

REGULAÇÃO E REMOÇÃO DE CONTEÚDOS E A INFLUÊNCIA DAS REDES SOCIAIS SOBRE AS CONCEPÇÕES POLÍTICAS¹

CONTENT REGULATION AND REMOVAL AND HOW IT INFLUENCES POLITICAL VIEWS IN SOCIAL MEDIA

Daniel Jorge Teixeira Cesar²

Resumo: O Facebook possui um conjunto de políticas e regulamentos internos definidos pelos responsáveis pela moderação de conteúdos na plataforma. Os Padrões da Comunidade, conjunto de regras da empresa que determinam o que pode ser postado na rede, colocam limites ao usuário para que não divulgue conteúdos proibidos como violência explícita, notícias falsas e discurso de ódio. Este trabalho analisa a aplicação destas regras a partir do estudo de documentos internos do Facebook obtidos pelo jornal inglês *The Guardian*, notícias publicadas sobre conteúdos removidos e depoimentos coletados em entrevistas com moderadores da empresa para examinar como se dá o processo de remoção de conteúdos. Os resultados da observação apontam que as políticas de regulação e o trabalho dos moderadores, além de insuficientes, são falhos em mais de um aspecto, com brechas para publicação de conteúdos preconceituosos e postagens que disrompem a esfera pública.

Palavras-Chave: Mídias Sociais. Esfera Pública. Regulação de Conteúdos.

Abstract: Facebook has its own set of policies and internal regulations defined by those responsible for moderating content on the platform. The Community Standards, Facebook's set of rules that determine what can be posted online, impose limits to the user on forbidden content such as explicit violence, fake news and hate speech. This paper analyses how these rules apply by studying Facebook's internal documents obtained by british newspaper *The Guardian*, published news on banned content and statements collected on interviews with the Facebook's moderators to examine how contents are removed. Observation results point that regulation policies and the work of moderators are insufficient and flawed in more than one aspect, with loopholes for publication of bigoted content and posts that disrupt the public sphere.

Keywords: Social Media. Public Sphere. Content Regulation.

¹ Trabalho apresentado ao Grupo de Trabalho Cultura Política, Comportamento e opinião pública do VIII Congresso da Associação Brasileira de Pesquisadores em Comunicação e Política (VIII COMPOLÍTICA), realizado na Universidade de Brasília (UnB), de 15 a 17 de maio de 2019.

² Doutorando do Programa de Pós-Graduação em Comunicação da Universidade de Brasília, danieljtc@gmail.com

1. Introdução

A literatura sobre tecnologias de informação e comunicação aponta que as mídias sociais online são espaços democráticos e informativos de construção de debates culturais e políticos em que é possível obter perspectivas e opiniões diferentes sobre um mesmo assunto e ganharam, nos anos recentes, um papel cada vez maior na formação de conceitos e influencia na composição da esfera pública (Lemos, 2004). Há um enorme volume de informação, visto que a estrutura e capacidade das mídias permite que mais pessoas possam ser ouvidas e divulgar ideias (Zamith & Lewis, 2014), mas apesar disso há problemas inerentes às estruturas das redes que, pelo modo como são construídas, estimulam o consumidor a buscar conteúdos e se informar apenas com o que concorde com sua tendência política e seu pensamento de maneira geral (Flaxman, Goel & Rao, 2016).

A informação não é democrática e não chega a todos da mesma maneira. As mídias sociais utilizam algoritmos para modular as informações acessadas pelos usuários. Além disso, as redes contém notícias falsas e constróem narrativas que não dialogam com a realidade objetiva, como os movimentos anti-vacinação e de negação do holocausto na segunda guerra mundial, por exemplo. A qualidade da informação é essencial para a esfera pública, e as modulações e desigualdades nos conteúdos que chegam aos usuários de mídias sociais influenciam a opinião pública pelo que é postado nas redes, que possuem algoritmos e políticas para determinar a curadoria de conteúdo e o modo como são filtrados segundo os interesses do usuário e formam bolhas (Pariser, 2011).

