Universidade Federal do Estado do Rio de Janeiro

Centro de Ciências Políticas e Jurídicas

Escola de Administração

A ANÁLISE DE TEXTO DOS TWITTER DOS CONSUMIDORES DO NETFLIX ATRAVÉS DO MÉTODO LATENT DIRICHLET ALLOCATION

VINICIUS MENCARINI DE MELLO

Orientador

ROSSANDRO RAMOS

Rio de Janeiro, RJ – Brasil

Dezembro de 2020

VINICIUS MENCARINI DE MELLO

ANÁLISE DE TWITTERS DE CONSUMIDORES DO NETFLIX ATRAVÉS DO MÉTODO LATENT DIRICHLET ALLOCATION

Monografia apresentada à Escola de Administração da Universidade Federal do Estado do Rio de Janeiro (UNIRIO) para obtenção do título de Bacharel em Administração Pública.

Orientador

ROSSANDRO RAMOS

Rio de Janeiro, RJ – Brasil

Dezembro de 2020

ANÁLISE DE TWITTERS DE CONSUMIDORES DO NETFLIX ATRAVÉS DO MÉTODO LATENT DIRICHLET ALLOCATION

VINICIUS MENCARINI DE MELLO

Monografia apresentada à Escola de Administração da Universidade Federal do Estado do Rio de Janeiro (UNIRIO) para obtenção do título de Bacharel em Administração Pública.

Aprovado por:

_____

ROSSANDRO RAMOS

_____

RICARDO LUIZ DO NASCIMENTO

_____

JULIO CÉSAR SILVA MACEDO

Rio de Janeiro, RJ – Brasil.

Dezembro de 2020

Agradecimentos

Dedico este trabalho a todos que me ajudaram ao longo do curso diretamente ou indiretamente. E agradeço ainda à minha família que me apoiou sempre durante todo o curso.

RESUMO

Este trabalho se utiliza da técnica Latent Dirichlet Allocation (LDA) a fim de analisar tweets coletados de respostas dos consumidores que responderam ao Twitter do Netflix Brasil. Além da análise dos resultados, são discutidas as forças e fraquezas desta técnica. Este trabalho então serve para ilustrar como o topic modeling, ou modelagem de tópico, pode se tornar uma ferramenta útil para se entender melhor os consumidores usuários de redes sociais e para auxiliar dirigentes de governos e empresas nas suas definições de políticas públicas ou empresariais. Sendo constatado a versatilidade do LDA em transformar uma grande quantidade de dados em algo que sirva para facilitar a tomada de decisões.

Palavras-chave: Latent Dirichlet Allocation, Redes Sociais, Text Mining, Topic modeling, Twitter.

ABSTRACT

This work uses the Latent Dirichlet Allocation (LDA) technique in order to analyze tweets collected from responses from consumers who responded to Netflix Brasil's Twitter. In addition to analyzing the results, the strengths and weaknesses of this technique are discussed. This work then serves to illustrate how topic modeling can become a useful tool to better understand consumers using social networks and to assist managers of governments and companies in their definitions of public or business policies. Being verified the versatility of the LDA in transforming a great amount of data in something that serves to facilitate the decision making.

Keywords: Latent Dirichlet Allocation, Social Media, Text Mining, Topic modeling, Twitter.

Sumário

Índice de Figuras

1 INTRODUÇÃO

1.1 CARACTERIZAÇÃO E IMPORTÂNCIA DO TEMA

A cada ano sobe o usuários de redes sociais, indo de 970 milhões de usuários em 2010 para estimados 2.77 bilhões em 2017, segundo dados do eMarketer (2019). Os governos e as empresas, então, cada vez mais, tentam se utilizar dessas redes a fim de definir políticas e estratégias públicas ou comerciais; buscar novas ideias de inovação; utilizar o feedback dessas redes tanto para atendimento do cidadão/consumidor como para auxílio na compra ou no pós-venda; reforçar a lealdade do consumidor com a marca; facilitar o posicionamento da marca em relação a assuntos discutidos em períodos recentes; e facilitador para o diálogo com o cliente de maneira geral. Estudos já apontavam para um aumento de 50% do número de CEOs utilizando redes sociais, crescendo de 42% em 2013 para 63% 2018. Isso mostra uma grande preocupação das empresas em estabelecer e se posicionar nessas redes sociais, como reportado por Weber Shandwick (2013). Portanto, cada vez mais, é essencial descobrir como obter, estruturar e analisar os dados de acordo com os objetivos estratégicos seja de governos, seja de empresas públicas ou privadas.

A rede social escolhida para o trabalho foi o Twitter, a qual é uma das maiores redes sociais da atualidade, tendo mais de 321 milhões de usuários ativos em 2018, segundo o eMarketer (2019). Por ter sido criado com foco em textos pequenos, o Twitter contém informações riquíssimas para que governos e empresas públicas ou privadas que utilizarem as ferramentas certas possam estruturar esses dados e gerar conhecimento. Afinal, esses textos pequenos, condensados de opinião, possibilitam múltiplas análises e estudos, caso sejam propriamente tratados. Como destaca Zhao et Al (2011), o Twitter é uma ótima rede social para que se possa mais facilmente retirar dados úteis dos consumidores sobre uma empresa.

Cabe destacar que a importância global das redes sociais torna fundamental que governos e empresas estabeleçam métodos para administrar a relação entre fornecedores e usuários. Uma das ferramentas para tratamento de dados dessa relação é a análise de conteúdo usando topic

modeling. O topic modeling trabalha com dados das redes sociais e permite a quantificação de opiniões e a análise de sentimentos por meio dos textos subjetivos publicados pelos internautas (CIRQUEIRA et al., 2017). Já o método Latent Dirichlet Allocation (LDA) combina as partes léxicas e de machine learning da modelagem de tópico, sendo um modelo híbrido com maior flexibilidade. Ambos os métodos foram utilizados neste trabalho.

1.2 OBJETIVOS

Objetivo Geral: realizar a análise de dados dessa dos principais tópicos e temas abordados pelos consumidores do Netflix Brasil no Twitter no período deste trabalho. Esses tópicos, sendo corretamente entendidos, podem levar a uma melhor compreensão dos consumidores do Netflix Brasil.

Objetivos específicos: proceder uma revisão de literatura da técnica LDA; analisar os uso de ferramentas de análise das redes sociais e plataformas por empresas públicas e privadas; e abordar o uso de modelagem de tópico em redes sociais.

1.3 ESTRUTURA DO TRABALHO

O presente trabalho encontra-se estruturado em quatro capítulos, além desta introdução. No capítulo 2 - Revisão de Literatura, será abordado o papel das redes sociais na modelagem de tópicos buscando entender o LDA em uma organização especifica na plataforma Twitter, fazendo uso do software Mallet.

No capítulo 3 - Metodologia, apresentamos o percurso metodológico da pesquisa. No capítulo 4 - Resultado e discussão, faz-se apresentação dos resultados obtidos e uma breve discussão. No capítulo 5 realiza-se as considerações finais.

