem para detecção de viés ideológico em portais de notícias com base em hiperlinks. Dessa forma, descrevemos a metodologia aplicada para o desenvolvimento do método, incluindo: obtenção de dados, construção das redes, análise da detecção de comunidades e de um método baseado em padrões de citações. Em seguida, listamos os resultados obtidos após a implementação desse método. Tais resultados incluem a caracterização de propriedades das redes que modelam as relações de portais de notícias de diferentes ideologias, mostrando que essas relações podem ser representadas como redes livres de escala. Além disso, vimos que portais influentes tiveram, de fato, alta centralidade na rede. Após, descrevemos os resultados da aplicação de algoritmos de detecção de comunidades nestas redes, mostrando que a abordagem obteve desempenho mais promissor em redes onde o viés dos portais é mais polarizado. Nestas redes, o alinhamento ideológico aparece mais associado às citações entre os portais.

Utilizar padrões de citações mostrou-se como uma estratégia mais bem sucedida para caracterizar viés ideológico em qualquer caso. Nesse experimento, utilizamos padrões de citações para construir atributos que modelam a dissimilaridade entre as citações de cada website e cada classe de viés. Esses atributos foram utilizados por um classificador, que assinalou o viés de cada portal de notícias. Como resultados, tivemos acurácia acima de 72% nas três redes, com destaque para a rede que representa de forma balanceada as quatro classes de viés, onde a acurácia foi de 76%. Além da acurácia, os valores de precisão, revocação e F1 mantiveram-se acima de 0,70 na maioria dos casos, demonstrando a eficácia do método. Portanto, em nossos experimentos, as características extraídas a partir das distribuições de links nas redes foram superiores para representar o viés ideológico de portais de notícias que os grupos formados na rede.

Porém, percebemos que a abordagem baseada em hiperlinks é limitada pela presença de links de saída. Ou seja, se um portal é apenas citado, mas não cita nenhuma outra fonte (ou cita pouco), não há como fazer afirmações sobre sua orientação ideológica. Além disso, notamos que uma possível melhoria seria uma classificação mais detalhada, incluindo a identificação especificamente de extremos e moderados em adição à esquerda e direita. Diante disso, partiremos para uma abordagem que permita uma análise e experimentos mais abrangentes. No próximo capítulo, apresentaremos a metodologia e os resultados obtidos no desenvolvimento do método baseado em conteúdo textual.

Capítulo 5

Detecção de viés ideológico baseada em conteúdo textual

Neste capítulo, apresentaremos outra estratégia para detecção de viés ideológico em portais de notícias, com base no conteúdo textual dos artigos. No desenvolvimento deste método, a hipótese é que websites de mesmo viés ideológico possuem discursos semelhantes, isto é, discursos alinhados ideologicamente. Assim, caso a hipótese seja verdadeira, podemos desenvolver um método para representar o discurso de portais de notícias, capturando particularidades de cada ideologia e fazer uso destes atributos para detectar o viés político.

O capítulo está organizado do seguinte modo: na Seção 5.1 descrevemos a metodologia desenvolvida para a abordagem baseada em conteúdo textual. Na Seção 5.2 apresentamos os resultados obtidos após a implementação do método, incluindo as bases de dados coletadas e o desempenho observado ao realizar três tarefas de classificação. Por fim, a Seção 5.3 contém as considerações deste capítulo.

5.1 Método baseado em conteúdo textual

A abordagem baseada em hiperlinks, apresentada anteriormente, não leva em consideração a opinião expressa nos artigos de notícias. Um portal de esquerda pode fazer uma citação a um portal de direita, por exemplo, de modo negativo ou crítico, discordando da opinião do artigo citado. Estas nuances não são capturadas apenas pela presença de hiperlinks entre portais. Diante disso, nossa segunda abordagem é baseada no conteúdo textual dos artigos de notícias, mais especificamente, no quão diferentes são os discursos ou vocabulários de portais de notícias de ideologias distintas. Mais especificamente, nossa estratégia baseia-se em conceitos de Teoria da Informação como

a de Shannon e divergência estatística para caracterizar os discursos de portais de notícias de diferentes orientações ideológicas. A Figura 5.1 ilustra em detalhes as etapas dessa abordagem, que serão descritas a seguir.

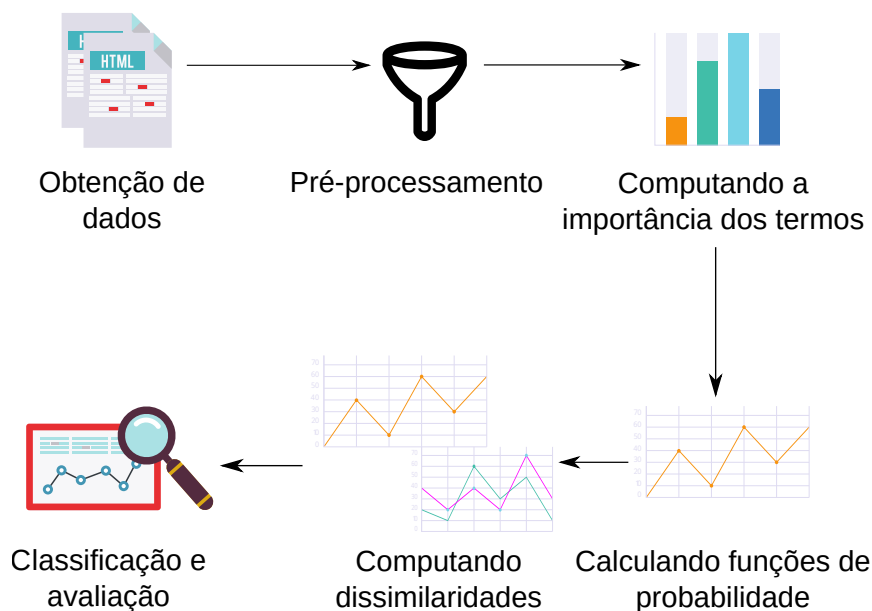


Figura 5.1: Visão geral da abordagem baseada em conteúdo.

5.1.1 Obtenção de dados

De forma análoga ao método anterior, o primeiro passo é obter artigos de portais de notícias cujo viés ideológico deve ser assinalado previamente por uma fonte manual. Para mais detalhes sobre esse processo, o leitor pode consultar a Seção 4.1.1 do Capítulo 4.

Portanto, após escolher uma amostra de fontes devidamente rotuladas, construímos uma coleção de documentos que representam o discurso de websites de diferentes vieses (esquerda/direita, extremos/moderados), contendo a seguinte estrutura:

- O nome (ou domínio) do portal de notícias.
- Um conjunto de páginas web (documentos HTML) que foram coletadas desse portal de notícias específico.
- A orientação ideológica desse portal de notícias, atribuída por outra fonte, que realize essa classificação de modo manual.

5.1.2 Preparando os dados

Como descrito anteriormente, nesse momento temos uma coleção de artigos de notícias provenientes de portais de notícias. Iniciamos o processo ao aplicar um pré-processamento nos dados. Estamos interessados em todo o texto das notícias, que pode ser extraído do HTML das páginas web. Consideraremos tanto título quanto o conteúdo que compõe o artigo. Em outros contextos, como no caso de notícias falsas, o título é o principal componente de interesse, uma vez que são inseridos elementos para chamar a atenção do leitor. Mas em nosso caso, para caracterizar o viés, acreditamos que uma representação mais ampla é capaz de levar a uma caracterização mais detalhada.

Para tratar esse texto extraído das páginas web, tratamos a capitalização, isto é, colocamos todas as letras em minúsculo, removemos números, caracteres especiais, pontuações e palavras que não fazem parte da língua inglesa. Isso é necessário para tratar possíveis ruídos, como *tags* HTML que porventura tenham sido extraídas do texto. Nessa etapa, decidimos manter as *stopwords*. Uma vez que nosso método é baseado em computar a importância dos termos, caso as *stopwords* não sejam relevantes, serão desconsideradas automaticamente.

5.1.3 Computando a importância dos termos

Para mensurar a importância de termos no vocabulário, calcularemos a entropia de Shannon, um quantificador de Teoria da Informação que mede a quantidade de informação associada a uma variável (ou aleatoriedade, de uma perspectiva estatística). Descrevemos previamente esta métrica no Capítulo 2.

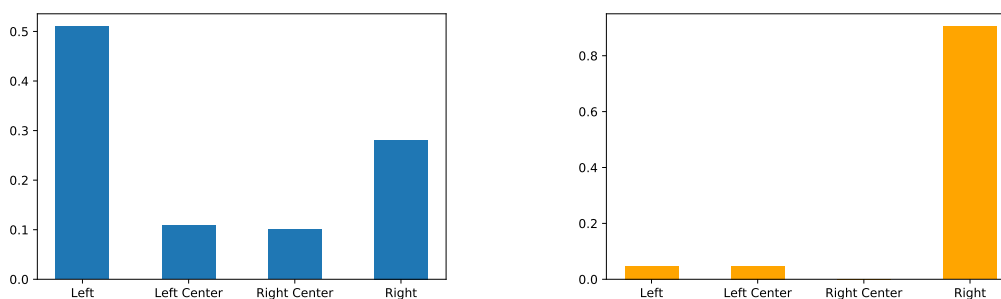
Em nosso contexto, utilizamos a entropia de Shannon para quantificar o quão útil um termo é para distinguir duas ou mais classes de viés através dos seguintes passos:

1. Computar a frequência de todos os termos em nosso corpus de referência e descartar aqueles de menor frequência, uma vez que tais termos possam ser ruído, gerando assim nosso vocabulário V .
2. Dado um problema de N classes de viés, para cada termo $t \in V$, computar $p^{(t)} = (p_1^{(t)}, p_2^{(t)}, \dots, p_N^{(t)})$, constituindo assim uma função de probabilidade do termo t sobre o espaço amostral das nossas classes de viés.
3. Para cada termo $t \in V$, computar $H(p^{(t)})$, representando a importância ou relevância do termo para distinguir entre as classes de viés.

4. Selecionar um subconjunto $V_R \subseteq V$, chamado vocabulário de referência, com m termos mais relevantes, com base nos valores $H(p^{(t)})$ computados no passo anterior. A estratégia mais ingênua seria selecionar os m termos de menor entropia. Neste trabalho, experimentamos com essa e outras estratégias, especificadas em detalhes no Capítulo 5.

Para esclarecer e ilustrar de que forma usamos a entropia neste trabalho, tomemos dois exemplos extremos. Um termo que é citado igualmente entre as classes de viés em nosso problema terá entropia máxima $\log N$, sendo assim inadequado para distinguir entre essas classes (ou seja, possui ocorrência aleatória). No outro extremo, um termo que é citado apenas por uma única classe terá entropia igual a zero (mínima), sendo assim útil para identificar aquela classe (assumindo que nosso corpus descreve perfeitamente a realidade).

A Figura 5.2 mostra um exemplo comparando a entropia dos termos *trump* e *soros* em uma das bases de dados em nossos experimentos. Nesse exemplo, vemos que *trump* é citado de forma mais uniforme que *soros*, cujas citações vêm de websites da classe direita (ou *right*). Assim, o termo *soros* deve ser mais útil que *trump* na tarefa de distinguir entre as quatro classes que estão representadas.



(a) Citação do termo *trump*, referência a Donald Trump. Esse termo teve entropia de 1,69.

(b) Citação do termo *soros*, referência a George Soros. Esse termo teve entropia de 0,55.

Figura 5.2: Exemplo da citação de dois termos por fontes de quatro classes de viés na base de dados Notícias-Julho (descrita no Capítulo 5). Os valores estão normalizados.

5.1.4 Calculando funções de probabilidade

A partir desta etapa, seguimos uma estratégia similar à aplicada na abordagem baseada em hiperlinks. Naquele caso, desenvolvemos uma estratégia para representar padrões de citações como funções de probabilidade, denotando assim a preferência de um portal

de notícias estabelecer hiperlinks/citar cada classe de viés. Aqui, seguimos uma ideia similar, porém o intuito é representar a probabilidade de um portal de notícias utilizar cada termo do vocabulário de referência V_R . Assim, teremos uma caracterização do uso dos termos por websites de cada classe de viés. A hipótese é que websites com a mesma orientação política utilizam um vocabulário mais similar a outros websites com a mesma ideologia. Se essa hipótese é verdadeira, podemos utilizar essa representação para caracterizar e identificar o discurso de cada classe de viés.