A informação que nos chega já é filtrada pelas empresas a partir do que querem promover para cada indivíduo conectado, mas também a partir dos interesses pessoais. Por um lado há uma curadoria algorítmica que produz câmaras de eco, enquanto na mesma medida há uma seleção por parte do indivíduo do que quer acessar (Spohr, 2017). A modulação de conteúdo descrita por Pariser e Spohr considera as sugestões de conteúdos baseados nos interesses individuais e ocorre a partir das informações oferecidas pelo usuário, mas além do algoritmo que sugere conteúdos há outros modos que as redes utilizam para promover ou remover postagens. Com o cenário de polarização ideológica que se desenha nas redes a partir da informação disponível nelas e tendo por base o Facebook, é possível

afirmar que há problemas na aplicação das políticas definidas pela empresa para regulação de conteúdos na plataforma.

O trabalho a seguir aborda as remoções de conteúdo realizadas pelo Facebook a partir do que está descrito nos Padrões da Comunidade³, documento que determina o que pode ser publicado na plataforma e permanecer na rede, das práticas de remoção dos moderadores e algoritmos que detectam e retiram da rede conteúdos proibidos, e como estes filtros contribuem para a formação de bolhas entre grupos de usuários da plataforma.

Como procedimento de pesquisa, além da revisão bibliográfica sobre o tema, foi realizada pesquisa com documentos internos do Facebook obtidos e publicados pelo jornal The Guardian em 2017⁴. Para examinar a aplicação dos Padrões da Comunidade e o trabalho dos moderadores de conteúdo, foram coletadas notícias que apontam os problemas no funcionamento da moderação da rede social considerando os tipos de conteúdos removidos ou mantidos pela plataforma.

O que se procura é analisar sob que condições ocorre a remoção de conteúdos utilizando pesquisa documental para levantar dados sobre páginas ou postagens e de que modo a circulação de informações como notícias falsas e postagens preconceituosas contribuem ativamente para alimentar e formar bolhas, seja pela leitura feita pelo algoritmo de controle das publicações ou pela intervenção de moderadores humanos que trabalham de acordo com demandas do Facebook.

Em um primeiro momento, discutiremos a formação de filtros a partir do conceito de Eli Pariser e incluiremos em sua definição os mecanismos de remoção de conteúdo além dos que criam o conjunto de informações editadas e personalizadas para cada indivíduo. Tomando o autor como ponto de partida, analisaremos em seguida os documentos da empresa e o trabalho dos moderadores na remoção de conteúdos.

³ Disponível em <https://www.facebook.com/communitystandards/> Acesso em 6 de Março de 2019

⁴ Disponível em <https://www.theguardian.com/news/series/facebook-files> Acesso em 6 de Março de 2019

2. Filtros, Bolhas e a Esfera Pública

Os conceitos de filtro e bolha elaborados por Eli Pariser apresentam o funcionamento de mecanismos de curadoria dos conteúdos acessados em plataformas online para exibir conjuntos de informações editadas e personalizadas para cada indivíduo a partir de seus interesses.

Os algoritmos de recomendação de conteúdo, que operam a partir de metadados e sugerem informações semelhantes às aquelas que o usuário acessa, são uma forma de filtrar o acesso a informação e agem encerrando o indivíduo em uma bolha, uma câmara de eco que apresenta apenas aquilo que está de acordo com as opiniões e ideologias do indivíduo. Em trecho de seu livro, Pariser cita Eric Schmidt, diretor-geral do Google, que certa vez declarou que “a tecnologia será tão boa que as pessoas terão muita dificuldade em assistir ou consumir qualquer coisa que não tenha sido minimamente adaptada para elas” (Pariser, 2011, p. 47), isto é, o consumo será moldado de acordo com os gostos individuais reconhecidos pelo sistema no banco de dados.

Cada empresa tem um algoritmo próprio, uma fórmula matemática que gera os resultados encontrados a partir da combinação de termos buscados e cliques que costumam ocorrer em contextos privados, como histórico de buscas e comportamento online, mas que ficam registrados para criar indicadores de comportamentos individuais e coletivos. De acordo com o mesmo autor, o Facebook utiliza outros parâmetros de medida, como os relacionamentos interpessoais e os conteúdos compartilhados. Estes indicadores criam filtros de conteúdos e, ao mesmo tempo que personalizam a experiência online, também servem como ferramenta de vigilância a partir das informações que fornecemos.