2 REVISÃO DA LITERATURA

A revisão de literatura foi dividida em cinco itens: rede social, que fala das redes sociais e sua importância para os administradores; topic modeling, que disserta sobre o assunto e como ele se relaciona com os estudos de rede sociais; Latent Dirichlet Allocation, explicando como a técnica LDA funciona; Netflix, expondo os detalhes do objeto de estudo; o programa usado neste trabalho, o Mallet; e o Twitter, que foi a rede social escolhida para se adquirir os dados.

2.1 REDES SOCIAIS

As redes sociais são espaços virtuais criados para que grupos de pessoas e empresas se relacionem através de mensagens, do compartilhamento de conteúdos, entre outros. As redes sociais abriram um mar de possibilidades novas para governos e empresas (DIJKSTRA *et al.*, 2018), sendo uma das mais importantes o fato de terem ganhado acesso direto a feedbacks dos usuários desses serviços virtuais. Além de tentar resolver problemas e aumentar o engajamento, essas plataformas servem como intermediárias entre empresas-consumidores e entre pessoas de diferentes grupos sociais, além de permitir a criação de conteúdos novos ou modificados. Por exemplo, uma pessoa pode pegar uma propaganda de determinada empresa e criar um vídeo, imagem ou texto, que pode gerar repercussões positivas ou negativas -- e de maneira muito veloz e abrangente. Por isso Berthon et al. (2012) ressalta que o ambiente digital pode promover grandes sucessos e desastres.

Um importante problema na administração das redes sociais é saber tratar e analisar os diferentes conteúdos produzidos internamente e por terceiros, devido à grande quantidade de informações à disposição das empresas. Cabe ao administrador encontrar uma forma de trabalhar esses dados não estruturados e de formatos diversos, fazendo isso de maneira eficaz para que se gere resultados compreensíveis e úteis para a tomada de decisões posteriores (CIRQUEIRA et al., 2017).

O artigo de Nelson (2018) apresenta a penetração das empresas da Fortune 500 nas redes sociais, tendo como base dados de 2018: 91% das empresas tinham conta no Twitter, 89% no Facebook e 63% no Instagram. Embora mostrem que a maioria das empresas já está acompanhando esta tendência, o fato é que mesmo entre as maiores empresas do mundo ainda há dificuldades em administrar a relação com os usuários das redes sociais. A situação piora, tanto em termos de cobertura quanto de relacionamento com os usuários, nas empresas de menor porte e com menos recursos para processar o gigantesco arsenal de conteúdos interativos.

Há casos também de empresas que aderem às redes sociais, mas por falta de uma estratégia real ligada ao projeto, acabam não aproveitando corretamente as vantagens das diferentes plataformas ou mesmo acabam deixando-as abandonadas após os primeiros momentos de sua criação.

2.2 TOPIC MODELING

A principal função do topic modelling é identificar as partes interessantes de um documento, sendo que, incialmente, foi mais utilizado em análises de revistas e jornais por eles terem conteúdos mais homogêneos (WAAL; MOUTON, 2013). A “Análise de Sentimentos” é uma das ferramentas do topic modeling e tem como objetivo lidar com a subjetividade e grande quantidade de dados produzidos pelas redes sociais. Basicamente existem três maneiras de se usar a “Análise de Sentimentos”: utilizando-se de um método léxico, em que se cria um dicionário em que várias palavras são associadas a um único tópico. O texto sendo então mostrado sob a ótica desse dicionário, podendo ser uma relação entre Negativo e Positivo por exemplo (SOUZA; PEREIRA; DALIP, 2017). A segunda maneira é o machine learning, em que algoritmos são especificados e calculados pela máquina para que ela chegue em um resultado. E, por fim, um método hibrido que mistura ambos.

Podemos usar a analogia de Graham, Weingart e Milligan (2012) em que se tem cestas com várias palavras de um tema, o programa iria então checar cada frase e tentar encontrar os tópicos da frase. Dessa maneira pode se achar os temas debatidos em quantidades altas de texto

de maneira matemática e possibilitando que computadores posam tirar modelos quantitativos de textos qualitativos.

Animais
leão
tigre
elefante
camelo
cobra

Refeitório
refrigerante
hambúrguer
bebida
comida
chocolate

Sentimentos Positivos
diversão
alegria
contente
feliz
correto

Sentimentos Negativos
crueldade
horrível
triste
desanimado
chateado

Figura 1 - Exemplo de tópicos que poderiam ser usados por um zoológico.
Fonte: autoria própria.

O topic modeling utiliza-se de modelos probabilísticos e cria uma relação entre textos e variáveis latentes (CIRQUEIRA et al., 2017). No exemplo acima, na Figura 1, um Zoológico hipotético poderia criar tópicos relacionados aos temas que se deseja entender. O processamento dos documentos em vista desses tópicos poderia então trazer resultados quantitativos dos documentos.

Um ponto importante do topic modeling é que ele pode ser utilizado em bancos de dados não estruturados, o que facilita sua utilização para lidar com textos de redes sociais que são analisados. Outra vantagem do topic modeling é o fato de se poder criar um resultado que se encaixe aos parâmetros do problema. Os processos automáticos e vetores geram resultados em praticamente qualquer conjunto de textos. Uma das técnicas de modelagem de tópico é o LDA, a qual é a utilizada neste trabalho, sendo uma das técnicas mais utilizadas segundo Ibrahim e Wang (2019).

2.3 LATENT DIRICHLET ALLOCATION

A principal característica da técnica LDA é a manipulação de dados para que cada documento consiga gerar um vetor de probabilidade de tópico (CAMPR;JEZEK 2013). Cada sentença então é analisada pela média do peso de cada vetor de tópico que se encontra no documento, como é visto na Figura 2. De maneira simplificada: um banco de dados dividido em tópicos e montado, os documentos são tokenizados e limpos. Esses documentos são então organizados de acordo com os tópicos e o peso referente destes tópicos. Originalmente, o próprio exemplo de seu criador sugeria um uso em banco de dados que continham livros e revistas (Blei et al., 2003), sendo que sua utilização para redes sociais aconteceu em momento posterior.