Para realizar essa tarefa, assim como na abordagem anterior, precisamos criar funções de probabilidade para representar portais de notícias e classes de viés. Essa modelagem é similar à descrita na abordagem de hiperlinks, no entanto, as equações a seguir estão adaptadas para o caso da modelagem de todos os termos $t \in V_R$.

Para representar um portal de notícias/website w , construímos uma função de probabilidade onde cada termo $t \in V_R$ é mapeado para uma variável da função, representando a probabilidade (frequência normalizada) de t ocorrer em um artigo publicado em w .

Digamos que $|V_R| = n$. Dado um website w com um corpus D de documentos/artigos coletados de w , para cada $t \in V_R$ e $d \in D$,

$$F_t^{(w)} = \sum_{d \in D, t \in d} f_{t,d}, \quad (5.1)$$

onde $f_{t,d}$ é o número de ocorrências (frequência) de t em d . Assim, o portal w é representado pela função $p^{(w)} = (p_1^{(w)}, p_2^{(w)}, \dots, p_n^{(w)})$, onde

$$p_t^{(w)} = \frac{F_t^{(w)}}{\sum_{t' \in V_R} F_{t'}^{(w)}} \quad (5.2)$$

e, conseqüentemente, $\sum_{t \in V_R} p_t^{(w)} = 1$.

Para a modelagem das classes de viés representadas por $B = \{b_1, b_2, \dots, b_N\}$, onde cada $b \in B$ é uma classe de viés (por exemplo, esquerda e direita). Cada b será representada por uma função computada de forma análoga à modelagem de cada portal de notícias, porém, em lugar de documentos de um único portal alvo, consideraremos documentos dos websites de uma classe alvo.

Digamos que $|V_R| = n$. Dada uma classe b com um corpus D_b de documentos/artigos coletados de portais de notícias na classe b , para cada $t \in V_R$ e $d \in D_b$,

$$F_t^{(b)} = \sum_{d \in D_b, t \in d} f_{t,d}, \quad (5.3)$$

onde $f_{t,d}$ é o número de ocorrências (frequência) de t em d . Assim, a classe b é representada pela função $p^{(b)} = (p_1^{(b)}, p_2^{(b)}, \dots, p_n^{(b)})$, em que

$$p_t^{(b)} = \frac{F_t^{(b)}}{\sum_{t' \in V_R} F_{t'}^{(b)}} \quad (5.4)$$

e, conseqüentemente, $\sum_{t \in V_R} p_t^{(b)} = 1$.

É importante destacar que as classes representadas pelas equações (5.3) e (5.4) incluem apenas documentos de referência. Os documentos do website w , representado pelas equações (5.1) e (5.2) não são incluídos para a computação das funções $p_t^{(b)}$ que representam as classes de viés, de modo análogo ao descrito na abordagem de hiperlinks.

5.1.5 Computando dissimilaridades entre portais de notícias e classes de viés

Uma vez computadas as funções de probabilidade para cada portal de notícias, calcularemos uma matriz de dissimilaridade, modelando o quão diferentes são os discursos de cada website e cada classe de viés. Aqui, utilizaremos divergências descritas no Capítulo 2, da mesma forma como na abordagem de hiperlinks.

Para cada website w e classe de viés b , computamos $D(w||b)$, onde $D(\cdot||\cdot)$ é uma divergência, ou seja, dado um espaço de distribuições de probabilidade S , $D(\cdot||\cdot) : S \times S \rightarrow \mathbf{R}$ é uma função onde:

- $D(p||q) \geq 0$, para cada $p, q \in S$ e
- $D(p||q) = 0$ se, e somente se, $p = q$.

O objetivo de $D(\cdot||\cdot)$ é medir a diferença entre duas funções de probabilidade.

5.1.6 Classificando e avaliando o desempenho

Após computar a matriz de dissimilaridade, representando assim as diferenças entre os discursos de cada portal de notícias e cada classe de viés, nós utilizamos essa matriz

como atributos em um classificador. A função desse classificador, portanto, será fazer uso dos padrões denotados pelos discursos para identificar o viés ideológico correto de cada website. Para avaliar o desempenho, empregaremos métricas como acurácia, precisão, revocação e F1 por classe, e a matriz de confusão. Caso a hipótese seja verdadeira e os discursos de cada website estejam relacionados ao viés ideológico, o classificador deve ser capaz de atribuir o viés de cada website corretamente.

5.2 Resultados obtidos

A seguir, descreveremos os resultados obtidos ao implementar o método descrito na seção anterior. Os experimentos e resultados estão listados da seguinte forma: descrição das bases de dados coletadas e utilizadas no processo de classificação; apresentação das diferentes configurações que compõem os experimentos; e desempenho do método ao desempenhar três tarefas de classificação.

5.2.1 Bases de dados obtidas

De modo similar ao método baseado em hiperlinks (Capítulo 4), nesta etapa é necessário obter artigos de notícias, isto é, páginas web de websites de notícias. O texto do título e conteúdo desses artigos será a principal fonte de nossa abordagem. Além disso, novamente usamos os rótulos atribuídos pelo website Media Bias Fact Check (2019), que atribui a um portal de notícias uma classificação relativa a viés ideológico e credibilidade. Dessa classificação, novamente utilizamos quatro classes de viés político: esquerda, centro-esquerda, centro-direita e direita.

Neste processo, consideramos que seria interessante analisar o desempenho do método em diferentes bases de dados. Essas bases deveriam ser compostas de artigos referentes a diferentes períodos temporais e não devem ser restritas a um tópico específico, para avaliar os resultados obtidos em diferentes cenários.

Dessa forma, selecionamos as bases listadas na Tabela 5.1. As bases de dados *Noticias-Julho* e *Noticias-Fevereiro* foram obtidas através do método de coleta de dados descrito na Seção 5.1. Além delas, coletamos uma versão reduzida da base de dados *News-Aggregator* (Dua e Graff, 2019), uma vez que parte dos artigos referenciados nesta base já não estão mais disponíveis. Na composição destas três bases de dados, buscamos balancear a quantidade de websites por classe de viés e o número de artigos por website. Para isso, em cada base, filtramos websites com poucos artigos em nossa amostragem e, dentre os restantes, contamos quantos pertenciam a cada classe. A menor quantidade

de artigos n em uma classe foi tomada como parâmetro para selecionar aleatoriamente n artigos de cada classe para compor a representação final daquela base.

Tabela 5.1: Bases de dados utilizadas no método baseado em texto.

Base de dados	Período das notícias	Nº de websites	Nº de artigos por website	Total de artigos
Noticias-Julho	17-19 de junho de 2019	248	20	4960
Noticias-Fevereiro	14-15 de fevereiro de 2020	576	20	11520
News-Aggregator	10 de março - 10 de agosto de 2014	120	15	1800

5.2.2 Metodologia experimental

Durante o desenvolvimento do método, pensamos em diferentes maneiras de realizar os experimentos aplicando o classificador às três bases de dados. Essas variações dizem respeito a aspectos como a modelagem do vocabulário, a estratégia para seleção de termos e métricas de dissimilaridade. A seguir, serão descritas as diferentes configurações que experimentamos e o que observamos em cada uma delas.

- **Modelagem do vocabulário:** dentre os possíveis modos de representar os termos dos textos, decidimos experimentar com **unigramas** e **bigramas**. A ideia era analisar se uma modelagem capaz de representar melhor o contexto possui impacto positivo no desempenho do método. De fato, observamos que a modelagem em bigramas levou a melhores resultados.
- **Estratégias para seleção dos termos:** como mencionado na Seção 5.1, precisamos construir funções de probabilidade baseadas em termos que melhor caracterizem os discursos de cada classe de viés. Para isso, computamos a entropia dos termos e definimos algumas estratégias para selecionar esses termos.
 1. Selecionar os k termos de menor entropia. Essa estratégia considera que os termos de menor entropia são os mais úteis para discriminar os discursos das diferentes ideologias. No entanto, percebemos que nesta seleção, as classes podem não ser igualmente representadas. Pode haver uma predominância de termos que caracterizem apenas uma classe, por exemplo, induzindo o modelo a não caracterizar corretamente as demais.
 2. Selecionar os k termos mais característicos de cada classe. No corpus, há termos que são citados majoritariamente por websites de uma única classe. A ideia nesta estratégia é tomar k termos exclusivos a cada classe de viés e

utilizá-los para as frequências de probabilidade. Assim, criamos uma representação mais balanceada e que caracteriza melhor o discurso dos websites de cada ideologia. Porém, nesse caso, percebemos que podem ser selecionados termos pouco citados mesmo pela classe que caracterizam, ou seja, a presença desses termos na função pode não ser tão relevante para a caracterização.

3. Selecionar termos em uma faixa de entropia. Para tratar o problema de termos pouco citados ou extremamente exclusivos, decidimos selecionar os termos que estão entre as 10% e 75% menores entropias. Esses seriam tanto termos exclusivos a uma ideologia, quanto termos citados por duas ou mais classes, que podem levar à construção de funções mais representativas. Dessa forma, a ideia seria equilibrar a capacidade de discriminação e a relevância dos termos. Mas neste caso, novamente pode haver desbalanceamento da representação do vocabulário de cada classe de viés.
 4. Combinar as estratégias 2 e 3. A última estratégia tem como objetivo criar uma função que represente igualmente o discurso das diferentes ideologias e considere também a inclusão de termos relevantes. Assim, o histograma de frequência conterá os k termos mais exclusivos de cada classe, e será complementado com os termos restantes, aqueles comuns a duas ou mais classes, mas que estão entre as 10% e 75% menores entropias.
- **Métricas de dissimilaridade:** como citado no Capítulo 2, decidimos experimentar duas métricas de divergência ou dissimilaridade: distância de Jaccard e divergência de Jensen-Shannon. Como detalhado na Seção 2.4, tais métricas foram selecionadas por serem mais sensíveis a pequenas diferenças em distribuições probabilísticas, algo que não está presente em outras métricas, como a distância do cosseno.
 - **Conjuntos de atributos:** uma vez computadas as funções de probabilidade e definida a métrica de dissimilaridade, definimos quais os atributos que irão compor a matriz de dissimilaridade. Nesse caso, pensamos em computar as dissimilaridades relativas às quatro classes de viés: esquerda, centro-esquerda, centro-direita e direita. Assim, representamos igualmente os atributos para os problemas bi-classe e multi-classe, garantindo uma melhor representação do contexto.

Deste modo, aplicamos uma combinação de cada uma dessas configurações às três bases de dados em três tarefas, buscando classificar websites:

1. Extremos e moderados.
2. De esquerda e direita.
3. De esquerda, centro-esquerda, centro-direita e direita.

O classificador utilizado foi o SVM implementado na biblioteca Scikit-Learn, com os seguintes parâmetros: kernel RBF, gamma *scale* e $C = 1.0$. Foi executada a validação cruzada seguindo a estratégia *leave-one-out* (LOOCV). As métricas de avaliação selecionadas foram precisão, revocação e F1 por classe, e acurácia.

Como baseline, utilizamos o trabalho de Baly et al. (2018), detalhado no Capítulo 3 e disponível publicamente¹. A razão para isso, como explicado anteriormente, é a similaridade entre os contextos de ambos os trabalhos: focando em classificação de portais de notícias utilizando as páginas web como fontes, não restringem os artigos a um único tópico em comum, e usam diferentes atributos de processamento de texto para a modelagem do discurso desses portais de notícias. Assim, podemos determinar como nosso método se compara com uma abordagem mais tradicional, aplicando um conjunto substancialmente menor de atributos independentes do contexto.

5.2.3 Desempenhando tarefas de classificação bi-classe e multi-classe

Na subseção anterior, apresentamos as diferentes configurações utilizadas nos experimentos. Todas foram, de fato, testadas nas três bases de dados. No entanto, como mencionado, em cada caso uma das configurações obteve melhor resultado. Deste modo, decidimos apresentar neste capítulo o resultado obtido em algumas configurações, em especial descrevendo o desempenho das quatro estratégias de seleção de termos. Nos demais aspectos, listaremos aqui os que levaram a melhores resultados: modelando o vocabulário em **bigramas**, calculando **divergência de Jensen-Shannon** e usando as dissimilaridades em relação às **quatro classes** de viés como atributos. O Apêndice B inclui resultados obtidos ao empregar unigramas, todas as estratégias e duas métricas de dissimilaridade e pode ser consultado para uma visão mais completa do desempenho do método. Esses resultados esclarecem que o desempenho da divergência de Jensen-Shannon é superior ao obtido ao calcular distância de Jaccard, além de deixar claro que, em geral, o uso de bigramas é superior aos unigramas, devido à melhor representação do contexto.