Segundo o autor, a experiência de acessar conteúdos pela Internet é determinada pelos interesses do usuário e os conteúdos que escolhe consumir, mas se considerarmos que o filtro funciona também na exibição de conteúdos ideológicos e culturais, selecionando da mesma maneira o que é acessado pelo usuário, é possível supor que os debates na esfera pública sofrem influência da qualidade da informação pela polarização e divulgação de conteúdos falsos, por exemplo. O acesso desigual à informação complementa o conceito de filtro pelas implicações trazidas pela regulação de conteúdos por algoritmos, para promoção e remoção de

informações. A exposição a opiniões políticas e ideias tende a criar câmaras de eco (Seargeant & Tagg, 2018) que apenas reproduzem conteúdos consonantes ao pensamento do usuário.

Apesar das mídias sociais possibilitarem que todos os indivíduos conectados à Internet possam participar das discussões na esfera pública, a existência de bolhas de informação com qualidade e alcances diferentes impõe dificuldades na mobilização de um debate positivo. A desigualdade no acesso à informação, resultado dos filtros baseados em interesses e na curadoria de conteúdos da plataforma são determinadas pela própria plataforma, de maneira opaca e que deve ser questionada.

A polarização em torno de temas como a negação do holocausto, movimentos antivacinação, conteúdos que pregam violência e divulgação de notícias falsas colabora para a desinformação na composição da esfera pública. Para compreender este fenômeno em que o acesso à informação é desigual, é preciso considerar e analisar como as modulações ocorrem. Aqui serão estudadas a remoção e manutenção de conteúdos em contextos de formação de opinião a respeito de temáticas culturais, políticas e ideológicas.

3. Os Padrões da Comunidade e as Remoções de Conteúdos

O Facebook opera no Brasil desde 2008 e se consolidou como uma das principais redes sociais acessadas pela população brasileira, com cerca de 130 milhões de usuários no País, segundo dados de Janeiro de 2019. O Brasil é o terceiro maior mercado do Facebook, atrás da Índia com 300 milhões de usuários, e Estados Unidos com 210 milhões⁵. Para regular o funcionamento da plataforma, há um conjunto de termos e políticas descritos nos documentos Política de Dados, que determina como a empresa coleta e utiliza os dados dos usuários; o Termo de Serviço, que delimita os compromissos da plataforma; e os Padrões da Comunidade, que regulam que conteúdos podem ser postados na rede.

⁵ Disponível em <https://www.statista.com/statistics/268136/top-15-countries-based-on-number-of-facebook-users/> Acesso em 19 de Março de 2019

A quantidade de usuários que se engajam e compartilham conteúdos no Facebook e as estruturas que definem seu funcionamento tornam a rede objeto de estudo interessante não apenas pelo modo como se utiliza dos dados dos usuários, mas também pela maneira como as pessoas se relacionam e que trocas de informações ocorrem neste espaço de interação. O que se procura aqui é debater os Padrões da Comunidade e a aplicação destes para a remoção ou manutenção de conteúdos segundo o trabalho dos moderadores.

No documento estudado estão descritos os tipos de conteúdos proibidos na plataforma. O Facebook descreve em sua política que prioriza a liberdade de expressão, a diversidade de vozes e a manutenção de um ambiente seguro. Para isso, define que postagens que exibam formas de violência contra indivíduos ou formas de violência organizada como terrorismo, exploração sexual de adultos e pornografia infantil devem ser removidas da rede. Para a aplicação destas normas, o Facebook conta com uma equipe de moderadores e utiliza algoritmos e machine learning para detectar e apagar conteúdos não permitidos.

Os Padrões da Comunidade e as informações sobre os tipos de conteúdos proibidos na plataforma vieram a público após o vazamento de documentos internos do Facebook utilizados no treinamento e capacitação de profissionais para remoção de conteúdo e obtidos pelo jornal inglês The Guardian em 2017. Anteriormente a isso não havia conhecimento público sobre a clareza e objetividade das remoções de postagens e perfis.