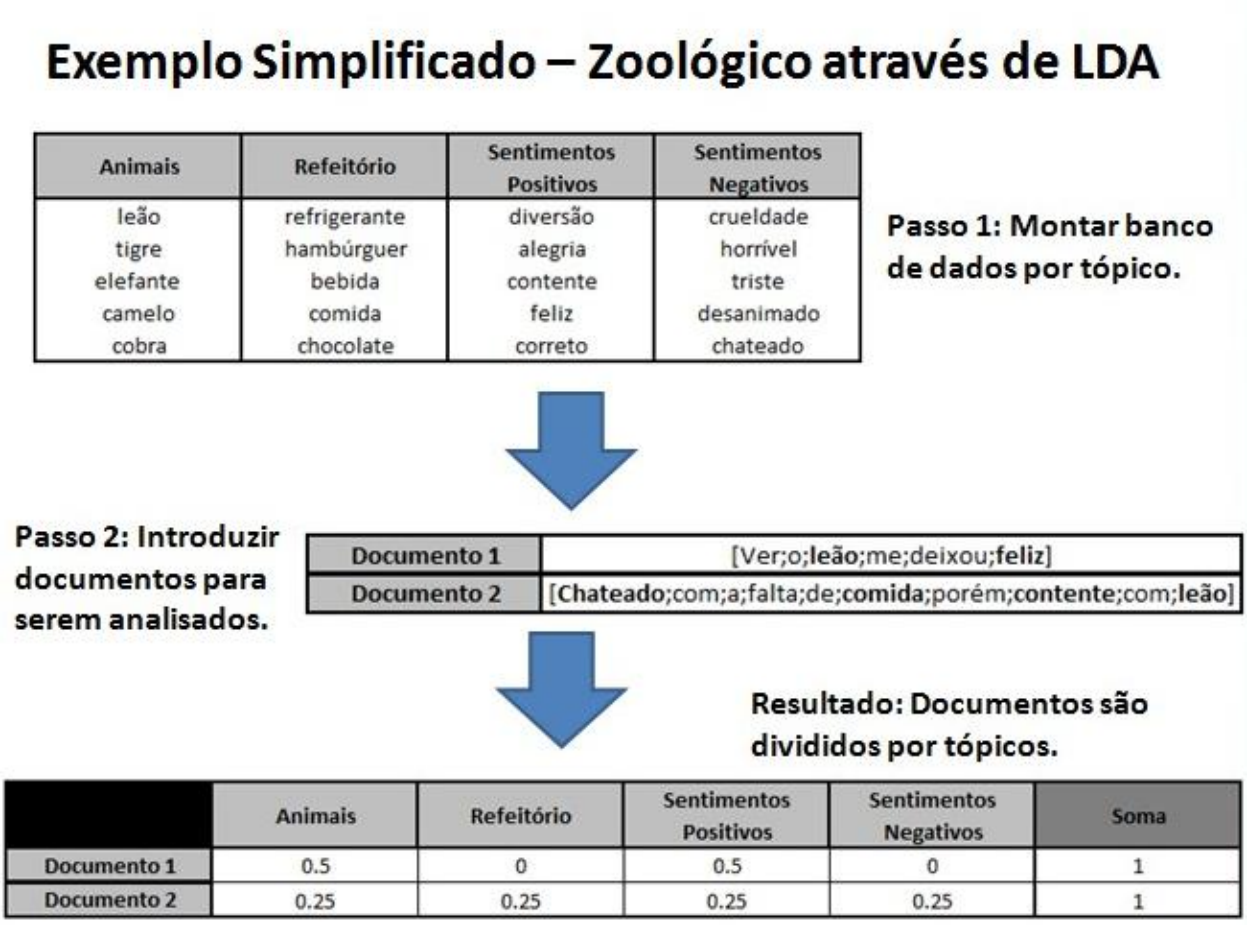


Figura 2 - Exemplo simplificado - Zoológico através do LDA.
Fonte: autoria própria.

Pode-se com a técnica, então, medir o peso que cada tópico tem em relação ao texto. Na Figura 2 o tema Animais comporia 50% (0.5) do Documento 1 e 25% (0.25) do Documento 2. Tudo isso feito de maneira não supervisionada por humanos, sendo todo o processo executado pelo computador automaticamente (Blei et al., 2003).

A técnica LDA tem então como figura chave a distribuição das palavras: cada documento é analisado de acordo com um tópico, cada palavra do documento é analisada em relação ao tópico e por fim se repete isso para cada palavra de cada documento (CIRQUEIRA et al., 2017). Repetido o processo pelo número de tópicos que existe.

As principais vantagens do modelo são sua modularidade, sendo facilmente conectado com outros programas e modelos; e sua capacidade de lidar com um número imenso de dados (Blei et al., 2003). No mundo moderno em que bancos de dados gigantescos são diariamente construídos com o uso diário da internet por usuários do mundo todo, modelos que possam lidar com isso estão se popularizando.

2.3.1 MALLET

O Mallet é um pacote de Java que ajuda em trabalhos que manipulam dados como os usados pelo topic modeling, text mining e outros manejos de dados. Neste trabalho foi utilizado para auxiliar na obtenção de dados a técnica LDA (LIU et al., 2016)

Após a tokenização dos dados, os vetores do documento foram calculados com auxílio do Mallet. Os dados obtidos puderam então ser transcritos para o Excel, a fim de possibilitar uma melhor visualização e a produção das imagens que se encontram neste trabalho. Na sequência, os tópicos foram manualmente tratados e categorizados. O número de tópicos foi encontrado a partir de diversos testes com o valor k, conforme o algoritmo de Gibbs (IBRAHIM; WANG, 2019).

2.4 NETFLIX

Netflix foi fundada em 1997, e hoje conta com mais de 195 milhões de assinantes, segundo eMarketer (2019), tendo causado grandes mudanças no setor televisivo ao trazer uma alternativa mais interativa que os modelos tradicionais de televisão aberta e a cabo.

Em 2007, seu serviço de *streaming* online foi extremamente disruptivo para a indústria televisiva (NOVER, 2017), pois possibilitou acesso fácil a qualquer aparelho ligado à internet, sendo ele smartphone, tablet, Smart TV ou TV regular com plugins de rede. Pode-se pagar a assinatura e assistir aos programas na hora que o assinante desejar e com conteúdo recomendado especialmente para ele, através de algoritmos desenvolvidos pela empresa para este fim. Principalmente, levando-se em conta o fato que, segundo o eMarketer (2019), existem 3,2 bilhões de usuários de smartphones no mundo no ano de 2019, contra 1,67 bilhões de casas com TV em 2018, os negócios na área de streamings estão cada vez mais atraindo empresas e consumidores.

Essa nova indústria de assinatura de *streaming* então acabou por mobilizar todas as grandes empresas dos tradicionais ramos de televisão, como a Disney e HBO (que lançaram em 2019 os serviços Disney+ e HBO Now, respectivamente). Assim como empresas não tradicionais, como a Apple e a Amazon (com a Apple TV+ e o Amazon Prime). Essa movimentação do mercado acabou gerando o movimento conhecido como "*cord cutters*" — pessoas que estão se desfazendo da TV a cabo ou nunca a tiveram — que têm muito rapidamente substituído a TV aberta e a cabo, especialmente o público mais jovem, na faixa de 25 a 34 anos (BARR, 2011).

Um dos problemas que Barr (2011) cita é a dificuldade de economistas, administradores e consultores externos terem dimensão dos efeitos do Netflix sobre o ramo, por falta de conhecimento. Nesse contexto as ferramentas de text mining podem brilhar, pois permitem a obtenção e a visualização de dados e padrões que de outra forma seriam difíceis, alguns até mesmo impossíveis, sem acesso às informações internas da provedora de filmes e séries via streaming.