¹<https://github.com/ramybaly/News-Media-Reliability>

Aplicando as estratégias de seleção de termos, obtivemos funções de probabilidade que modelam a quantidade de bigramas indicada pela Tabela 5.2. No caso da estratégia 1, estimamos o número k de bigramas empiricamente, com base no número total de termos no vocabulário daquela base de dados. No caso da estratégia 2, o número de termos que caracteriza cada classe está balanceado, ou seja, foram selecionados 325 termos de cada classe na base *Noticias-Julho*, 854 em *Noticias-Fevereiro* e 48 em *News-Aggregator*. A estratégia 3 toma termos que estão em uma faixa de entropia, e portanto os valores variam entre as bases. A estratégia 4, combinando as duas anteriores, inclui os mesmos termos selecionados na estratégia 2 e termos que foram citados de forma equilibrada por duas ou mais classes e que estão na faixa de entropia, portanto, também variando entre as bases. A seguir, apresentaremos os resultados obtidos por esta configuração final do método ao realizar as tarefas de classificação definidas anteriormente.

Tabela 5.2: Número de bigramas selecionados para a construção dos histogramas de frequência em cada estratégia para cada base de dados.

Estratégias	Bases de dados		
	Noticias-Julho	Noticias-Fevereiro	News-Aggregator
Estratégia 1	10000	10000	5000
Estratégia 2	1300	3416	192
Estratégia 3	4982	8997	877
Estratégia 4	4427	8704	810

Classificando extremos e moderados

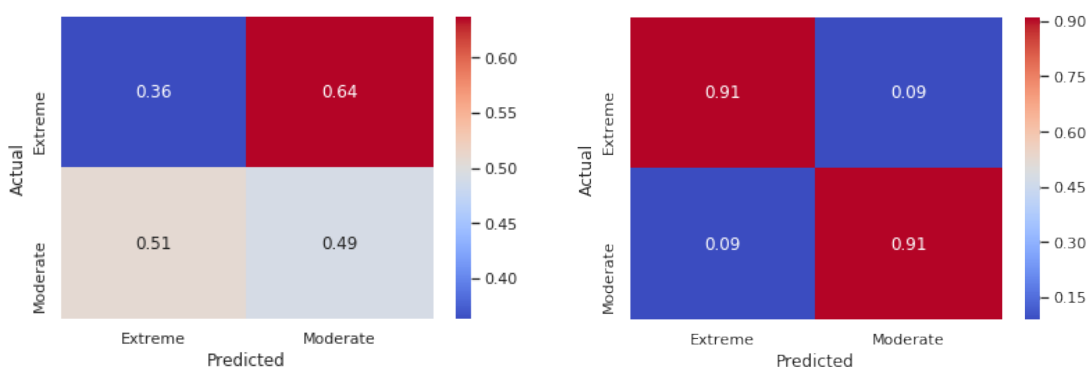
A primeira tarefa de classificação é distinguir entre portais de notícias com ideologias mais extremas (esquerda e direita) ou moderadas (centro-esquerda e centro-direita). O objetivo é verificar se há maior similaridade entre posicionamentos mais radicais e menos enviesados, respectivamente, e qual o desempenho do método neste cenário.

As Tabelas 5.3, 5.4 e 5.5 sumarizam o desempenho obtido pelo método ao empregar as quatro estratégias de seleção de termos e como se comparam ao desempenho do baseline nas três bases de dados. Aqui, destacamos de que, em todos os casos, nosso método supera o baseline, obtendo pelo menos o dobro de acurácia. Isso ocorre mesmo nosso método gerando apenas 4 atributos para representar os dados, enquanto o baseline implementa 141 diferentes atributos textuais. Este resultado pode indicar o sucesso de nossa estratégia para caracterização do texto com relação ao viés ideológico.

Analisando mais detalhadamente o desempenho em casa base de dados, na base *Noticias-Julho* (Tabela 5.3), as melhores estratégias foram 3 e 4, com destaque para a estratégia 3, que obteve maiores valores de precisão, revocação e F1 por classe, além de

Tabela 5.3: Resultados do método classificando extremos (Ex) e moderados (M) na base de dados *Noticias-Julho*. Em negrito, o melhor resultado em cada métrica.

Métodos	Precisão		Revocação		F1		Acurácia
	Ex	M	Ex	M	Ex	M	
Baseline	0,42	0,44	0,36	0,49	0,39	0,46	43%
Estratégia 1	0,87	0,84	0,84	0,88	0,86	0,86	86%
Estratégia 2	0,91	0,84	0,83	0,92	0,87	0,88	88%
Estratégia 3	0,91	0,91	0,91	0,91	0,91	0,91	91%
Estratégia 4	0,90	0,88	0,88	0,90	0,89	0,89	89%



(a) Baseline.

(b) Estratégia 3.

Figura 5.3: Matrizes de confusão obtidas pelo baseline e a melhor estratégia de seleção de termos na base *Noticias-Julho*. Os valores estão normalizados por linhas.

maior acurácia. A Figura 5.3 compara as matrizes de confusão obtidas pelo baseline e pelo nosso método, empregando a estratégia 3. Enquanto o baseline tem dificuldade em classificar corretamente portais moderados, nosso método realiza uma boa diferenciação entre ambas as classes.

Tabela 5.4: Resultados do método classificando extremos (Ex) e moderados (M) na base de dados *Noticias-Fevereiro*. Em negrito, o melhor resultado em cada métrica.

Métodos	Precisão		Revocação		F1		Acurácia
	Ex	M	Ex	M	Ex	M	
Baseline	0,43	0,43	0,44	0,42	0,43	0,42	43%
Estratégia 1	0,86	0,83	0,83	0,87	0,84	0,85	85%
Estratégia 2	0,83	0,78	0,76	0,84	0,79	0,81	80%
Estratégia 3	0,85	0,83	0,83	0,86	0,84	0,84	84%
Estratégia 4	0,86	0,82	0,82	0,86	0,84	0,84	84%

No caso da base de dados *Noticias-Fevereiro* (Tabela 5.4), o resultado foi um pouco diferente. Nesse caso, embora as estratégias 3 e 4 tenham se aproximado, o destaque foi a estratégia mais simples, separando um número fixo de termos para compor as funções de probabilidade. Acreditamos que esse desempenho está relacionado ao

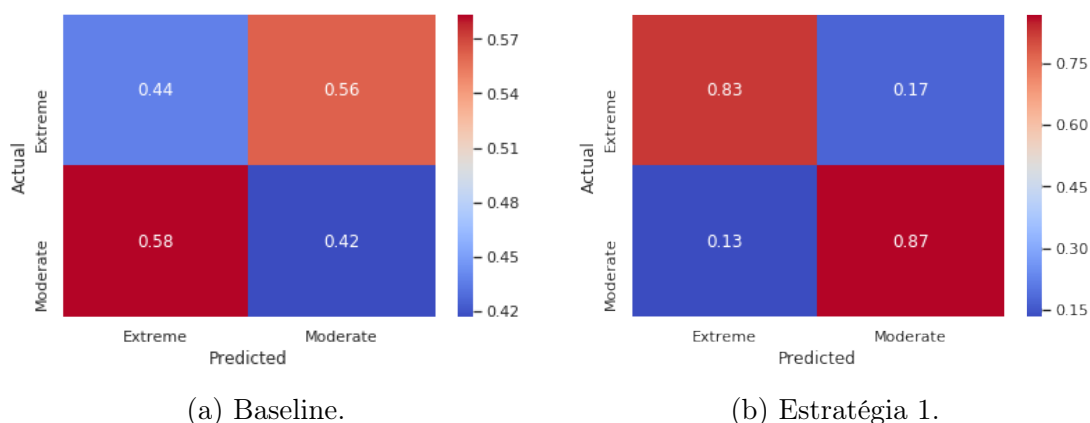


Figura 5.4: Matrizes de confusão obtidas pelo baseline e a melhor estratégia de seleção de termos na base *Noticias-Fevereiro*. Os valores estão normalizados por linhas.

fato desta ser a maior base de dados. O número de termos selecionados nas estratégias 3 e 4 já é muito próximo ao número de termos da estratégia 1. Portanto, os termos mais relevantes e característicos das classes extrema e moderada provavelmente estão presentes nas funções das três estratégias, levando a classificações muito similares. A Figura 5.4 esclarece que nesta base de dados, o baseline teve um desempenho mais equilibrado entre as classes, no entanto, ainda houve confusão. Nosso método foi mais bem sucedido em distinguir entre os portais extremos e moderados.

Tabela 5.5: Resultados do método classificando extremos (Ex) e moderados (M) na base de dados *News-Aggregator*. Em negrito, o melhor resultado em cada métrica.

Métodos	Precisão		Revocação		F1		Acurácia
	Ex	M	Ex	M	Ex	M	
Baseline	0,46	0,45	0,50	0,42	0,48	0,43	46%
Estratégia 1	0,75	0,71	0,68	0,77	0,71	0,74	73%
Estratégia 2	0,81	0,75	0,72	0,83	0,76	0,79	78%
Estratégia 3	0,88	0,77	0,73	0,90	0,80	0,83	82%
Estratégia 4	0,85	0,78	0,75	0,87	0,80	0,82	81%

Por fim, na menor base de dados, *News-Aggregator* (Tabela 5.5), ainda tivemos um desempenho bem consistente, com destaque para a estratégia 3 em precisão, revocação, F1 e acurácia. Vemos na Figura 5.5 que nosso método teve um pouco mais de dificuldade para classificar a classe de portais mais extremos. Ainda assim, em comparação ao baseline, o desempenho foi mais adequado, visto que este confundiu totalmente ambas as classes.

Portanto, os resultados da classificação de portais extremos e moderados foram muito promissores. Enquanto o baseline têm dificuldades em diferenciar as classes, todas as nossas estratégias foram melhor sucedidas nessa tarefa. Em especial, selecionar

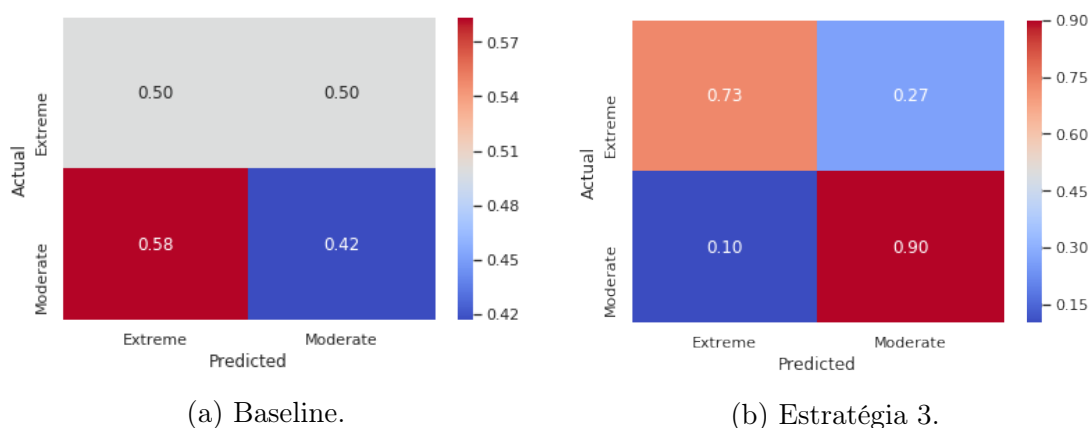


Figura 5.5: Matrizes de confusão obtidas pelo baseline e a melhor estratégia de seleção de termos na base *News-Aggregator*. Os valores estão normalizados por linhas.

termos em uma faixa de entropia levou a resultados mais satisfatórios nas três bases de dados. Esse resultado é um indicativo de que equilibrar a capacidade de caracterização e a relevância dos termos pode ser a melhor abordagem. Isso pode ser reforçado pelo fato de a estratégia 2 ter tido o pior desempenho nas três bases, o que pode mostrar que termos muito exclusivos podem não ser as melhores escolhas para a construção de funções mais representativas.

Classificando portais de esquerda e direita

A segunda tarefa de classificação é separar portais de notícias com ideologias mais à esquerda (esquerda e centro-esquerda) e mais à direita (centro-direita e direita). Nesse caso, o objetivo é medir o desempenho do método em um cenário mais polarizado, no qual é esperado que haja maior similaridade entre os discursos alinhados no espectro de viés político.