As matérias sobre remoções de conteúdo no The Guardian causaram controvérsia ao expor o funcionamento do Facebook, que em resposta à sociedade aumentou o número de moderadores – de 4.500 para 30.000, e criou um documento para delimitar os objetos das remoções de conteúdos. Os Padrões da Comunidade foram publicados pela primeira vez em 2018,⁶ em resposta às cobranças da sociedade quanto a intransparência dos critérios de remoção de conteúdo – que continuam opacos pois os Padrões da Comunidade apenas listam os tipos de conteúdos proibidos, sem dar detalhes sobre as sanções e casos de aplicação da

⁶ Disponível em <https://www.digitaltrends.com/social-media/facebook-community-standards/>
Acesso em 19 de Março de 2019

norma aos usuários que postem discurso de ódio, por exemplo. O texto nos Padrões da Comunidade define discurso de ódio da seguinte maneira⁷:

“Definimos discurso de ódio como um ataque direto a pessoas com base no que chamamos de características protegidas: raça, etnia, nacionalidade, filiação religiosa, orientação sexual, casta, sexo, gênero, identidade de gênero e doença ou deficiência grave. Também oferecemos proteções para o status migratório. Definimos ataques como discursos violentos ou degradantes, declarações de inferioridade ou incentivo à exclusão ou segregação.” (FACEBOOK, 2019)

Estes documentos apontam que a empresa prioriza a defesa do indivíduo mais do que a de uma ideia ou conceito. A análise do documento que detalha as remoções de conteúdos com discurso de ódio aponta que uma postagem que pregue ódio aos cristãos ou homossexuais, por exemplo, é deletada, mas o direcionamento do ódio ao cristianismo ou a homossexualidade enquanto conceitos é tolerado.

Instituições e crenças não estão protegidas na política contra discurso de ódio, e nessa categoria estão incluídas ideologias políticas de todos os espectros. Clamores pela exclusão e segregação de migrantes – definidos no documento como uma categoria quase protegida, são toleradas enquanto não fizerem menção a violência ou degradação e desumanização do indivíduo, comparando-o com animais, por exemplo. O documento destaca ainda o que o Facebook define como categoria quase protegida, que não recebe o mesmo tratamento na análise e remoção de conteúdos relacionados à crença e ao povo. Um exemplo destacado no texto são os muçulmanos.

O Facebook define em outro documento o conceito de violência plausível como situações em que a postagem possua características no discurso que façam crer que há ameaça de dano real contra indivíduos. É utilizado especialmente para tratar de pessoas públicas como líderes de estado, ativistas, jornalistas ou uma pessoa identificada por nome, imagem ou referência. Nestes casos são tratados

⁷ Disponível em https://www.facebook.com/communitystandards/objectionable_content Acesso em 6 de Março de 2019

como público protegido e a postagem contendo a ameaça é removida. Em casos de maior risco percebido, autoridades podem ser contatadas pelo Facebook.

Publicações que incitam violência de maneira geral são removidas apenas quando há um contexto identificado de violência plausível, o que significa que postagens informativas ou de humor, que não apresentam contexto de violência crível, são permitidas, o que abre brechas para a possibilidade de reprodução de discursos de ódio sob estas formas. Postagens que glorifiquem a violência de grupos terroristas, com frases de incentivo ou preconceituosas são proibidas, mas a mesma publicação é permitida se o usuário substitui a legenda com um contexto que condena os atos ou se assumem neutralidade na postagem.

Os documentos estudados não definem as sanções aplicadas para além da remoção da postagem, como bloqueio ou remoção de perfis, nem estabelecem regras para a remoção de notícias falsas. Sabe-se que há situações em que a plataforma pode bloquear o acesso do usuário por um período de tempo determinado ou deletar a conta permanentemente de acordo com a gravidade da ofensa aos Padrões da Comunidade, ou se o algoritmo da empresa perceber algo que não deve estar no Facebook. Um dos casos em que a conta pode ser apagada é a partir da comprovação de que se trata de um perfil falso, pois o Facebook estabelece que cada usuário deve utilizar seu nome e oferecer informações reais sobre si, mas para além disso não são claras as condições para outras sanções além da remoção de publicações.