2.5 TWITTER

O Twitter é uma rede social fundada em 2006 (SMITH; FISCHER; YONGJIAN, 2012), que se caracteriza por ter sua comunicação sobretudo textual e com limite de caracteres, o que faz com que os usuários abordem assuntos de maneira mais condensada (IBRAHIM; WANG, 2019). A vantagem da utilização de topic modeling no tweets é que eles possuem uma natureza mais textual e o API do próprio Twitter é excelente para fazer estudos.

Por funcionar como uma espécie de micro blog para divulgação de informações, opiniões, ideias, criações próprias, reclamações, e pequenas coisas que aconteceram no dia-a-dia; e, por trabalhar com cada publicação tendo um número limitado de caracteres, o Twitter possibilita a obtenção de informações sobre diversos temas e opiniões por meio do tratamento de poucas palavras (ATEFEH; KHREICH, 2015)

O diferencial do Twitter em relação a outras redes sociais é a conexão do usuário com a plataforma. Enquanto que em redes sociais tradicionais o usuário se comunica com comunidades e pessoas próximas, no Twitter existe o retweet, o compartilhamento de um tweet, basicamente uma pessoa compartilha algo já publicado levando a informação para novas redes dentro da rede social. No estudo de Recuero e Zago (dez. de 2009) se mostra que 62,2% dos tweets coletados eram conteúdo informativo, 54,1% eram conteúdos informativos mas que tinham informações complementares e 25,3% dos tweets eram opinativos.

Portanto, a cada retweet e comentário, mais informações e opiniões são acrescidas ao tweet original. O que leva opiniões muito além dos círculos que divulgaram tal opinião originalmente. Assim hashtags, que são palavras chaves com # na frente para frisar um tema, podem criar trendings, que são hashtags populares que aparecem em rank para todos os usuários, e essas trendings geram visualização e exposição para o assunto tratado (BENEVENUTO; MAGN; RODRIGUES; ALMEIDA, July 13-14, 2010).

2.6 ADMINISTRAÇÃO PÚBLICA

A administração pública também tem se beneficiado das novas ferramentas de análise de dados das redes sociais. Afinal, as redes sociais representam um canal aberto de comunicação entre instituições públicas e a população. É uma maneira de ser transparente, divulgar informações pertinentes e ter um feedback direto da população. (ALVES; PINHEIRO; LEMOS, 2014).

Essa posição é reforçada legalmente pela publicação de atos normativos, como a Portaria nº 59, de 1º de fevereiro de 2019, a qual prevê no Art. 4º: "É responsabilidade de todos que trabalham no Ministério Público Federal zelar pela boa imagem da Instituição, inclusive nas redes sociais, e cuidar para que os processos de comunicação social se realizem conforme os objetivos institucionais." (BRASIL, 2019). A portaria dita que as instituições devem fazer divulgações, monitoramentos e gestão das redes sociais. Reforçando a ideia da importância das redes sociais com a frase: "deverá permitir o acesso dos usuários às redes sociais, como instrumento importante de aproximação com o cidadão e ferramenta de divulgação institucional."

Cabe destacar que existe lei que estabelece a obrigatoriedade de divulgação de informações de interesse coletivo nos meios virtuais:

> A Lei 12.527 estabelece que órgãos e entidades públicas devem divulgar informações de interesse coletivo, salvo aquelas cuja confidencialidade esteja prevista no texto legal. Isto deverá ser feito através de todos os meios disponíveis e obrigatoriamente em sítios da internet. Entre as informações a serem disponibilizadas estão:
>
> -Endereços e telefones das unidades e horários de atendimento ao público.
>
> -Dados gerais para acompanhamentode programas, ações, projetos e obras.
>
> -Respostas a perguntas mais frequentes da sociedade. (CONTROLADORIA-GERAL DA UNIÃO, 2011, p. 15)

Twitter	@Min_Agricultura	@embrapa	@mmeioambiente	@brasil_Ibama
Site	www.agricultura. gov.br	www.embrapa.br	www.mma.gov.br	www.ibama.gov.br
Perfil	Perfil oficial do Ministério da Agri-cultura Pecuária e Abastecimento do Brasil.	Twitter oficial. Nossa missão é viabilizar soluções de pesquisa, desenvolvimento e inovação para a sustentabilidade da agricultura.	Perfil oficial do Ministério do Meio Ambiente do Brasil.	Ibama – Instituto Brasileiro do Meio Ambiente e dos Recursos Naturais Renováveis.
Seguidores	7.212	747	11.166	27.133
Seguindo	97	28	147	13
Tweets entre 10 e 20/06	182	64	284	21

Tabela 1 - Dados referentes aos perfis das empresas e coletados em 21 jun. 2012.
Fonte: Nascimento (2013)

Conforme constatou Nascimento (2013), e mostrado aqui na Tabela 1, alguns órgãos públicos já garantiram presença nas redes sociais e possuem um bom número de seguidores, como nas quatro instituições públicas que ela pesquisou: Ministério da Agricultura, Embrapa, Ministério do Meio Ambiente e Ibama. O problema constatado foi a falta de interação. Em 10 dias de pesquisa, ela encontrou 551 tweets, sendo 450 de divulgação institucional, 95 retweets e 6 respostas. Em um total de cerca de 13,8 tweets por instituição por dia. Alves, Pinheiro e Lemos (2014) apuraram que dos 26 órgãos do Poder Judiciário trabalhista brasileiro estudados, 24 tinham Twitter. Foram publicados 288 tweets nos 4 dias pesquisados por ele. O que totaliza uma média diária de 3 tweets por instituição.

Um outro estudo interessante é o realizado sobre a Controladoria-Geral da União (CGU) feita por Farranha e Santos (2015) que mostra que em maio de 2015 a página do CGU tinha 218 mil seguidores e foram feitos 48 postagens, ou 1,6 por dia, com uma média de 538 curtidas. Se reforça aqui que existe público para redes sociais da administração pública, o problema no caso é a necessidade de manutenção constante dessas redes.

Ferramentas de topic modeling poderiam ser usadas para melhorar o entendimento das necessidades e anseios dos cidadãos, o que poderia levar a uma comunicação mais eficiente e interativa e alavancar a avaliação positiva das instituições e dos próprios administradores públicos. Essa comunicação, como citado acima, já faz parte das leis e discussões sobre acesso a informação. O Twitter e outras redes sociais dessas instituições podem ser bem mais do que meras ferramentas rotineiras de publicação de links e anúncios como falaram Silva e Rocha (2011).