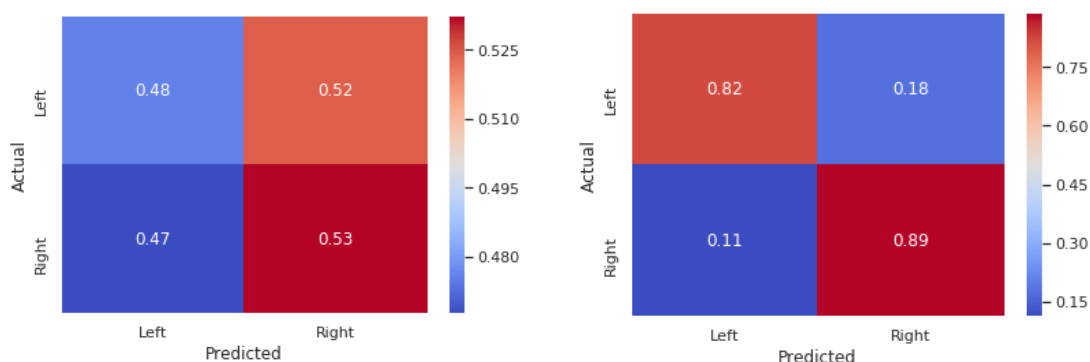
As Tabelas 5.6, 5.7 e 5.8 resumizam o desempenho do método usando diferentes estratégias comparado ao baseline nas três bases de dados. Assim como na primeira tarefa, aqui novamente evidencia-se a superioridade de nossa abordagem em relação ao baseline. Esta tarefa parece ser um pouco mais desafiadora para nosso método que a primeira, visto que os valores de acurácia foram um pouco inferiores, exceto na segunda base. Ainda assim, mantivemos um desempenho bem consistente, com a menor acurácia igual a 78%.

Fazendo uma análise por base de dados, observamos que na base *Noticias-Julho* (Tabela 5.6) a estratégia 3 superou as demais em valores de precisão, revocação e F1 por classe, além da acurácia de 85%. Essa é mais uma evidência que termos na faixa de entropia escolhida são os mais relevantes para nosso contexto. A Figura 5.6 ilustra

que, enquanto o baseline não soube distinguir entre as duas classes, nosso método teve um desempenho equilibrado, sendo especialmente bem sucedido ao classificar portais de direita. Acreditamos que essa distinção tenha ocorrido devido à alta similaridade dos discursos de centro-esquerda e centro-direita, como vimos na tarefa anterior nessa mesma base de dados.

Tabela 5.6: Resultados do método classificando websites de esquerda (E) e direita (D) na base de dados *Noticias-Julho*. Em negrito, o melhor resultado em cada métrica.

Métodos	Precisão		Revocação		F1		Acurácia
	E	D	E	D	E	D	
Baseline	0,50	0,50	0,48	0,53	0,49	0,52	50%
Estratégia 1	0,84	0,78	0,76	0,85	0,80	0,82	81%
Estratégia 2	0,79	0,72	0,69	0,82	0,74	0,77	75%
Estratégia 3	0,88	0,83	0,82	0,89	0,85	0,86	85%
Estratégia 4	0,87	0,81	0,80	0,88	0,83	0,84	84%



(a) Baseline.

(b) Estratégia 3.

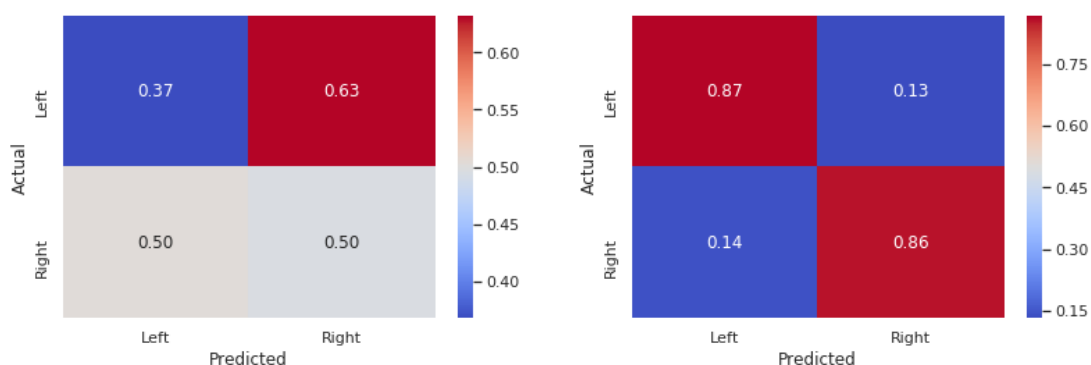
Figura 5.6: Matrizes de confusão obtidas pelo baseline e a melhor estratégia de seleção de termos na base *Noticias-Julho*. Os valores estão normalizados por linhas.

No caso da base de dados *Noticias-Fevereiro* (Tabela 5.7), a base que contém mais portais e artigos, houve um empate entre as estratégias 1, 3 e 4 em termos de acurácia (86%). Porém, o destaque maior foi da estratégia 1, com maiores valores de precisão, revocação e F1. Esse desempenho é análogo ao que observamos no caso da primeira tarefa e também pode estar relacionado ao tamanho do vocabulário desta base, que em nossa estratégia mais ingênua acaba já incluindo os termos mais relevantes. A Figura 5.7 mostra a superioridade de nosso método em relação ao baseline, que teve aqui seu pior desempenho, obtendo alta confusão em ambas as classes. Nosso método, por sua vez, foi capaz de distinguir de maneira equilibrada entre as classes.

Os resultados na menor base de dados, *News-Aggregator* (Tabela 5.8), também foram consistentes, com a melhor estratégia atingindo 78% de acurácia. Novamente,

Tabela 5.7: Resultados do método classificando websites de esquerda (E) e direita (D) na base de dados *Noticias-Fevereiro*. Em negrito, o melhor resultado em cada métrica.

Métodos	Precisão		Revocação		F1		Acurácia
	E	D	E	D	E	D	
Baseline	0,42	0,44	0,37	0,50	0,39	0,47	43%
Estratégia 1	0,86	0,87	0,87	0,86	0,86	0,86	86%
Estratégia 2	0,71	0,71	0,71	0,71	0,71	0,71	71%
Estratégia 3	0,85	0,87	0,87	0,85	0,86	0,86	86%
Estratégia 4	0,86	0,85	0,85	0,86	0,85	0,86	86%



(a) Baseline.

(b) Estratégia 1.

Figura 5.7: Matrizes de confusão obtidas pelo baseline e a melhor estratégia de seleção de termos na base *Noticias-Fevereiro*. Os valores estão normalizados por linhas.

a estratégia 3 destacou-se dentre as demais. Na Figura 5.8 observamos a diferença de desempenho entre nosso método e o baseline, que novamente teve dificuldades em distinguir as classes (embora tenha sido mais bem sucedido que nas bases anteriores), enquanto atingimos uma classificação bem equilibrada mais uma vez.

Tabela 5.8: Resultados do método classificando websites de esquerda (E) e direita (D) na base de dados *News-Aggregator*. Em negrito, o melhor resultado em cada métrica.

Métodos	Precisão		Revocação		F1		Acurácia
	E	D	E	D	E	D	
Baseline	0,56	0,56	0,55	0,57	0,55	0,56	56%
Estratégia 1	0,50	0,50	0,53	0,47	0,52	0,48	50%
Estratégia 2	0,61	0,71	0,80	0,48	0,69	0,57	64%
Estratégia 3	0,78	0,77	0,77	0,78	0,77	0,78	78%
Estratégia 4	0,75	0,75	0,75	0,75	0,75	0,75	75%

Sumarizando os resultados observados nessa tarefa, nosso método continua capaz de classificar corretamente as classes, porém com um pouco menos de facilidade que na primeira tarefa. O motivo disso é o maior alinhamento entre as classes menos enviesadas, que estão mais relacionadas entre si que com as extremas correspondentes.

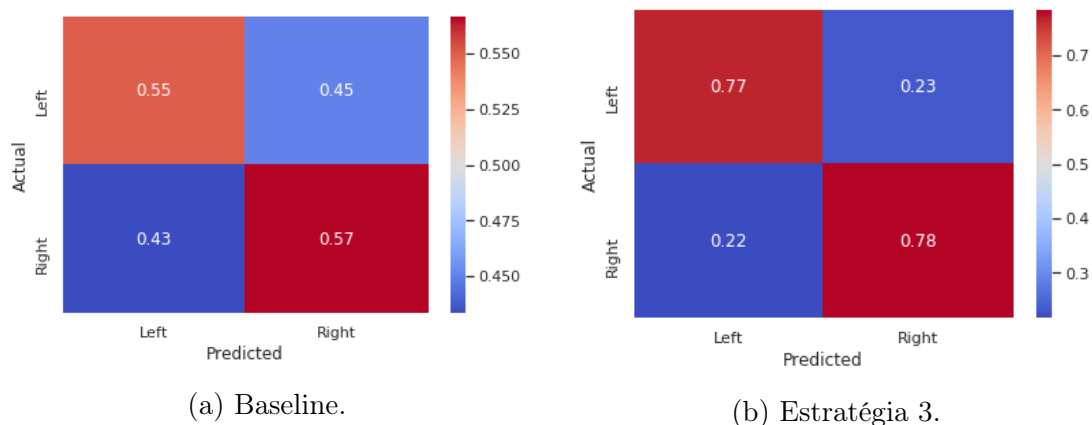


Figura 5.8: Matrizes de confusão obtidas pelo baseline e a melhor estratégia de seleção de termos na base *News-Aggregator*. Os valores estão normalizados por linhas.

Além disso, também há maior similaridade entre os discursos de ambos os extremos. Vimos que selecionar termos em uma faixa de entropia trouxe bons resultados em todos os casos, mas especialmente nas bases de dados menores. Além disso, a estratégia 2, selecionar k termos mais característicos de cada classe, parece ainda ser a menos interessante dentre nossas abordagens.

Classificando portais de esquerda, centro-esquerda, centro-direita e direita

A terceira e última tarefa de classificação é distinguir entre as quatro classes de viés (esquerda, centro-esquerda, centro-direita e direita). O objetivo é verificar o desempenho do método em um cenário multi-classe, mais desafiador que uma tarefa bi-classe, e como se compara ao baseline experimentalado.

As Tabelas 5.9, 5.10 e 5.11 descrevem o desempenho obtido pelo baseline e nossas estratégias. Nessa tarefa, o baseline teve seu pior desempenho. Nosso modelo também teve maiores dificuldades em comparação às tarefas binárias, no entanto, ainda obteve um desempenho consistente e interessante, com acurácia de até 79%.

Tabela 5.9: Resultados do método classificando websites de esquerda (E), centro-esquerda (CE), centro-direita (CD) e direita (D) na base de dados *Noticias-Julho*. Em negrito, o melhor resultado em cada métrica.

Métodos	Precisão				Revocação				F1				Acurácia
	E	CE	CD	D	E	CE	CD	D	E	CE	CD	D	
Baseline	0,29	0,26	0,34	0,25	0,26	0,31	0,34	0,23	0,27	0,28	0,34	0,24	28%
Estratégia 1	0,77	0,75	0,63	0,79	0,79	0,66	0,73	0,74	0,78	0,70	0,68	0,77	73%
Estratégia 2	0,84	0,72	0,56	0,88	0,84	0,63	0,73	0,73	0,84	0,67	0,63	0,80	73%
Estratégia 3	0,86	0,84	0,64	0,84	0,87	0,68	0,81	0,77	0,86	0,75	0,71	0,81	78%
Estratégia 4	0,82	0,84	0,68	0,88	0,85	0,68	0,84	0,79	0,83	0,75	0,75	0,83	79%

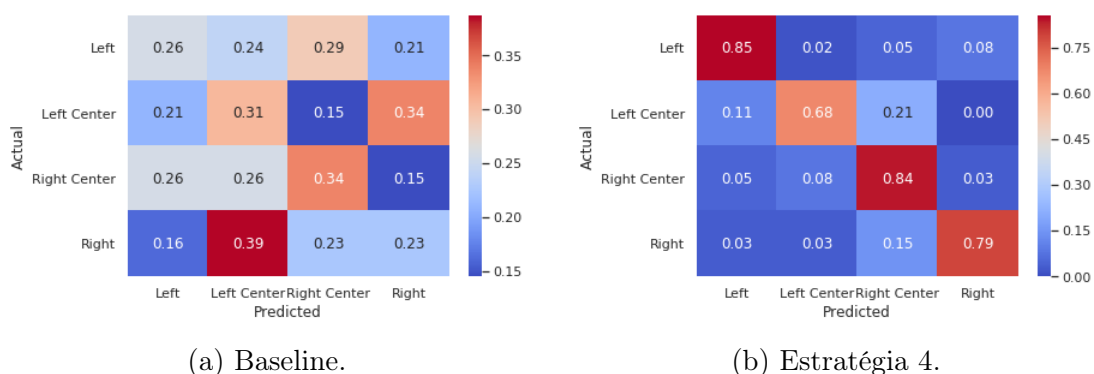


Figura 5.9: Matrizes de confusão obtidas pelo baseline e a melhor estratégia de seleção de termos na base *Noticias-Julho*. Os valores estão normalizados por linhas.