A análise dos documentos aponta também que há brechas para publicação de conteúdos proibitivos e mesmo as definições nos documentos internos sobre que conteúdos devem ser removidos são vagas. Um exemplo que ilustra essas questões é a postagem de conteúdos sobre a negação do holocausto de seis milhões de judeus pelo nazismo, que são reguladas apenas nos países em que a divulgação destas ideias é proibida por lei. Trata-se de uma desinformação e uma afirmação preconceituosa protegida pelos regulamentos do Facebook sob o manto da liberdade de expressão. Mark Zuckerberg, em entrevista realizada em 2018⁸, afirmou que não acredita que conteúdos como este deveriam estar na plataforma, mas não

⁸ Disponível em <https://www.recode.net/2018/7/18/17575158/mark-zuckerberg-facebook-interview-full-transcript-kara-swisher> Acesso em 15 de Março de 2019

concorda que sejam retirados pois poderia ferir a liberdade de cada indivíduo de expressar uma opinião, mesmo que, segundo ele, esteja errada.

Esse argumento pode abrir precedentes para divulgação de notícias falsas e conteúdos que criam polarizações políticas e ideológicas e causar danos à disrupção na esfera pública. A pluralidade e liberdade de pensamento são pilares da sociedade democrática moderna, mas é preciso considerar que há contaminação da esfera pública quando fontes de informação como o movimento anti-vacinação e aqueles que dizem que o planeta Terra é plana possuem voz e divulgam suas ideias para um volume grande de pessoas. Para compreender a extensão do problema é preciso analisar também que conteúdos o Facebook remove e como se dá o trabalho dos moderadores de conteúdo.

Definidos dos Padrões da Comunidade e os tipos de postagens proibidas, o Facebook recebe diariamente milhares de publicações contendo pornografia e violência, detectadas e avaliadas por algoritmos e revisadas por moderadores humanos que possuem uma rotina de trabalho estressante. Matéria recente publicada no site The Verge⁹ detalha o treinamento e trabalho de uma funcionária terceirizada pelo Facebook para regular conteúdos na plataforma. Os moderadores de conteúdo são expostos a todo tipo de conteúdo degradante. De formas de violência como suicídio e automutilação a abuso sexual e pornografia infantil, entre outros conteúdos sensíveis. Segundo a informante, muitos desenvolvem distúrbios e abusam de drogas no local de trabalho para lidar com a exposição e quantidade de conteúdo avaliado. O Facebook possui atualmente 30.000 funcionários – a maioria deles terceirizados, para remover milhões de posts por dia em 100 idiomas. Cada moderador pode revisar até 3.500 conteúdos por dia¹⁰.

Os resultados do trabalho realizado pelos moderadores de conteúdo estão descritos em relatório publicado pelo Facebook em novembro de 2018¹¹ e apontam que houve um aumento da proatividade da empresa em remover discurso de ódio

⁹ Disponível em <https://www.theverge.com/2019/2/25/18229714/cognizant-facebook-content-moderator-interviews-trauma-working-conditions-arizona> Acesso em 26 de Fevereiro de 2019

¹⁰ Disponível em <https://g1.globo.com/tecnologia/noticia/checava-se-alguem-se-mataria-ao-vivo-a-rotina-de-um-moderador-brasileiro-de-posts-denunciados-no-facebook.ghtml> Acesso em 19 de Março de 2019

¹¹ Disponível em <https://newsroom.fb.com/news/2018/11/enforcing-our-community-standards-2/> Acesso em 7 de Março de 2019

pelo uso de algoritmos, antes mesmo do conteúdo ser denunciado por um usuário. Dados do último trimestre de 2017 mostram que 76% dos conteúdos de discurso de ódio foram removidos a partir de denúncias, e que o número caiu para 48% no terceiro trimestre de 2018. Isso sinaliza para maior eficiência da empresa de trabalhar com algoritmos e inteligência artificial para detectar, reconhecer e remover postagens deste tipo com maior precisão. O relatório destaca também que 97% de conteúdos com violência gráfica são detectados pela plataforma antes de serem denunciados.

Outros números consolidados em relatório de maio de 2018 e levantados por Thais Dias (Dias, 2018) mostram que 21 milhões de conteúdos de pornografia adulta foram removidos da plataforma no primeiro trimestre de 2018. 96% deste conteúdo foi encontrado antes de ser denunciado. Publicações contendo discurso de ódio somaram 2,5 milhões de remoções. Deste total, 38% foram sinalizados pelos usuários. No mesmo período analisado foram derrubados 837 milhões de spams e 583 milhões de contas falsas desativadas.