3 METODOLOGIA

O primeiro passo para realizar esse trabalho foi a aquisição dos dados, no caso os tweets postados no Netflix Brasil. Para isso foi utilizado o próprio API (Application programming interface) do Twitter, usando o programa R como ferramenta para obtenção dos dados. Ao todo 2000 tweets foram coletados entre 05 e 13 de Novembro de 2019. Cabe destacar que o API não permite o processamento de mais que 200 tweets por vez.

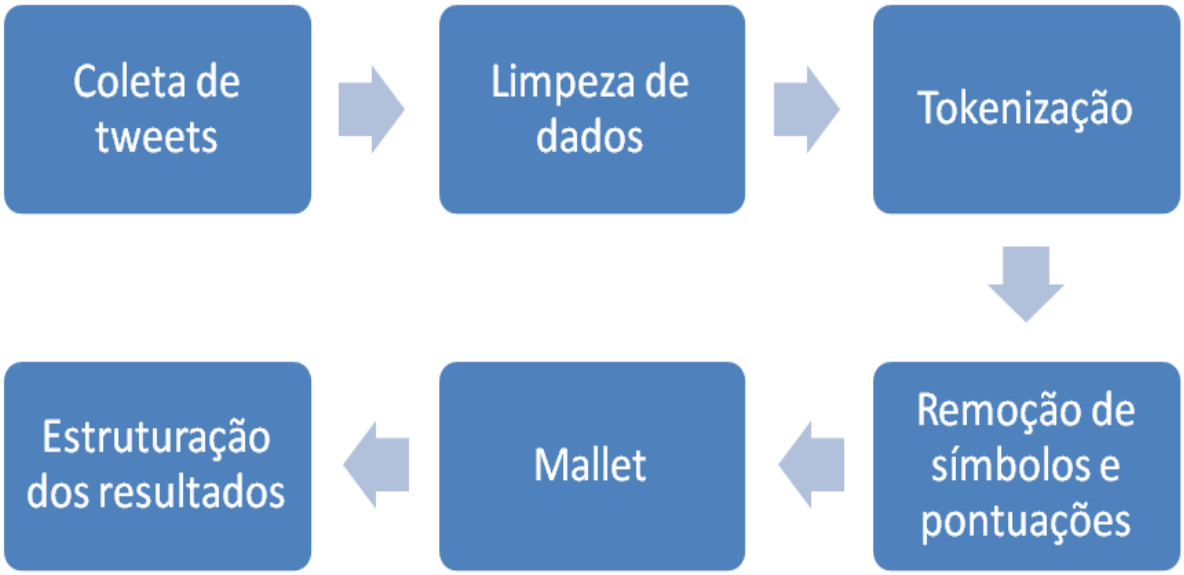


Figura 3 - Visualização geral do manuseamento de dados.
Fonte: autoria própria.

O processamento dos dados foi ilustrado na Figura 3. De maneira geral após a aquisição dos dados eles tiveram de ser "limpos", retirando-se a identificação das pessoas, tweets duplicados, bloqueados ou sem textos. Nessa etapa foram retirados pontuações, símbolos e números. Os tweets então foram tokenizados, ou seja, cada palavra da frase foi devidamente separada de forma a facilitar seu processamento pelo software. Para isso foi utilizado o Word,

que possibilitou a limpeza e a tokenização desses dados. Foi criado ainda um macro do Word para verificação do número de vezes que cada palavra apareceu no banco de dados deste trabalho. Essa limpeza levou à retirada de alguns tweets, tendo restado como amostra do presente trabalho um total de 1944 tweets, em processo similar ao trabalho de Ibrahim e Wang (2019).

Uma segunda limpeza ocorre para formatar todas as palavras em letras minúsculas para evitar divergências por conta de letra maiúscula, além de uma segunda varredura para se certificar que todas as pontuações, sinais, números, espaços em branco ou nome privado estão presentes. O texto sendo modificado como mostra a Figura 4 de maneira geral.

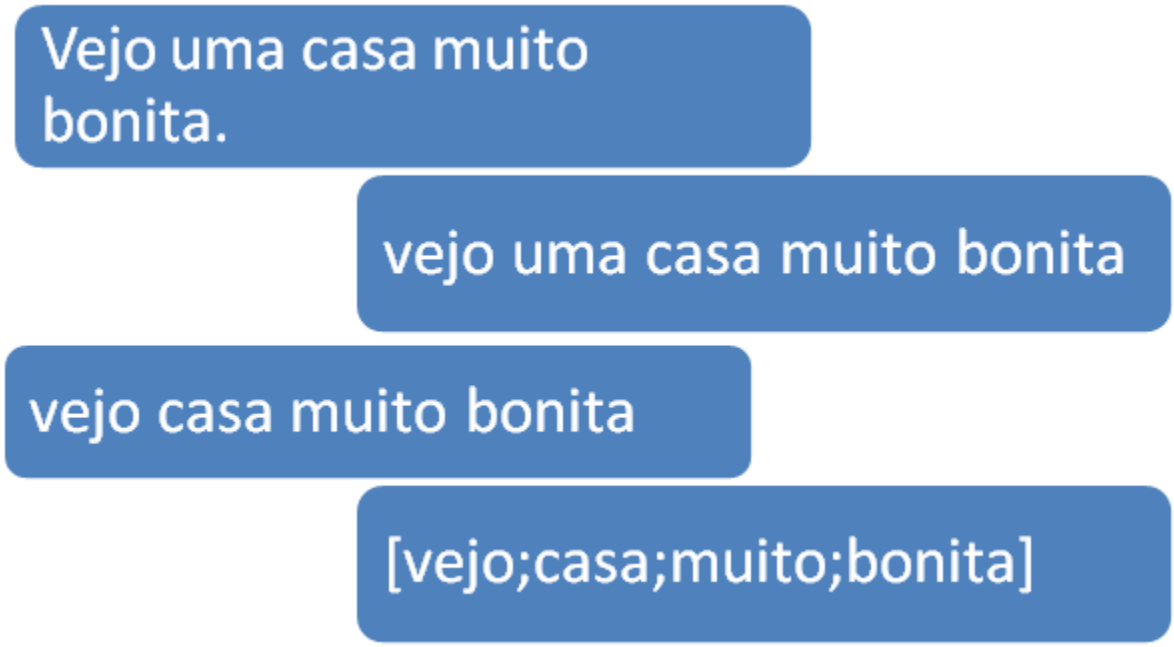


Figura 4 - Visualização da tokenização das frases.
Fonte: autoria própria.

Neste ponto se processa os dados pelo Mallet para se obter o melhor número de tópicos baseado no banco de dados. O software divulga dois documentos: um que apresenta o peso percentual de cada tópico em relação ao banco de dados e o outro com as palavras chaves de cada tópico (GRAHAM; WEINGART; MILLIGAN, 2012). Essas informações podem ser então organizadas e unidas em planilhas de Excel para melhor visualização. As informações estão nesta parte estruturadas e prontas para a análise.