Na primeira base de dados, *Noticias-Julho* (Tabela 5.9), observamos um desempenho muito próximo das estratégias 3 e 4 (em termos de precisão, revocação e F1), com leve superioridade desta última (79% de acurácia). Comparando o desempenho do baseline e dessa estratégia, mostrados na Figura 5.9, vemos a superioridade de nosso método. Enquanto o baseline, mesmo tendo mais atributos, não é capaz de distinguir entre as classes, nosso método consegue uma classificação bem equilibrada, gerando uma diagonal de acertos na matriz de confusão.

Na segunda base de dados, *Noticias-Fevereiro* (Tabela 5.10), assim como nas tarefas anteriores, os melhores resultados foram obtidos pela estratégia 1. A Figura 5.10 novamente evidencia o bom desempenho de nosso método. Enquanto o baseline têm seu pior desempenho até aqui, incorretamente classificando a maioria das instâncias como sendo da classe de esquerda, nosso método é capaz de distinguir corretamente entre as quatro classes de viés. Este é mais um indicativo da robustez de nossa estratégia para um problema multi-classe. É interessante destacar novamente que esse resultado foi atingido com um conjunto de apenas quatro atributos contra 282 empregados pelo baseline.

Tabela 5.10: Resultados do método classificando websites de esquerda (E), centro-esquerda (CE), centro-direita (CD) e direita (D) na base de dados *Noticias-Fevereiro*. Em negrito, o melhor resultado em cada métrica.

Métodos	Precisão				Revocação				F1				Acurácia
	E	CE	CD	D	E	CE	CD	D	E	CE	CD	D	
Baseline	0,26	0,22	0,19	0,22	0,30	0,22	0,19	0,19	0,28	0,22	0,19	0,21	22%
Estratégia 1	0,82	0,73	0,69	0,80	0,78	0,73	0,77	0,75	0,80	0,73	0,73	0,77	76%
Estratégia 2	0,72	0,57	0,54	0,72	0,70	0,59	0,56	0,69	0,71	0,58	0,55	0,70	63%
Estratégia 3	0,79	0,72	0,69	0,80	0,78	0,69	0,78	0,73	0,78	0,71	0,73	0,76	75%
Estratégia 4	0,81	0,72	0,70	0,78	0,76	0,70	0,79	0,74	0,79	0,71	0,74	0,76	75%

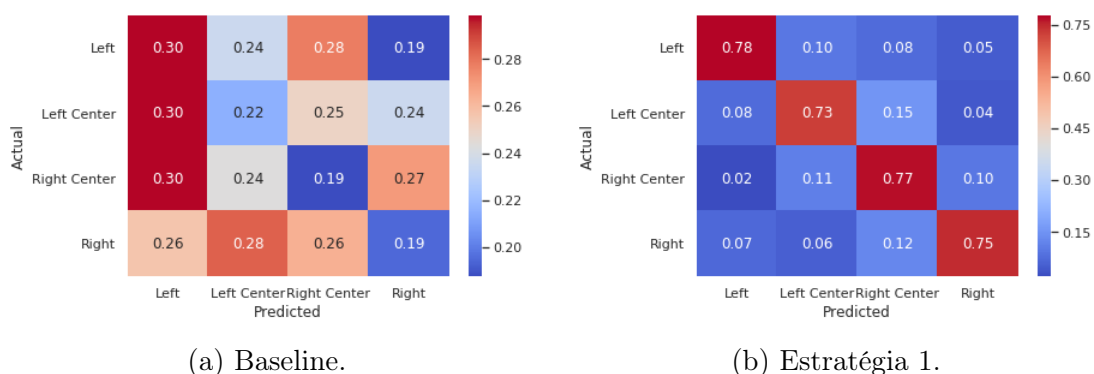


Figura 5.10: Matrizes de confusão obtidas pelo baseline e a melhor estratégia de seleção de termos na base *Noticias-Fevereiro*. Os valores estão normalizados por linhas.

Na terceira base de dados, *News-Aggregator* (Tabela 5.11), nosso método teve o resultado menos interessante, atingindo 63% de acurácia ao empregar a estratégia 3. Acreditamos que isso ocorra por se tratar da menor base de dados, e, assim, conter o menor número de exemplos para o classificador. Ainda assim, é interessante observar que, ainda mais que nos outros casos, nesta base a escolha de termos em uma faixa de entropia possui ainda mais impacto, principalmente comparado à estratégia 1 (que obteve apenas 31% de acurácia). Isso ocorre devido ao menor número disponível de termos, fazendo com que a escolha dos mais relevantes seja ainda mais importante. Comparado ao baseline, como mostra a Figura 5.11, nosso método ainda obteve uma classificação bastante equilibrada.

Tabela 5.11: Resultados do método classificando websites de esquerda (E), centro-esquerda (CE), centro-direita (CD) e direita (D) na base de dados *News-Aggregator*. Em negrito, o melhor resultado em cada métrica.

Métodos	Precisão				Revocação				F1				Acurácia
	E	CE	CD	D	E	CE	CD	D	E	CE	CD	D	
Baseline	0,27	0,36	0,29	0,26	0,30	0,40	0,30	0,20	0,29	0,38	0,30	0,23	30%
Estratégia 1	0,44	0,17	0,12	0,71	0,57	0,23	0,10	0,33	0,49	0,20	0,11	0,45	31%
Estratégia 2	0,70	0,40	0,46	0,68	0,63	0,53	0,43	0,57	0,67	0,46	0,45	0,62	54%
Estratégia 3	0,74	0,43	0,64	0,87	0,67	0,60	0,60	0,67	0,70	0,50	0,62	0,75	63%
Estratégia 4	0,75	0,39	0,61	0,90	0,70	0,50	0,67	0,63	0,72	0,44	0,63	0,75	62%

Dessa forma, observamos que além de obter alta acurácia em problemas bi-classe, nosso método, que emprega um conjunto de quatro atributos, mantêm-se estável em um problema multi-classe, atingindo valores interessantes de acurácia e muito superiores a um baseline que emprega 282 atributos. Os resultados mostram que o maior desafio para nossa abordagem é distinguir entre as classes de centro-esquerda e centro-direita, em especial nas bases *Noticias-Julho* e *News-Aggregator*. Além disso, observamos um impacto ainda maior de escolher termos em uma faixa de entropia, em especial na

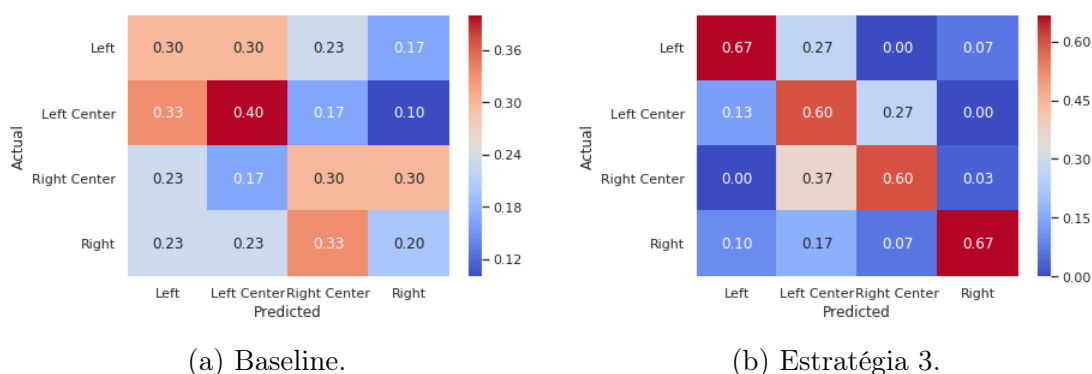


Figura 5.11: Matrizes de confusão obtidas pelo baseline e a melhor estratégia de seleção de termos na base *News-Aggregator*. Os valores estão normalizados por linhas.

menor base de dados. A abordagem menos interessante permanece a da estratégia 2, selecionar o mesmo número de termos que mais caracterizam cada classe.

5.3 Considerações do capítulo

Neste capítulo, apresentamos nossa abordagem para detecção de viés ideológico em portais de notícias com base em conteúdo textual. Iniciamos descrevemos a metodologia aplicada para o desenvolvimento do método, incluindo: obtenção de dados, pré-processamento, seleção de termos mais relevantes através da computação da entropia de Shannon, cálculo da dissimilaridade entre os discursos de cada fonte e cada classe de viés, e classificação. Após, apresentamos os resultados obtidos ao implementar esse método. Os resultados incluem a coleta de três bases de dados contendo notícias de períodos diferentes e sobre temas variados e a aplicação de um classificador a essas bases. O classificador recebeu como atributos uma matriz que mede a dissimilaridade entre os portais de notícias de quatro classes de viés político diferentes: esquerda, centro-esquerda, centro-direita e direita. Essa dissimilaridade foi medida através da divergência de Jensen-Shannon, que mediu o quão diferentes são funções de probabilidade que modelam os discursos de cada portal de notícias e cada classe de viés. Essas funções são construídas usando diferentes estratégias para selecionar os termos que as compõem e considerando o vocabulário em bigramas. Por fim, comparamos o desempenho obtido por nosso método com um baseline que emprega 282 atributos de processamento textual, contra apenas quatro atributos em nossa abordagem (a dissimilaridade com cada classe de viés ideológico).

Analisamos três tarefas de classificação: classificando extremos e moderados; esquerda e direita; e as quatro classes de viés. Em todas elas, observamos alguns padrões

diferentes para as três bases de dados. Enquanto nas bases menores a melhor estratégia foi selecionar termos em uma faixa de entropia (entre as 10% e 75% menores entropias), na base que contém mais artigos uma estratégia ingênua, selecionando os 10000 termos de menor entropia, foi suficiente. Essa diferença pode estar relacionada ao número de artigos e, assim, ao número de exemplos disponíveis para o treino do modelo.

Falando especificamente dos resultados da classificação das três tarefas, observamos que distinguir entre duas classes foi mais fácil para o classificador que uma tarefa multi-classe. Os resultados mostram que os discursos mais próximos são os das ideologias extremas (esquerda e direita) e moderadas (centro-esquerda e centro-direita), onde obtivemos uma acurácia de até 91%. Essa maior similaridade parece ter impactado na classificação de ideologias de esquerda (esquerda e centro-esquerda) e direita (centro-direita e direita), onde tivemos um desempenho um pouco inferior, mas ainda estável, atingindo acurácia de 86%. Por fim, essa maior similaridade também impacta o problema multi-classe, onde o método confundiu mais portais de ideologias de centro-esquerda e centro-direita. Ainda assim, tivemos uma classificação equilibrada, com até 79% de acurácia, um desempenho muito acima do baseline.

Diante disso, nossos resultados demonstram a eficácia de empregar conceitos de Teoria da Informação (entropia, divergências) para construir atributos mais relevantes para a tarefa de classificação de viés ideológico em portais de notícias. Com um conjunto de apenas quatro atributos, nosso método é capaz de superar um baseline que emprega 282 atributos comuns em processamento de texto e em contexto como o de *fake news*, por exemplo, atingindo aproximadamente o dobro da acurácia obtida pelo baseline em todos os cenários. Mostramos também que nosso método é acurado tanto para classificações bi-classe quanto para um problema mais desafiador, multi-classe. Isso deixa claro que o conjunto de atributos construído é realmente mais útil para caracterizar cada classe de viés que estratégias mais tradicionais.