A remoção de conteúdo da plataforma, apesar dos avanços demonstrados pelos números acima, tornou-se um problema para a empresa¹² pois não há pessoal suficiente para lidar com o volume de trabalho e há brechas no modo como são conduzidas as remoções de arquivos, especialmente pelos algoritmos. A programação destes pode vir a apagar conteúdos não por infringem os Padrões da Comunidade, mas em razão do volume de denúncias recebidas por uma postagem, por exemplo. Grupos podem se organizar para denunciar uma postagem de forma massiva e retirar um conteúdo do ar.

Outro aspecto que vale ser detalhado é que o Facebook não menciona outras formas de lidar com conteúdo, como a manipulação do algoritmo que compõe o feed do usuário: ao invés de remover da plataforma, dificulta o acesso a determinadas informações. Não há transparência para explicar quais os critérios para seguir essa atuação ao invés de remover conteúdos como notícias falsas. Um exemplo são as

¹² Disponível em https://motherboard.vice.com/en_us/article/xwk9zd/how-facebook-content-moderation-works Acesso em 6 de Março de 2019

postagens de movimentos anti-vacinação¹³, que não têm seu conteúdo retirado da plataforma e apenas perdem prioridade nas linhas do tempo dos usuários.

Para completar esse cenário, um relatório de fevereiro de 2019 da comissão da União Europeia¹⁴ aponta para a falta de transparência das atividades do Facebook na regulação de desinformação na plataforma. Segundo o documento, a empresa não fez nenhum progresso em combater notícias falsas, apesar de declarar nos Padrões da Comunidade:

“Reduzir a disseminação de notícias falsas no Facebook é uma responsabilidade que levamos a sério. Também reconhecemos que essa é uma questão desafiadora e delicada. Queremos ajudar as pessoas a se manter bem informadas sem deixar de lado o discurso público produtivo. Existe uma linha tênue entre notícias falsas e sátiras ou opiniões. Por esse motivo, não removemos notícias falsas do Facebook, mas, em vez disso, reduzimos significativamente sua distribuição, mostrando-as mais abaixo no Feed de Notícias.” (FACEBOOK, 2019)

Com base nos conceitos estabelecidos e considerando o cenário de aplicação e remoção de conteúdos, há exemplos noticiados pela grande mídia em diferentes países com casos de remoções devidas e casos em que a omissão da rede social em curar conteúdos mostrou-se falha.

Em 2015 a promotoria de Hamburgo abriu investigação contra executivos do Facebook na Alemanha por incitar ódio racial ao não tomar medidas contra a divulgação de mensagens de ódio do partido neonazista contra imigrantes que chegavam ao país¹⁵.

Durante o período das eleições brasileiras o Facebook removeu, em outubro de 2018, 68 páginas e 43 contas ligados a grupos de direita voltados para divulgação de anúncios e conteúdo anti-PT não pelo conteúdo publicado, mas por se tratar de

¹³ Disponível em <https://www.theverge.com/2019/3/7/18255107/facebook-anti-vaccine-misinformation-measles-outbreaks-group-page-recommendations-removal> Acesso em 19 de Março de 2019

¹⁴ Disponível em http://europa.eu/rapid/press-release_STATEMENT-19-1379_en.htm Acesso em 6 de Março de 2019

¹⁵ Disponível em https://brasil.elpais.com/brasil/2015/10/19/internacional/1445273931_947958.html Acesso em 21 e Março de 2019

contas falsas para divulgação de propaganda¹⁶. A mesma justificativa, que segundo as normas da empresa demandam que os usuários utilizem nome real e forneçam informações verídicas, foi utilizada para apagar 196 páginas e 87 perfis ligados ao Movimento Brasil Livre em Julho do mesmo ano¹⁷. O conteúdo divulgado pelas contas é secundário na análise dos moderadores, e as remoções não forma suficientes para reduzir a divulgação de notícias falsas.