4 RESULTADOS

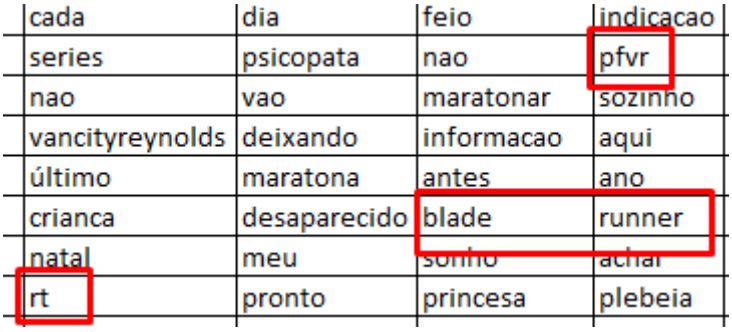
Uma das coisas que logo no processo de limpeza de dados se mostrou interessante foi a constatação de que havia uma diversidade imensa de assuntos publicados pelos usuários do Twitter. No trabalho de Ibrahim e Wang (2019), mesmo trabalhando com uma quantidade maior de tweets, os assuntos eram muito mais homogeneizados. Anderson et al. (2010) descreve o Netflix como o maior serviço de aluguel de filmes e séries do mundo, podendo-se então especular que o grande número de diferentes conteúdos tenha gerado um maior número de assuntos nos tweets pesquisados.

Peso	Nº Tópico	Palavras									
2.68%	34	gabriel	chega	logo	blackmirrorbr	mais	esperado	vindo	perfil	chamada	black
2.65%	11	visto	boca	dela	eu	qual	justica	chega	minha	assim	espero
2.55%	14	estao	muitos	conexoes	aquele	pessoas	irmandade	faz	todos	bem	ver
2.54%	33	ver	seu	pouco	bruxao	ser	uma	pessoa	normal	nao	muito
2.42%	51	eu	netflixit	vancityreynolds	um	grande	sonho	meu	olhar	meu	ta
2.39%	29	ligar	video	claro	ainda	ele	sonhou	smartphones	forma	popularizada	telefone
2.34%	62	mim	uma	agonia	vejo	eles	destruidor	waffle	tem	formato	poster
2.29%	46	voces	rapazes	pensando	maior	emocao	natal	seria	toma	uva	passa
2.27%	38	pausa	harrison	ford	sim	hum	coisa	estranho	hoje	alegria	voce
2.14%	56	isto	acontece	mim	quase	cada	dia	feio	indicacao	feio	apontar
2.01%	63	credo	maio	gostar	isto	series	psicopata	nao	pfvr	bem	lembrar
1.96%	37	eu	quer	estes	series	nao	vao	maratonar	sozinho	seja	bonita
1.93%	17	ser	chegando	pessoas	netflixit	vancityreynolds	deixando	informacao	aqui	como	onde
1.90%	12	pennbadgley	younetflix	gosto	garantia	último	maratona	antes	ano	acabar	seriesbrasil
1.90%	27	ontem	bem	dia	nao	crianca	desaparecido	blade	runner	classico	ficcao
1.88%	0	nosso	voce	odeia	muito	natal	meu	sonho	achar	mostarda	pensou
1.86%	61	macacao	vermelho	mascara	ali	rt	pronto	princesa	plebeia	nao	sabem
1.84%	52	sim	tem	também	netflixit	vancityreynolds	tal	redencao	coracao	pequeno	natal
1.81%	68	todos	qual	will	passou	tem	coragem	dizer	ele	garra	lutar
1.73%	13	veio	voando	ja	aqui	todos	prontinha	minha	coroa	prometi	mimo

Tabela 2 - Relação de tópicos por número do tópico, seus pesos e as palavras chaves do tema.
Fonte: autoria própria.

Neste trabalho, usando o LDA, foram encontrados 70 tópicos. Os 20 tópicos com maior peso, que somam cerca de 43% dos assuntos segundo o Mallet, foram postos na Tabela 2 e, em cada um desses tópicos, as 10 palavras de maior relevância foram transcritas, assim como o peso relativo do assunto em relação ao total geral (IBRAHIM; WANG, 2019).

Após a priorização das 10 palavras com mais peso de cada tópico pode-se então tentar uma demarcação dos seriados, filmes ou assunto geral de cada tópico, cada tópico podendo então mostrar seu peso na mente dos assinantes e quais as palavras foram mais associadas ao tópico.



cada	dia	feio	indicacao
series	psicopata	nao	pfvr
nao	vao	maratonar	sozinho
vancityreynolds	deixando	informacao	aqui
último	maratona	antes	ano
crianca	desaparecido	blade	runner
natal	meu	sonho	achar
rt	pronto	princesa	plebeia

Figura 5 - Ampliação da Tabela 2 para ilustrar as dificuldades do estudo.
Fonte: autoria própria.

Na Figura 5 foi ampliado um pedaço da Tabela 2 para ilustrar uma das dificuldades de estudo de redes sociais: o uso de gírias, que podem dificultar o entendimento tanto do pesquisados quanto da máquina que analisa os dado; o uso de palavras estrangeiras, que são usadas em certas situações e que o pesquisador tem de achar a melhor forma de trabalhar com elas; e por fim, a utilização de expressões de cada rede social, no caso a palavra "rt" que significa retweet, ou seja, quando se compartilha um tweet feito por outro (BOYD; GOLDER; LOTAN, 2010).

Nº Tópico	Tema Geral
34	Black Mirror (Seriado)
11	Breaking Bad (Seriado)
14	Irmandade (Seriado)
33	The Witcher (Seriado)
51	Ryan Reyonold (Ator)
29	Stranger Things (Seriado)
62	Sex Education (Seriado)
46	Natal (Feriado)
38	Harrison Ford
56	Harry Potter
63	Interações entre usuarios
37	Interações entre usuarios
17	Ryan Reyonold (Ator)
12	Penn Badgley (Ator)
27	Blade Runner (Filme)
0	Natal (Feriado)
61	La Casa de Papel
52	Ryan Reyonold (Ator)
68	Stranger Things (Seriado)
13	Interações entre usuarios

Tabela 3 - Temas associados a cada um dos 10 tópicos com maior peso.
Fonte: autoria própria.

A Tabela 3 mostra os temas associados com cada tópico. Embora algumas análises sejam intuitivas como a do tópico 34 que explicitamente fala sobre Black Mirror e o personagem Gabriel, ou mesmo o tópico 11 que tem como palavras de maior peso número 14, 15 e 16 as palavras -ozark, breaking, bad. Ozark sendo uma série mais recente que é bastante comparada com a série de sucesso Breaking Bad (RAMOS, 2017).

Outros tópicos como o 68 fazem alusão a séries através do nome do personagem, no caso Will. Polêmicas sobre a série como a palavra waffle da série Sex Education (SMYTHE, 2020), a qual se refere a uma doença venérea e que motivou diversas postagens. Ou alusões gerais como o macacão vermelho e máscara presentes no seriado La Casa de Papel (ALVES, 2019).