Ainda assim, percebemos no decorrer da pesquisa que, apesar dos bons resultados, este método possui algumas limitações. Em primeiro lugar, o método é dependente do vocabulário dos artigos de notícias, e, portanto, precisa ser construído novamente para um novo conjunto de notícias. Além disso, essa dependência também é relacionada aos tópicos discutidos nos artigos da coleção e ao período dessas notícias, uma vez que os tópicos mudam com o tempo. Por exemplo, o cenário político sofre alterações constantes a cada eleição, mudando as notícias sobre esse tema. Uma possível melhoria para estas questões seria combiná-lo a métodos que avaliam a estrutura das páginas web, como o método baseado em hiperlinks proposto no Capítulo 4. No próximo capítulo, apresentamos as considerações finais dessa pesquisa, incluindo comentários sobre os resultados das abordagens de hiperlinks e conteúdo textual.

Capítulo 6

Considerações finais

Nesta dissertação, apresentamos abordagens para detecção de alinhamento ideológico de portais de notícias na Web. Descrevemos em detalhes duas estratégias: um método baseado em hiperlinks, e um método baseado em conteúdo textual. A seguir, apresentaremos um sumário do que foi feito em cada uma das abordagens e as conclusões mais interessantes observadas em cada caso (Seções 6.1 e 6.2), seguido das publicações geradas neste trabalho na Seção 6.3 e indicações de possíveis trabalhos futuros na Seção 6.4.

6.1 Detecção de viés ideológico via hiperlinks

O primeiro método fez uso dos hiperlinks presentes em páginas web de artigos de notícias para desenvolver uma forma de caracterizar portais de notícias. Realizamos a modelagem de artigos de notícias de websites de diferentes classes de viés político/ideológico através da construção de redes, onde os nós foram os portais de notícias e os links, o número de hiperlinks de um website para outro presentes nas páginas web. Construímos redes para avaliar três cenários diferentes, considerando: todos os portais presentes em nossa base de dados; uma amostra balanceada desses portais; e uma amostra contendo apenas portais de vieses extremos. Ao caracterizá-las, observamos que todas possuem propriedades de redes complexas, e as diferentes medidas de centralidade demonstraram a alta influência de portais de centro-esquerda e centro-direita, o que indica alta citação para fontes desse tipo.

Além de caracterizar estas redes, tínhamos como objetivo avaliar se o estabelecimento de hiperlinks estava relacionado ao viés ideológico, isto é, se as fontes fazem citações a portais ideologicamente alinhados a elas. Se essa hipótese fosse verdadeira, seria possível utilizar características estruturais das redes para detectar viés ideológico.

Assim, a primeira característica explorada foram as comunidades, ou grupos, presentes nas redes. Através da aplicação de algoritmos de detecção de comunidades, vimos que apenas no caso de uma rede totalmente polarizada foi possível verificar uma associação entre os grupos e o alinhamento ideológico das fontes que os compunham. Observamos que isso ocorreu devido aos portais menos enviesados preferirem citar-se entre si que às fontes mais extremas, além de outros fatores, como uma citação para adquirir credibilidade.

A segunda estratégia foi modelar os padrões de citações, ou links, dos portais de notícias como funções de probabilidade. A ideia era que portais de mesmo viés fariam citações de modo similar, preferencialmente a portais da mesma ideologia. Assim, computamos uma matriz de dissimilaridade, que media o quão diferentes são os padrões de citações de cada website em relação a cada classe de viés. Essa matriz foi dada como entrada a um classificador, que foi capaz de obter acurácias entre 72% e 76% para as três redes ao assinalar o viés dos portais como sendo de esquerda ou direita. Assim, os padrões de citações mostram-se mais úteis para esse problema que a análise de comunidades, sendo interessantes para analisar o viés ideológico de portais de notícias. Como limitações, percebemos que a estratégia direcionada pode dificultar a classificação de websites que fazem poucas citações. Além disso, uma melhoria seria experimentar uma análise mais detalhada, adicionando vieses extremos e moderados em adição à esquerda e direita.

6.2 Detecção de viés ideológico via conteúdo textual

A segunda abordagem foi o desenvolvimento de um método baseado no texto dos artigos de notícias. Nesse contexto, focamos em construir um método mais eficaz e menos complexo que outros modelos mais tradicionais. A hipótese, nesse caso, era que portais de notícias de um certo viés possuem discurso mais semelhante aos de websites desse mesmo viés.

Assim, de modo semelhante aos padrões de citações, desenvolvemos uma estratégia para modelar o texto através de funções de probabilidade. O diferencial está na escolha dos termos que compõem essas funções. Empregamos o conceito de entropia para selecionar os termos mais relevantes, ou seja, aqueles mais úteis para caracterizar os discursos de um portal de notícias. Construímos uma matriz de dissimilaridade, que mede o quão diferentes são os discursos (ou as funções de probabilidade) de cada website e cada classe de viés (portanto, construímos um conjunto de quatro atribu-

tos). Essa matriz foi dada a um classificador, que desempenhou três tarefas: distinguir entre extremos e moderados; esquerda e direita; e todas as quatro classes (esquerda, centro-esquerda, centro-direita e direita).

Conduzimos experimentos em três bases de dados, construídas de modo a representar um conjunto balanceado de portais de notícias das quatro classes de viés. Além disso, comparamos diferentes estratégias que usam a entropia para a seleção de termos. Os resultados foram comparados também aos obtidos por um baseline (Baly et al. (2018)), que emprega 282 atributos textuais mais tradicionais, aplicados por exemplo em detecção de *fake news*.

Como resultados, observamos que nosso método é capaz de superar o baseline nas três tarefas de classificação, atingindo até o dobro da acurácia deste. Quanto às estratégias de seleção de termos, a mais bem sucedida foi escolher aqueles entre as 10% e 75% menores entropias. Essa estratégia destacou-se especialmente nas bases de menor tamanho. Em seguida, a estratégia mais ingênua, selecionar k termos de menor entropia, teve o melhor desempenho na base de dados que contém mais websites e artigos. Por fim, mostramos que nosso modelo é capaz de obter um bom desempenho, distinguindo entre as classes de modo equilibrado, em especial nos problemas bi-classe. Ao classificar extremos e moderados, tivemos até 91% de acurácia e 0,91 de F1, contra acurácia de 43% e F1 de 0,46 obtidos pelo baseline no mesmo caso. Classificando portais de esquerda e direita, obtivemos até 86% de acurácia e 0,87 de F1, contra acurácia de 43% e F1 de 0,47 obtidos pelo baseline no mesmo caso. No entanto, é no problema multi-classe que nosso método realmente se destaca, atingindo acurácia de até 79% e F1 de 0,83, enquanto o baseline obteve apenas 28% de acurácia e até 0,34 de F1 nesse mesmo contexto. Assim, ficou claro a robustez de nosso modelo, tanto em problemas bi-classe quanto multi-classe. A principal limitação do método é a dependência do vocabulário e período, uma vez que tópicos de artigos de notícias evoluem com o tempo. Portanto, uma melhoria para tornar o método ainda mais adequado seria combiná-lo a abordagens que sejam independentes de vocabulário, como a estratégia baseada em hiperlinks.

6.3 Publicações

Este trabalho gerou as seguintes publicações:

- *A Link-based Approach to Detect Media Bias in News Websites* no workshop LA-Web (*Latin America Web*), realizado em conjunto com a conferência *The Web Conference* de 2019 em San Francisco, California (Aires et al., 2019). Neste

artigo, descrevemos a abordagem baseada em hiperlinks e alguns resultados preliminares.

- *An Information Theory Approach to Detect Media Bias in News Websites*, artigo aceito e, brevemente, publicado no workshop WISDOM (*Workshop on Issues of Sentiment Discovery and Opinion Mining*), parte dos workshops da conferência SIGKDD (*Conference on Knowledge Discovery and Data Mining*), que será realizado de modo virtual em agosto de 2020. Neste artigo, descrevemos experimentos da abordagem baseada em conteúdo textual, tratando especificamente da estratégia de selecionar k termos de menores entropias.

6.4 Trabalhos futuros

Como trabalhos futuros, acreditamos que existem algumas direções interessantes. No caso da abordagem baseada em hiperlinks, fazer experimentos complementares considerando o caso da rede não-direcionada, uma vez que tratamos a direção das citações ao modelar seus padrões. No caso da abordagem baseada em conteúdo, possíveis extensões podem ser feitas ao experimentar outras métricas de divergências, além de um conjunto mais amplo de artigos de notícias. Além disso, uma última extensão seria a análise do desempenho obtido ao combinar ambas as estratégias, e como se compara ao desempenho individual de cada uma em termos de eficácia e eficiência.

No que tange a essa última abordagem, acreditamos que se trata de uma interessante oportunidade. Considerando que os tópicos de notícias evoluem com o tempo, a possível combinação das abordagens propostas neste trabalho poderia tornar o método agnóstico a tópicos. Ao analisar características independentes do texto, como no caso de citações, é possível incorporar aspectos mais robustos à variação de tópicos tratados, permitindo a caracterização do viés de portais de notícias. Tal abordagem híbrida precisaria levar em consideração a construção de bases de dados contendo tanto texto quanto conteúdo HTML completo de maneira a representar de forma balanceada cada classe de viés político. Além das citações, outras características estruturais das páginas web podem ser exploradas, como proposto no trabalho de Castelo et al. (2019) para o contexto de identificação de *fake news*. Diante dos aspectos positivos, da relevância do tema e da precedência de trabalhos na literatura, ainda que em outros contextos, sugerimos fortemente que essa possibilidade seja trabalhada em uma futura continuação desta pesquisa.

Referências Bibliográficas

- Aires, V. P. e Nakamura, F. G. (2017). Detecção de comunidades em redes sociais: Relacionando o método de louvain a medidas de centralidade. Em *Proceedings of the 36th Concurso de Trabalhos de Iniciação Científica*.
- Aires, V. P., Nakamura, F. G., e Nakamura, E. F. (2019). A link-based approach to detect media bias in news websites. Em *Companion Proceedings of The 2019 World Wide Web Conference, WWW '19*, pgs. 742–745, New York, NY, USA. ACM.
- AllSides (2020). Allsides media bias ratings. Acesso em: Junho de 2020. Disponível em: <https://www.allsides.com/media-bias/media-bias-ratings/>.
- Baeza-Yates, R. e Davis, E. (2004). Web page ranking using link attributes. Em *Proceedings of the 13th international World Wide Web conference on Alternate track papers & posters*, pgs. 328–329. ACM.
- Baeza-Yates, R., Ribeiro-Neto, B., et al. (1999). *Modern information retrieval*, volume 463. ACM press New York.
- Baly, R., Karadzhov, G., Alexandrov, D., Glass, J., e Nakov, P. (2018). Predicting factuality of reporting and bias of news media sources. Em *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, Brussels, Belgium. Association for Computational Linguistics.
- Barabási, A.-L. (2003). *Linked: The new science of networks*.
- Barabási, A.-L. et al. (2016). *Network science*. Cambridge university press.
- Bedi, P. e Sharma, C. (2016). Community detection in social networks. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 6(3):115–135.
- Bishop, C. M. (2006). *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Springer-Verlag, Berlin, Heidelberg.

- Blom, J. e Reinecke Hansen, K. (2015). Click bait: Forward-reference as lure in online news headlines. *Journal of Pragmatics*, 76:87–100.
- Blondel, V. D., Guillaume, J.-L., Lambiotte, R., e Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Journal of statistical mechanics: theory and experiment*, 2008(10):P10008.
- Bollobás, B. e Riordan, O. M. (2003). Mathematical results on scale-free random graphs. *Handbook of graphs and networks: from the genome to the internet*, pgs. 1–34.
- Brandes, U. (2008). On variants of shortest-path betweenness centrality and their generic computation. *Social Networks*, 30(2):136–145.
- Broder, A., Kumar, R., Maghoul, F., Raghavan, P., Rajagopalan, S., Stata, R., Tomkins, A., e Wiener, J. (2000). Graph structure in the web. *Computer networks*, 33(1-6):309–320.
- Castelo, S., Almeida, T., Elghafari, A., Santos, A., Pham, K., Nakamura, E., e Freire, J. (2019). A topic-agnostic approach for identifying fake news pages. Em *Companion Proceedings of The 2019 World Wide Web Conference, WWW '19*, pg. 975–980, New York, NY, USA. Association for Computing Machinery.
- Chen, W.-F., Wachsmuth, H., Al-Khatib, K., e Stein, B. (2018). Learning to flip the bias of news headlines. Em *Proceedings of the 11th International Conference on Natural Language Generation*, pgs. 79–88, Tilburg University, The Netherlands. Association for Computational Linguistics.
- Chiang, C.-F. e Knight, B. (2011). Media bias and influence: Evidence from newspaper endorsements. *The Review of Economic Studies*, 78(3):795–820.
- Clauset, A., Newman, M. E., e Moore, C. (2004). Finding community structure in very large networks. *Physical review E*, 70(6):066111.
- D'Alessio, D. e Allen, M. (2000). Media bias in presidential elections: A meta-analysis. *Journal of Communication*, 50(4):133–156.
- Dallmann, A., Lemmerich, F., Zoller, D., e Hotho, A. (2015). Media bias in german online newspapers. Em *Proceedings of the 26th ACM Conference on Hypertext & Social Media*, pgs. 133–137. ACM.