Ainda em 2018 o Facebook foi a ferramenta utilizada pelos militares de Mianmar para divulgar notícias falsas e discurso de ódio e estimular o genocídio aos Rohingya, uma etnia muçulmana sob ataque no país de maioria budista desde os anos 1970. A análise realizada pela empresa aponta que não fizeram o suficiente para evitar as mortes apesar da violência crível nas postagens¹⁸. É curioso notar que no país o Facebook é sinônimo de Internet para parte significativa da população, e a falta de pessoal capacitado para reconhecer e remover conteúdos preconceituosos contra os Rohingya permitiu que agressões virtuais se tornassem reais.

Casos mais recentes como o ataque terrorista na Nova Zelândia, em que o atirador transmitiu ao vivo pelo Facebook seus atos, também repercutiram na plataforma, que teve que bloquear ou remover 1,5 milhão de reproduções do vídeo por seu conteúdo de violência explícita¹⁹. Os números mostram o esforço do Facebook para remover conteúdos como este e a demora em conseguir apagar ou reduzir o alcance dos vídeos.

Os moderadores e algoritmos que o Facebook utiliza para as remoções de conteúdo não são suficientes e são passíveis de falhas na remoção ou manutenção de conteúdos na rede. Apesar do sucesso em remover conteúdos pornográficos ou de violência, a dificuldade em reconhecer contextos de violência crível e a resposta tardia à divulgação de notícias falsas pelo Facebook colocam em dúvida o funcionamento do sistema de remoção de conteúdo. Nota-se também a pouca

¹⁶ Disponível em <https://gizmodo.uol.com.br/facebook-paginas-anti-pt-fazenda-anuncios/> Acesso em 21 de Março de 2019

¹⁷ Disponível em <https://politica.estadao.com.br/noticias/eleicoes,facebook-retira-do-ar-rede-de-contas-ligadas-ao-mbl-antes-das-eleicoes-2018,70002414717> Acesso em 21 de Março de 2019

¹⁸ Disponível em <https://www.gazetadopovo.com.br/mundo/facebook-reconhece-falha-em-impedir-violencia-em-mianmar-8tlgs4smu1dacm3n8bcqkw0ew/> Acesso em 21 de Março de 2019

¹⁹ Disponível em <https://olhardigital.com.br/noticia/facebook-diz-ter-retirado-de-1-5-milhoes-de-videos-do-ataque-na-nova-zelandia/83784> Acesso em 21 de Março de 2019

iniciativa para combater conteúdos como notícias falsas e a falta de transparência sobre o funcionamento dos algoritmos da empresa e do trabalho dos moderadores.

4. Considerações Finais

Considerando este panorama em que as redes sociais se estabeleceram como um importante foco de trocas e socialização capazes de moldar opiniões a partir da informação consumida, e tendo o Facebook como um dos principais espaços de interação, devemos considerar que papel a empresa possui na divulgação de ideias e como se posiciona no debate em torno da liberdade de expressão que é provocado quando conteúdos prejudiciais à formação da esfera pública estão presentes na rede.

As redes são democráticas no sentido de que todos podem participar e opinar, e a liberdade de expressão deve ser preservada juntamente à pluralidade de pensamento, mas o modo como as plataformas funcionam não garante um acesso igualitário à informação e a polarização ideológica pode encerrar os indivíduos em bolhas, dificultando os debates na esfera pública. A informação não é acessada por todos de maneira igual, e essa desigualdade colabora para que o debate político seja prejudicado e que se reproduza desinformação e preconceito. Somam-se a isso as restrições impostas sobre conteúdos específicos e a falta de sanções sobre outros. A falta de transparência e simetria no tratamento de conteúdos pode prejudicar os debates e a formação da opinião pública.

As mídias sociais modulam os conteúdos acessados pelos usuários a partir do que é promovido e interpretado por algoritmos programados para exibir conteúdos baseados nos interesses do usuário, mas também age removendo conteúdos de páginas e das linhas de tempo. A moderação do Facebook, que realiza seu trabalho a partir das definições nos Padrões da Comunidade, tem tido sucesso em excluir postagens com violência, pornografia, terrorismo segundo dados recentes. Não obstante, a moderação feita a partir dos algoritmos e da análise dos funcionários se mostra insuficiente e questionável pois não há transparência sobre as sanções e não

se medem os erros nos conteúdos retirados indevidamente ou nos que são mantidos na plataforma apesar de infringir as normas.