Houve também tópicos falando sobre atores como Ryan Reyonold, que está em 3 tópicos provavelmente por suas aparições em La Casa de Papel, Detetive Pikachu e Deadpool 2. Além de ser um ator com alto engajamento no twitter, tendo grande interatividade com seus fans na

rede, o que lhe gerou inclusive o apelido de Rei do Twitter, como consta no artigo de Iveta (2017).

Por fim, nota-se nos tópicos 63, 37 e 13 interações diversas dos usuários principalmente falando sobre maratonas de filmes e séries, bem como diversas alusões ao feriado de natal. Lembrando que a coleta de dados ocorreu em Novembro de 2019, o que faz o tema natalino ser pertinente. Sendo isso também uma boa leitura pois mostra engajamento dos usuários para além dos serviços e produtos da Netflix, o que realça a diferença do papel dos usuários das redes sociais. Os usuários não são apenas observadores passivos e sim participantes ativos que criam, modificam, reusam e tornam a plataforma melhor pela soma de todas as partes (DOLAN ET AL, 2019).

Ou seja, a ferramenta permite que empresas tenham maior facilidade de entender quais são os temas, serviços ou produtos que estão sendo mais mencionados pelos usuários em determinada rede social. E também entender se essa menção é positiva ou negativa, possibilitando que as empresas possam então satisfazer melhor o consumidor, como nota Ibrahim e Wang (2019).

CONSIDERAÇÕES FINAIS

O que se constatou neste estudo foi como o LDA pode tratar uma grande quantidade de dados e se transformar em algo que sirva para facilitar a tomada de decisões por parte de administradores públicos e privados. O LDA cumpre seu papel de transformar centenas de tweets em tabelas de rápido entendimento. Interessante pensar que a grande quantidade de dados na verdade aprimora cada vez mais o LDA, já que essa técnica se caracteriza por fornecer resultados mais precisos e abrangentes quanto maior a base de dados processada.

Para quem deseja usar essa ferramenta o presente trabalho mostrou que ela é eficaz e pode ser aplicada tanto no setor público quanto no setor privado. Porém, cabe destacar, que ainda faltam materiais de treinamento e softwares na língua portuguesa. Ressalto que, devido às características do Twitter, da grande variedade de conteúdo disponível na Netflix e à forma como o Mallet processa os dados, houve uma certa quantidade de ruídos na estruturação de dados para uma melhor análise.

Esse estudo pode mostrar então como com um recolhimento de informações de redes sociais somado ao uso de ferramentas apropriadas, como no caso do software Mallet que foi utilizado aqui, é possível obter um melhor perfil do consumidor ou usuário. O nível de entrada para um pesquisador conseguir obter e trabalhar grandes bancos de dados digitais se torna cada mais fácil através: da facilidade de obter dados públicos em grande quantidade através de plataformas e redes; e ferramentas, teorias e modelos que tornam a manipulação desses dados cada vez mais refinada e eficiente.

Por fim, destaco algumas sugestões para futuros trabalhos com base no tema apresentado: pesquisas mais amplas que possam relacionar empresas públicas brasileiras com grande participação nas redes sociais, fazendo a comparação, por exemplo, com empresas privadas do mesmo segmento, visando a extração de dados comparativos. Sugiro também pesquisas com Twitters de agências ou órgãos de governos que possam mostrar que essa ferramenta, por seu caráter de tentar entender o público mais diretamente, não faz parte apenas do repertório de administradores privados, mas pode sim ser utilizado também por administradores públicos.

REFERÊNCIAS BIBLIOGRÁFICAS

ALVES, Camila Magalhães; PINHEIRO, Hugo Cardim; LEMOS, Daniel Dantas. **A comunicação da Justiça do Trabalho em redes sociais digitais: uma análise da presença do judiciário trabalhista brasileiro no Facebook, Twitter e Youtube.** Fortaleza: Intercom, 2014.

ALVES, Paulo. **Melhores filmes da Netflix em 2019: veja os títulos mais populares do ano**. TechTudo, 2019. Disponível em: https://www.techtudo.com.br/noticias/2019/12/melhores-filmes-da-netflix-em-2019-veja-os-titulos-mais-populares-do-ano.ghtml. Acesso em: 30 nov. 2020.

ANDERSON, Matt et al. **The Rise of Social Apponomics How Social Media and Apps Are Transforming E-Commerce**. S. L: Booz & Company Inc, 2010.

ATEFEH, Farzindar; KHREICH, Wael. **A SURVEY OF TECHNIQUES FOR EVENT DETECTION IN TWITTER.** Montrea: Wiley Periodicals, v. 31, n. 1, 2015.

BARR, Trevor. **TELEVISION'S NEWCOMERS: NETFLIX, APPLE,GOOGLE AND FACEBOOK.** Melbourne: Telecommunications Journal Of Australia, v. 61, pp. 60.1-60.10, nov 2011.

BENEVENUTO, Fabrício; MAGN, Gabriel; RODRIGUES, Tiago; ALMEIDA, Virgílio. **Detecting Spammers on Twitter**. Redmond: Ceas 2010 - Seventh Annual Collaboration, Electronic Messaging, Anti-Abuse And Spam Conference, 13-142010.

BERTHON, Pierre R. et al. **Marketing meets Web 2.0, social media, and creative consumers: Implications for international marketing strategy**. Indiana: Elsevier, 2012.

BLEI, David M.; NG, Andrew Y.; JORDAN, Michael I. **Latent Dirichlet Allocation**. Journal of machine Learning research 3. pp. 993-1022, 2003.

BOYD, Danah; GOLDER, Scott; LOTAN, Gilad. **Tweet, Tweet, Retweet: Conversational Aspects of Retweeting on Twitter**. 43. ed. Honolulu: Ieee, 2010. (2010 43rd Hawaii International Conference on System Sciences).

BRASIL. Ministério Público da União. Atos da Procuradora-Geral da República. Portaria nº 59, de 1 de fevereiro de 2019. Altera a Portaria PGR/MPF nº 918/2013, que dispõe sobre a Política Nacional de Comunicação Social do Ministério Público Federal. **Diário Oficial da União**, Brasília, DF, 8 abr. 2019. p. 59.

CAMPR, Michael; JEZEK, Karel. **Challenges Comparative Summarization via Latent Dirichlet Allocation**. Plzen: Dateso pp. 80-86, 2013.

CIRQUEIRA, Douglas et al. **Improving Relationship Management in Universities with Sentiment Analysis and Topic Modeling of Social Media Channels: Learnings from UFPA.** Leipzig: Wi 2017. 2017 IEEE/WIC/ACM International Conference on Web Intelligence. pp. 998-1005., 2017

CONTROLADORIA-GERAL DA UNIÃO. **Acesso à Informação Pública: Controladoria-Geral da União: Uma introdução à Lei nº 12.527, de 18 de novembro de 2011**. Brasília: Controladoria-Geral da União, 2011.