- Dehmer, M. e Emmert-Streib, F. (2009). *Analysis of complex networks: from biology to linguistics*. John Wiley & Sons.
- DellaVigna, S. e Kaplan, E. (2007). The fox news effect: Media bias and voting. *The Quarterly Journal of Economics*, 122(3):1187–1234.
- DellaVigna, S. e Kaplan, E. (2008). The political impact of media bias. *Information and Public Choice*, 79.
- Diestel, R. (2018). *Graph theory*. Springer Publishing Company, Incorporated.
- Dietterich, T. G. (2000). Ensemble methods in machine learning. Em *International workshop on multiple classifier systems*, pgs. 1–15. Springer.
- Dua, D. e Graff, C. (2019). Uci machine learning repository.
- Efron, M. (2004). The liberal media and right-wing conspiracies: using cocitation information to estimate political orientation in web documents. Em *Proceedings of the Thirteenth ACM International Conference on Information and Knowledge Management*, pgs. 390–398. ACM.
- Elejalde, E., Ferres, L., e Herder, E. (2017). The nature of real and perceived bias in chilean media. Em *Proceedings of the 28th ACM Conference on Hypertext and Social Media*, pgs. 95–104. ACM.
- Endres, D. M. e Schindelin, J. E. (2003). A new metric for probability distributions. *IEEE Transactions on Information Theory*, 49(7):1858–1860.
- Fortuna, B., Galleguillos, C., e Cristianini, N. (2009). Detection of bias in media outlets with statistical learning methods. Em *Text Mining*, pgs. 57–80. Chapman and Hall/CRC.
- Gentzkow, M. e Shapiro, J. M. (2006). Media bias and reputation. *Journal of Political Economy*, 114(2):280–316.
- Girvan, M. e Newman, M. E. (2002). Community structure in social and biological networks. *Proceedings of the national academy of sciences*, 99(12):7821–7826.
- Golbeck, J. (2015). *Introduction to social media investigation: a hands-on approach*. Syngress.
- Gordon, J., Babaeianjelodar, M., e Matthews, J. (2020). Studying political bias via word embeddings. Em *Companion Proceedings of the Web Conference 2020, WWW '20*, pg. 760–764, New York, NY, USA. Association for Computing Machinery.

- Gottfried, J. e Shearer, E. (2016). *News Use Across Social Medial Platforms 2016*. Pew Research Center.
- Griffiths, D. J. e Schroeter, D. F. (2018). *Introduction to quantum mechanics*. Cambridge University Press.
- Harper, R. A. (2010). The social media revolution: Exploring the impact on journalism and news media organizations. 2(03).
- Heinze, A. R. (2006). The political economy of mass consumption. *Journal of Urban History*, 32(4):607–618.
- Kernighan, B. W. e Lin, S. (1970). An efficient heuristic procedure for partitioning graphs. *Bell system technical journal*, 49(2):291–307.
- Knoche, M., Popović, R., Lemmerich, F., e Strohmaier, M. (2019). Identifying biases in politically biased wikis through word embeddings. Em *Proceedings of the 30th ACM Conference on Hypertext and Social Media*, HT '19, pgs. 253–257, New York, NY, USA. Association for Computing Machinery.
- Krestel, R., Wall, A., e Nejd, W. (2012). Treehugger or petrolhead?: identifying bias by comparing online news articles with political speeches. Em *Proceedings of the 21st International Conference on World Wide Web*, pgs. 547–548. ACM.
- Levandowsky, M. e Winter, D. (1971). Distance between sets. *Nature*, 234(5323):34–35.
- Lin, J. (1991). Divergence measures based on the shannon entropy. *IEEE Transactions on Information Theory*, 37(1):145–151.
- Lin, Y.-R., Bagrow, J. P., e Lazer, D. (2011). More voices than ever? quantifying media bias in networks. Em *Fifth International AAAI Conference on Weblogs and Social Media*.
- Lipkus, A. H. (1999). A proof of the triangle inequality for the tanimoto distance. *Journal of Mathematical Chemistry*, 26(1-3):263–265.
- Media Bias Fact Check (2019). The most comprehensive media bias resource. Acesso em: Julho de 2019. Disponível em: <https://mediabiasfactcheck.com/>.
- Mitchell, T. M. (1997). *Machine Learning*. McGraw-Hill, Inc., New York, NY, USA, 1 edition.

- Morstatter, F., Wu, L., Yavanoglu, U., Corman, S. R., e Liu, H. (2018). Identifying framing bias in online news. *ACM Transactions on Social Computing*, 1(2):5.
- Newman, M. E. (2006). Finding community structure in networks using the eigenvectors of matrices. *Physical review E*, 74(3):036104.
- Newman, M. E. J. (2010). *Networks: an introduction*. Oxford University Press.
- Paliouras, G., Karkaletsis, V., e Spyropoulos, C. D., editores (2001). *Machine Learning and Its Applications, Advanced Lectures*, London, UK, UK. Springer-Verlag.
- Pang, B., Lee, L., e Vaithyanathan, S. (2002). Thumbs up? sentiment classification using machine learning techniques. *arXiv preprint cs/0205070*.
- Park, S., Ko, M., Kim, J., Liu, Y., e Song, J. (2011). The politics of comments: predicting political orientation of news stories with commenters' sentiment patterns. Em *Proceedings of the ACM 2011 conference on Computer supported cooperative work*, pgs. 113–122. ACM.
- Perra, N. e Fortunato, S. (2008). Spectral centrality measures in complex networks. *Physical Review E*, 78(3):036107.
- Rao, A. e Spasojevic, N. (2016). Actionable and political text classification using word embeddings and lstm. Em *Proceedings of the 5th International Workshop on Issues of Sentiment Discovery and Opinion Mining (WISDOM)*.
- Reichardt, J. e Bornholdt, S. (2006). Statistical mechanics of community detection. *Physical Review E*, 74(1):016110.
- Ribeiro, F. N., Henrique, L., Benevenuto, F., Chakraborty, A., Kulshrestha, J., Babaei, M., e Gummadi, K. P. (2018). Media bias monitor: Quantifying biases of social media news outlets at large-scale. Em *Twelfth International AAAI Conference on Web and Social Media*.
- Saez-Trumper, D., Castillo, C., e Lalmas, M. (2013). Social media news communities: gatekeeping, coverage, and statement bias. Em *Proceedings of the 22nd ACM international conference on Information & Knowledge Management*, pgs. 1679–1684. ACM.
- Scholz, M. (2012). Node degree distribution. Disponível em: www.network-science.org/powerlaw_scalefree_node_degree_distribution.html. Acesso em: 08/2019.

- Shannon, C. (1948). A mathematical theory of communication. *Bell System Technical Journal*, 27(3):379–423.
- Shen, H.-W. (2013). *Community structure of complex networks*. Springer Science & Business Media.
- Shoemaker, P. e Reese, S. (1996). *Mediating the message: Theories of influence on mass media content..* hoboken.
- Shu, K., Sliva, A., Wang, S., Tang, J., e Liu, H. (2017). Fake news detection on social media: A data mining perspective. *ACM SIGKDD Explorations Newsletter*, 19(1):22–36.
- The Batch (2020). Facebook’s unruly algorithm, ai that does the dishes, new life for old data, models that take shortcuts, yolo returns. Acesso em: Junho de 2020. Disponível em: <https://blog.deeplearning.ai/blog/the-batch-facebooks-unruly-algorithm-ai-that-does-the-dishes-new-life-for-old-data-models-that-take-shortcuts-yolo-returns>.
- Tiropanis, T., Hall, W., Crowcroft, J., Contractor, N., e Tassiulas, L. (2015). Network science, web science, and internet science. *Communications of the ACM*, 58(8):76–82.
- Travers, J. e Milgram, S. (1967). The small world problem. *Psychology Today*, 1:61–67.
- Van Der Hofstad, R. (2016). *Random graphs and complex networks*, volume 1. Cambridge University Press.
- Vinodhini, G. e Chandrasekaran, R. (2012). Sentiment analysis and opinion mining: a survey. *International Journal*, 2(6):282–292.
- Wasserman, S., Faust, K., et al. (1994). *Social network analysis: Methods and applications*, volume 8. Cambridge university press.
- West, D. B. et al. (2001). *Introduction to graph theory*, volume 2. Prentice hall Upper Saddle River.
- Wong, T.-T. (2015). Performance evaluation of classification algorithms by k-fold and leave-one-out cross validation. *Pattern Recognition*, 48(9):2839–2846.
- Wu, F.-Y. (1982). The potts model. *Reviews of modern physics*, 54(1):235.
- Yang, Z., Algesheimer, R., e Tessone, C. J. (2016). A comparative analysis of community detection algorithms on artificial networks. *Scientific reports*, 6:30750.

- Yigit-Sert, S., Altingovde, I. S., e Ulusoy, Ö. (2016). Towards detecting media bias by utilizing user comments. Em *Proceedings of the 8th ACM Conference on Web Science*, pgs. 374–375. ACM.

Apêndice A

Resultados adicionais da classificação baseada em padrões de citações

Neste apêndice, incluímos resultados obtidos ao computar as dissimilaridades do método baseado em padrões de citações através da distância de Jaccard. Como mencionado no Capítulo 4, tais resultados são muito similares aos observados ao calcular usando a divergência de Jensen-Shannon. Apesar do desempenho próximo, as métricas deixam claro que a divergência de Jensen-Shannon é mais adequada para esse contexto, obtendo melhores valores. Assim, decidimos listar esses resultados obtidos ao usar distância de Jaccard apenas como complemento, para prover uma visão geral do método.

Tabela A.1: Valores de acurácia e precisão, revocação e F1 por classe para cada rede ao classificar o viés ideológico usando padrões de citações e distância de Jaccard.

Redes	Precisão		Revocação		F1		Acurácia
	Esquerda	Direita	Esquerda	Direita	Esquerda	Direita	
Rede 1	0,73	0,79	0,92	0,47	0,81	0,59	75%
Rede 2	0,76	0,73	0,71	0,77	0,73	0,75	74%
Rede 3	0,72	0,72	0,72	0,72	0,72	0,72	72%

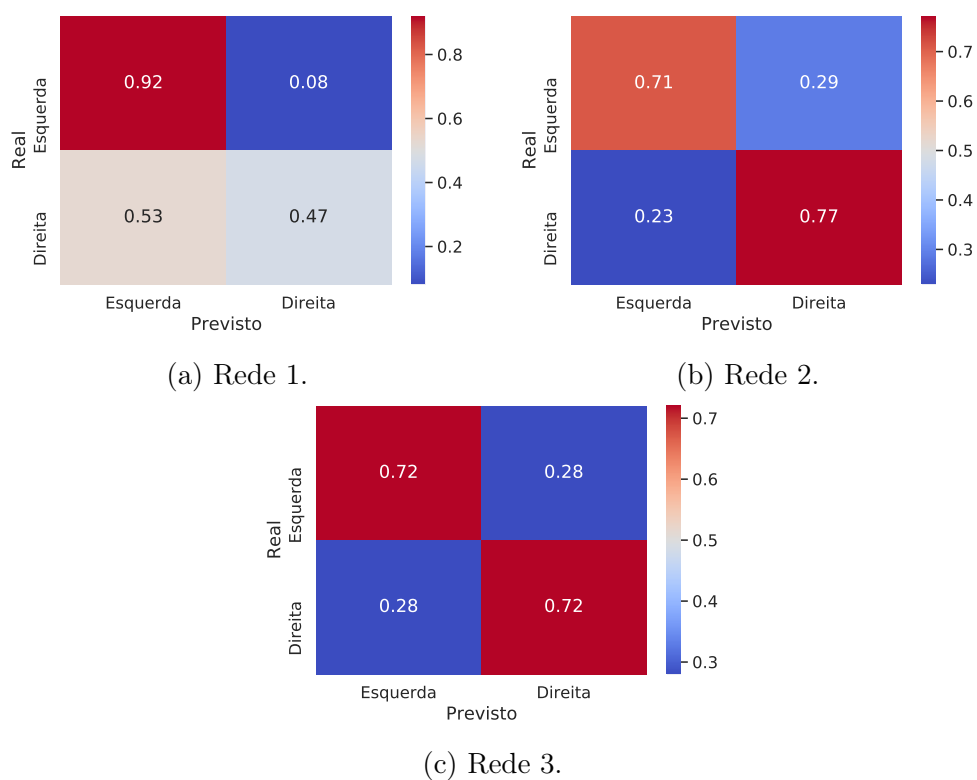


Figura A.1: Matrizes de confusão obtidas pelo método ao classificar as redes 1, 2 e 3, quando a dissimilaridade foi calculada através da distância de Jaccard. Os valores estão normalizados por linhas.