Há relatos de postagens com conteúdos que foram retirados do ar mesmo sem ferir os padrões da comunidade, apenas porque denúncias em massa indicariam ao algoritmo para retirar a publicação, por exemplo. A intransparência não nos permite ter certeza de como funciona o sistema de remoção do Facebook e se as regras definidas pela empresa são seguidas.

O Facebook colabora para a polarização de opiniões e o acirramento de disputas políticas ao permitir brechas em suas regras para postagem de conteúdos que promovem violência contra minorias, notícias falsas e não as retira da plataforma por não considerar que ferem os padrões da comunidade delineados pela empresa.

A formação de bolhas contribui para que não seja possível dialogar com ambos os lados e conhecer a realidade política e social como um todo a partir do que é apresentado e pela dificuldade de discernir informação de desinformação. As bolhas são fechadas a ponto do usuário não se expor a uma opinião diferente da sua, e é necessário mais transparência da empresa sobre as políticas de remoção de conteúdos e sua aplicação para determinar a influência das mídias sociais na esfera pública.

Referências

- BAYAMHHOGLU, Emre. **Depoliticization in the Digital Infosphere: When communication runs counter-democratic** in Digital Democracy in a Globalized World. Edward Elgar Publishing, 2017
- CARLSON, , Caitlin Ring. ROUSELLE, Hayley. **Report and Repeat: Investigating Facebook's Hate Speech Removal Process**. Seattle University, 2018
- CARVALHO, Camila Neves. **A eficácia da estratégia do Facebook na promoção e percepção de segurança da comunidade virtual**. Faculdade de Ciências Sociais e Humanas, Universidade Nova de Lisboa. 2017
- CRAWFORD, Kate. GILLESPIE, Tarleton. **What is a flag for? Social media reporting tools and the vocabulary of complaint** in New Media & Society, 2014
- DIAS, Thais Irene. **Análise do Comportamento dos Usuários em Relação à Diminuição Geral do Uso do Facebook**. Monografia de graduação. 2018. Departamento de Ciências Administrativas da Universidade Federal do Rio Grande do Sul Porto Alegre, 2018.
- FACEBOOK. **Understanding the Community Standards Enforcement Report**. 2018
- FLAXMAN, Seth. GOEL, Sharad. GAO, Justin M. **Filter Bubbles, Echo Chambers and Online News Consumption** in Public Opinion Quarterly Vol. 80, 2016
- FUCHS, Christian. **Mídias Sociais e a Esfera Pública** in Revista Contracampo, Vol. 34, Nº 3, 2015
- KRUSE, Lisa M. NORRIS, Dawn R. FLINCHUM, Jonathan R. **Social Media as a Public Sphere? Politics on Social Media** in The Sociological Quarterly, 2017
- LEMOS, André. **Nova Esfera Conversacional** in Dimas A. Künsch, D.A, da Silveira, S.A., et al, Esfera pública, redes e jornalismo., Rio de Janeiro, Ed. E-Papers, 2009
- OLIVEIRA, José Aparecido de. **Redes Sociais e a Participação Política na Esfera Pública** in Revista de Economia Política de las Tecnologias de la Informacion e Comunicacion Vol. XIV Nº 1, 2012
- PARISER, Eli. **O Filtro Invisível: o que a Internet está escondendo de você**. Rio de Janeiro: Zahar, 2012
- RABELO, Leon. **As Mídias Sociais e a Esfera Pública: Mudanças de Paradigma na Comunicação Contemporânea** apresentado no XII Congresso de Ciências da Comunicação a Região Centro-Oeste, 2010
- SERGEANT, Philip. TAGG, Caroline. **Social Media and the Future of Open Debate: A user-oriented approach to Facebook's filter bubble conondrum** in Discourse, Context & Media, 2018
- SPOHR, Dominic. **Fake News and Ideological Polirization: Filter bubbles and selective exposure on social media** in Business Information Review Vol. 34 Nº 3, 2017
- WORLD WIDE WEB FOUNDATION. **The Invisible Curation of Content: Facebook's News Feed and our Information Diets**, 2018