DIJKSTRA, Suzan *et al*. **Possibilities and Pitfalls of Social Media for Translational Medicine**. Frontiers In Medicine, 2018. 345 p. 5 v.

DOLAN, Rebecca, CONDUIT, Jodie, FRETHEY-BENTHAM, Catherine, FAHY, John and GOODMAN, Steve. **"Social media engagement behavior: A framework for engaging customers through social media content"**, European Journal of Marketing, Vol. 53 No. 10, pp. 2213-2243, out. 2019.

eMarketer. (n.d.). **Leading retail and consumer merchandise brands on Twitter as of July 2018, by followers (in millions).** [s.L.]: Statista, 2018. Disponível em: <http://www.statista.com/statistics/281370/most-popular-retailers-on-twitter-ranked-by-number-of-followers/>. Acesso em: 01 dez. 2019.

eMarketer. (n.d.). **Number of Netflix paid subscribers worldwide from 3rd quarter 2011 to 3rd quarter 2020.** [s. L.]: Statista, 2020. Disponível em: < https://www.statista.com/statistics/250934/quarterly-number-of-netflix-streaming-subscribers-worldwide/>. Acesso em: 30 nov.2020.

eMarketer. (n.d.). **Number of smartphone users worldwide from 2016 to 2021 (in billions).** [s. L.]: Statista, 2019. Disponível em: < https://www.statista.com/statistics/330695/number-of-smartphone-users-worldwide/>. Acesso em: 29 nov.2019.

eMarketer. (n.d.). **Number of social media users worldwide from 2010 to 2021 (in billions)**. [s. L.]: Statista, 2019. Disponível em: < https://www.statista.com/statistics/278414/number-of-worldwide-social-network-users/>. Acesso em: 29 nov.2019.

eMarketer. (n.d.). **Number of TV households worldwide from 2010 to 2018 (in billions).** [s. L.]: Statista, 2019. Disponível em: < https://www.statista.com/statistics/268695/number-of-tv-households-worldwide/>. Acesso em: 29 nov.2019.

FARRANHA, Ana Cláudia; SANTOS, Leonardo Tadeu dos. **ADMINISTRAÇÃO PÚBLICA, DIREITO E REDES SOCIAIS: O CASO DA CGU NO FACEBOOK**. Revista Eletrônica do Curso de Direito da Ufsm, 10 v. n. 2, 2015.

GRAHAM, Shawn; WEINGART, Scott; MILLIGAN, Ian. **Getting Started with Topic Modeling and MALLET**. The Programming Historian 1, 2012.

IBRAHIM, NF; WANG, Xiaojun. **A text analytics approach for online retailing service improvement: Evidence from Twitter.** Decision Support Systems. v. 121, pp. 37-50, 2019.

IVETA. **93 Times Ryan Reynolds Was The King Of Twitter**. Boredpanda, 2017. Disponível em: https://www.boredpanda.com/polite-ryan-reynolds-twitter-replies/. Acesso em: 30 nov. 2020.

KAPLAN, Andreas; HAENLEIN, Michael. **Users of the World, Unite! The Challenges and Opportunities of Social Media.** Business Horizons.vol. 53, pp 59-68, 2010.

LEEFLANG, PSH et al. **Challenges and solutions for marketing in a digital era**. European Management Journal, vol. 32, no. 1, pp. 1-12, 2014.

LIU, Cindy. **US DIGITAL USERS:** The eMarketer Forecast for 2016. S. L: Emarketer, 2016.

LIU, Lin et al. **An overview of topic modeling and its current applications in bioinformatics.** Yunnan: Springer Open, 2016.

NELSON, Nick. **Tapping Key Takeaways from Recent Research on Fortune 500 Social Media Usage**. Minneapolis: Topranking Marketing, 2018. Disponível em: https://www.toprankblog.com/2018/11/fortune-500-social-media-adoption/. Acesso em: 30 nov. 2020.

NOVER, Scott. **Netflix vs. The World:A Study of Competitive Trends in the Modern American TV Industry.** Ann Arbor: Proquest, 2017.

RAMOS, Guilherme. **"Ozark", da Netflix, é o irmão caçula de "Breaking Bad"**. New Order, 2017. Disponível em: https://medium.com/neworder/ozark-da-netflix-e-o-irmao-cacula-de-breaking-bad-715b408f8e1e. Acesso em: 30 nov. 2020.

RECUERO, Raquel; ZAGO, Gabriela. **Em busca das "redes que importam": redes sociais e capital social no Twitter**. São Paulo: Líbero, 2009. 12 v. (N. 24, p. 81-94).

SILVA, Danilo Morais da; RIBEIRO, Ana Claudia Dias; SILVA FILHO, Esiomar Andrade. **AS REDES SOCIAIS COMO FERRAMENTA PARA ACESSO À INFORMAÇÃO NA ADMINISTRAÇÃO PÚBLICA**. Belo Horizonte: Perspectivas em Políticas Públicas, 2018. v. XI, nº 21, p. 267-294.

SILVA, Maurílio Luiz Hoffmann da; ROCHA, Liana Vidigal. **TWITTER E CIBERCULTURA: UM ESTUDO SOBRE AS FUNCIONALIDADES DA FERRAMENTA DE COMUNICAÇÃO.** S. L: Intercom, 2011.

SMITH, Andrew N.; FISCHER, Eileen; YONGJIAN, Chen. **How Does Brand-related User-generated Content Differ across YouTube, Facebook, and Twitter?** Ontario: Elsevier, 2012.

SMYTHE, Polly. **'It answered my weird questions': what do teens really think of Netflix's Sex Education?** The Guardian, 2020. Disponível em: https://www.theguardian.com/tv-and-radio/2020/jan/22/it-answered-my-weird-sex-questions-what-teens-really-think-of-sex-education-netflix. Acesso em: 01 dez. 2020.

SOUZA, Karine França de; PEREIRA, Moisés Henrique Ramos; DALIP, Daniel Hasan. **Unilex:Método léxico para análise de sentimentos textuais sobre conteúdo de tweets em português brasileiro.** Belo Horizonte: Abakós, 2017.

WAAL, Alta de; MOUTON, Francois. **Topic modelling in the information warfare domain.** Pretoria: Ieee, 2013.

WAYNE, Michael L.. **Global streaminging platforms and national pay-television markets: a case study of Netflix and multi-channel providers in Israel. Rotterdam: Department Of Media And Communication**, 2019.

Weber Shandwick. **The Social CEO: Executives Tell All.** [s. L.]: Weber Shandwick, 2013.

ZHAO, Wayne Xin, et al. **Comparing twitter and traditional media using topic models." European conference on information retrieval.** Springer, Berlin, Heidelberg, 2011

ZHAO, Weizhong et al. **A heuristic approach to determine an appropriate number of topics in topic modeling.** Little Rock: Bmc Bioinformatics v. 16, S8, 2015.