Apêndice B

Resultados adicionais da classificação baseada em conteúdo textual

Neste apêndice, incluímos resultados observados ao experimentar o método usando **unigramas** para modelar o vocabulário e comparando os resultados ao computar dissimilaridades usando **distância de Jaccard** além da divergência de Jensen-Shannon. Devido à acurácia obtida por essas combinações, e ao desempenho superior ao empregar Jensen-Shannon, decidimos prosseguir utilizando bigramas e divergência de Jensen-Shannon, resultados apresentados no Capítulo 5. A seguir, descrevemos os valores observados em cada base de dados, para cada tarefa de classificação.

B.1 Classificando extremos e moderados

Tabela B.1: Resultados do método classificando websites de vieses extremos (Ex) e moderados (M) na base de dados Noticias-Julho usando unigramas. Em negrito, o melhor resultado em cada métrica.

Métodos	Dissimilaridades	Precisão		Revocação		F1		Acurácia
		Ex	M	Ex	M	Ex	M	
Estratégia 1	Jaccard	0,81	0,81	0,81	0,81	0,81	0,81	81%
	JSD	0,82	0,82	0,82	0,82	0,82	0,82	82%
Estratégia 2	Jaccard	0,79	0,72	0,69	0,81	0,73	0,77	75%
	JSD	0,84	0,78	0,76	0,85	0,80	0,82	81%
Estratégia 3	Jaccard	0,80	0,79	0,78	0,81	0,79	0,80	79%
	JSD	0,82	0,80	0,80	0,82	0,81	0,81	81%
Estratégia 4	Jaccard	0,83	0,79	0,78	0,84	0,80	0,82	81%
	JSD	0,85	0,82	0,81	0,85	0,83	0,84	83%

Tabela B.2: Resultados do método classificando websites de vieses extremos (Ex) e moderados (M) na base de dados Noticias-Fevereiro usando unigramas. Em negrito, o melhor resultado em cada métrica.

Métodos	Dissimilaridades	Precisão		Revocação		F1		Acurácia
		Ex	M	Ex	M	Ex	M	
Estratégia 1	Jaccard	0,68	0,67	0,66	0,69	0,67	0,68	68%
	JSD	0,73	0,68	0,65	0,75	0,69	0,72	70%
Estratégia 2	Jaccard	0,68	0,63	0,56	0,73	0,61	0,68	65%
	JSD	0,72	0,69	0,66	0,75	0,69	0,72	70%
Estratégia 3	Jaccard	0,69	0,65	0,60	0,73	0,64	0,69	67%
	JSD	0,77	0,71	0,68	0,79	0,72	0,75	74%
Estratégia 4	Jaccard	0,70	0,65	0,61	0,73	0,65	0,69	67%
	JSD	0,77	0,71	0,67	0,80	0,72	0,75	74%

Tabela B.3: Resultados do método classificando websites de vieses extremos (Ex) e moderados (M) na base de dados News-Aggregator usando unigramas. Em negrito, o melhor resultado em cada métrica.

Métodos	Dissimilaridades	Precisão		Revocação		F1		Acurácia
		Ex	M	Ex	M	Ex	M	
Estratégia 1	Jaccard	0,76	0,72	0,70	0,78	0,73	0,75	74%
	JSD	0,80	0,71	0,67	0,83	0,73	0,77	75%
Estratégia 2	Jaccard	0,80	0,72	0,68	0,83	0,74	0,78	76%
	JSD	0,85	0,76	0,73	0,87	0,79	0,81	80%
Estratégia 3	Jaccard	0,83	0,76	0,73	0,85	0,78	0,80	79%
	JSD	0,83	0,77	0,75	0,85	0,79	0,81	80%
Estratégia 4	Jaccard	0,82	0,72	0,67	0,85	0,73	0,78	76%
	JSD	0,82	0,78	0,77	0,83	0,79	0,81	80%

B.2 Classificando portais de esquerda e direita

Tabela B.4: Resultados do método classificando websites de vieses de esquerda (E) e direita (D) na base de dados Noticias-Julho usando unigramas. Em negrito, o melhor resultado em cada métrica.

Métodos	Dissimilaridades	Precisão		Revocação		F1		Acurácia
		E	D	E	D	E	D	
Estratégia 1	Jaccard	0,71	0,62	0,52	0,78	0,60	0,69	65%
	JSD	0,73	0,63	0,53	0,80	0,61	0,70	67%
Estratégia 2	Jaccard	0,61	0,61	0,60	0,62	0,61	0,62	61%
	JSD	0,66	0,67	0,69	0,65	0,67	0,66	67%
Estratégia 3	Jaccard	0,77	0,71	0,67	0,80	0,72	0,75	73%
	JSD	0,81	0,74	0,70	0,83	0,75	0,78	77%
Estratégia 4	Jaccard	0,74	0,70	0,68	0,77	0,71	0,73	72%
	JSD	0,77	0,72	0,69	0,80	0,73	0,76	75%

Tabela B.5: Resultados do método classificando websites de vieses de esquerda (E) e direita (D) na base de dados Noticias-Fevereiro usando unigramas. Em negrito, o melhor resultado em cada métrica.

Métodos	Dissimilaridades	Precisão		Revocação		F1		Acurácia
		E	D	E	D	E	D	
Estratégia 1	Jaccard	0,71	0,69	0,68	0,72	0,69	0,71	70%
	JSD	0,77	0,76	0,75	0,77	0,76	0,77	76%
Estratégia 2	Jaccard	0,64	0,65	0,62	0,62	0,65	0,64	64%
	JSD	0,65	0,67	0,68	0,64	0,67	0,65	66%
Estratégia 3	Jaccard	0,70	0,70	0,70	0,70	0,70	0,70	70%
	JSD	0,76	0,78	0,78	0,75	0,77	0,76	77%
Estratégia 4	Jaccard	0,71	0,71	0,71	0,71	0,71	0,71	71%
	JSD	0,75	0,77	0,77	0,74	0,76	0,75	76%

Tabela B.6: Resultados do método classificando websites de vieses de esquerda (E) e direita (D) na base de dados News-Aggregator usando unigramas. Em negrito, o melhor resultado em cada métrica.

Métodos	Dissimilaridades	Precisão		Revocação		F1		Acurácia
		E	D	E	D	E	D	
Estratégia 1	Jaccard	0,54	0,53	0,48	0,58	0,51	0,56	53%
	JSD	0,57	0,55	0,50	0,62	0,53	0,58	56%
Estratégia 2	Jaccard	0,62	0,73	0,82	0,50	0,71	0,59	66%
	JSD	0,62	0,70	0,77	0,53	0,69	0,60	65%
Estratégia 3	Jaccard	0,75	0,77	0,78	0,73	0,76	0,75	76%
	JSD	0,80	0,85	0,87	0,78	0,83	0,82	82%
Estratégia 4	Jaccard	0,72	0,72	0,72	0,72	0,72	0,72	72%
	JSD	0,79	0,80	0,82	0,78	0,80	0,80	80%

B.3 Classificando portais de esquerda, centro-esquerda, centro-direita e direita

Tabela B.7: Resultados do método classificando websites de esquerda (E), centro-esquerda (CE), centro-direita (CD) e direita (D) na base de dados Noticias-Julho usando unigramas. Em negrito, o melhor resultado em cada métrica.

Métodos	Dissimilaridades	Precisão				Revocação				F1				Acurácia
		E	CE	CD	D	E	CE	CD	D	E	CE	CD	D	
Estratégia 1	Jaccard	0,57	0,50	0,45	0,57	0,47	0,29	0,68	0,63	0,51	0,37	0,54	0,60	52%
	JSD	0,61	0,49	0,44	0,59	0,50	0,27	0,66	0,65	0,55	0,35	0,53	0,62	52%
Estratégia 2	Jaccard	0,58	0,52	0,36	0,63	0,60	0,55	0,47	0,39	0,59	0,53	0,41	0,48	50%
	JSD	0,68	0,60	0,47	0,77	0,76	0,56	0,60	0,53	0,72	0,58	0,53	0,63	61%
Estratégia 3	Jaccard	0,72	0,63	0,48	0,65	0,66	0,53	0,68	0,53	0,69	0,58	0,56	0,58	60%
	JSD	0,81	0,64	0,50	0,67	0,74	0,55	0,65	0,63	0,77	0,59	0,56	0,65	64%
Estratégia 4	Jaccard	0,70	0,62	0,47	0,56	0,65	0,56	0,60	0,52	0,67	0,59	0,53	0,54	58%
	JSD	0,76	0,60	0,52	0,60	0,68	0,56	0,56	0,66	0,72	0,58	0,54	0,63	62%

Tabela B.8: Resultados do método classificando websites de esquerda (E), centro-esquerda (CE), centro-direita (CD) e direita (D) na base de dados Noticias-Fevereiro usando unigramas. Em negrito, o melhor resultado em cada métrica.

Métodos	Dissimilaridades	Precisão				Revocação				F1				Acurácia
		E	CE	CD	D	E	CE	CD	D	E	CE	CD	D	
Estratégia 1	Jaccard	0,53	0,45	0,48	0,60	0,60	0,40	0,53	0,54	0,56	0,42	0,50	0,57	52%
	JSD	0,64	0,52	0,55	0,64	0,62	0,48	0,64	0,61	0,63	0,50	0,59	0,63	59%
Estratégia 2	Jaccard	0,45	0,36	0,41	0,58	0,45	0,33	0,51	0,47	0,45	0,34	0,46	0,52	44%
	JSD	0,41	0,46	0,44	0,62	0,56	0,39	0,51	0,56	0,53	0,42	0,47	0,58	50%
Estratégia 3	Jaccard	0,53	0,45	0,44	0,61	0,48	0,44	0,52	0,58	0,51	0,45	0,48	0,59	51%
	JSD	0,63	0,51	0,52	0,70	0,58	0,56	0,55	0,66	0,60	0,53	0,53	0,68	59%
Estratégia 4	Jaccard	0,51	0,42	0,42	0,64	0,49	0,38	0,53	0,58	0,50	0,40	0,47	0,61	49%
	JSD	0,60	0,49	0,51	0,73	0,56	0,51	0,55	0,67	0,58	0,50	0,53	0,70	57%

Tabela B.9: Resultados do método classificando websites de esquerda (E), centro-esquerda (CE), centro-direita (CD) e direita (D) na base de dados News-Aggregator usando unigramas. Em negrito, o melhor resultado em cada métrica.

Métodos	Dissimilaridades	Precisão				Revocação				F1				Acurácia
		E	CE	CD	D	E	CE	CD	D	E	CE	CD	D	
Estratégia 1	Jaccard	0,53	0,33	0,36	0,72	0,63	0,23	0,53	0,43	0,58	0,27	0,43	0,54	46%
	JSD	0,53	0,35	0,37	0,76	0,60	0,27	0,57	0,43	0,56	0,30	0,45	0,55	47%
Estratégia 2	Jaccard	0,54	0,51	0,67	0,65	0,50	0,67	0,67	0,50	0,52	0,58	0,67	0,57	58%
	JSD	0,50	0,62	0,68	0,64	0,50	0,77	0,63	0,53	0,50	0,69	0,66	0,58	61%
Estratégia 3	Jaccard	0,69	0,48	0,55	0,76	0,60	0,50	0,70	0,63	0,64	0,49	0,62	0,69	61%
	JSD	0,74	0,55	0,66	0,84	0,77	0,53	0,77	0,70	0,75	0,54	0,71	0,76	69%
Estratégia 4	Jaccard	0,65	0,50	0,57	0,78	0,57	0,50	0,70	0,70	0,61	0,50	0,63	0,74	62%
	JSD	0,69	0,53	0,66	0,77	0,67	0,53	0,77	0,67	0,68	0,53	0,71	0,71	66%