

HASSREDE IN DIGITALEN UMGEBUNGEN

Definitionen, Spezifitäten und Kontext
der Online-Diskriminierung in Brasilien
ausgehend von Twitter und Facebook



Rio de Janeiro, März 2021

Unterstützt von



Embaixada
da República Federal da Alemanha
Brasília

HASSREDE IN DIGITALEN UMGEBUNGEN:

Definitionen, Spezifitäten und Kontext der
Online-Diskriminierung in Brasilien ausgehend von Twitter
und Facebook

Rio de Janeiro

FGV DAPP

2021

Dados Internacionais de Catalogação na Publicação (CIP)
Ficha catalográfica elaborada pela Biblioteca Mario Henrique Simonsen/FGV

Hassrede in digitalen umgebungen : definitionen, spezifitäten und kontext der online-diskriminierung in Brasilien ausgehend von Twitter und Facebook / Koordinierung Marco Aurélio Ruediger, Amaro Grassi. – Rio de Janeiro : FGV DAPP, 2021.
1 recurso online (60 p.) : PDF

Inclui bibliografia.
ISBN: 978-65-86845-13-6

1. Mídia digital - Censura. 2. Políticas públicas. 3. Discurso de ódio na Internet. 4. Liberdade de expressão. 5. Redes sociais on-line. 6. Discriminação. I. Ruediger, Marco Aurelio, 1959- . II. Grassi, Amaro. III. Fundação Getulio Vargas. Diretoria de Análise de Políticas Públicas.

CDD – 302.23

Wie man zitiert

RUEDIGER, M. A.; GRASSI, A. (Koord.). **Hassrede in digitalen umgebungen : definitionen, spezifitäten und kontext der online-diskriminierung in Brasilien ausgehend von Twitter und Facebook.** Policy paper. Rio de Janeiro: FGV DAPP, 2021

HERAUSGEBER



Die Fundação Getulio Vargas (Stiftung Getulio Vargas) wurde 1944 mit dem Ziel gegründet, die sozioökonomische Entwicklung Brasiliens mittels der Ausbildung qualifizierter Manager in privaten sowie öffentlichen Bereichen zu fördern. Im Lauf der Zeit erweiterte die FGV ihren Tätigkeitsbereich um Wissensgebiete wie Sozialwissenschaften, Rechtswissenschaft, Wirtschaft, Geschichte, Angewandte Mathematik und Internationale Beziehungen. Mit zehn Fakultäten zeichnet sie sich durch ihre Qualität und Exzellenz aus.

Edifício Luiz Simões Lopes (Hauptsitz)

Praia de Botafogo 190, Rio de Janeiro

RJ - PLZ 22250-900

Postfach 62.591 PLZ 22257-970

Tel +55 21 3799-5498 | www.fgv.br

Erster Präsident und Gründer

Luiz Simões Lopes

Präsident

Carlos Ivan Simonsen Leal

Vizepräsidenten

Francisco Oswaldo Neves Dornelles (Lizenziat)

Marcos Cintra Cavalcanti de Albuquerque (Lizenziat)



Direktor

Marco Aurelio Ruediger

Doktor der Soziologie und Direktor des Vorstandes für Evaluation öffentlicher Politiken der Stiftungshochschule Getulio Vargas (FGV/DAPP). Seine Hauptinteressengebiete sind Politische Soziologie, Kommunikation und soziale Netzwerke sowie Technische Innovation und ihre Auswirkungen auf die Demokratie. Er ist Mitglied der Beratungsausschüsse der *D4D-Coalition* des *National Democratic Institute* und der *Social Science One*. Derzeit leitet er die Gründung der Fakultät für Kommunikations-, Medien- und Informationswissenschaften der Fundação Getulio Vargas und das Projekt „Digitalisation and Democracy in Brazil“, das mit der Unterstützung des deutschen Auswärtigen Amtes und der Deutschen Botschaft Brasília realisiert wird.

E-Mail: marco.ruediger@fgv.br

FGV DAPP

+55 21 37996208

www.dapp.fgv.br | dapp@fgv.br

Forschungskoordination

Marco Aurelio Ruediger

Amaro Grassi

Forscher

Dalby Hubert

Danielle Sanches

Eurico O. Matos Neto

Luiza C. Santos

Lucas Roberto da Silva

Polyana Sampaio Barbosa

Renata Tomaz

Fachliche Prüfung

Renata Tomaz

Grafikdesign

Luis Gomes

Daniel Cunha

INHALTSVERZEICHNIS

EXECUTIVE SUMMARY	5
ZUSAMMENFASSUNG DER ERGEBNISSE	5
VORSTELLUNG DES PROJEKTS	6
METHODEN	8
THEORETISCHE DEFINITIONEN UND KONZEPTUALISIERUNGEN	10
1. Was ist Hassrede?	10
2. Juristische Ansätze und Rahmenbestimmungen bezüglich der Hassrede weltweit und in Brasilien	14
3. Hassrede in digitalen Umgebungen	20
4. Richtlinien der digitalen Plattformen bezüglich Hassrede	25
ANALYSE DER DEBATTE ÜBER HASSREDEN UND ZENSUR AUF TWITTER UND FACEBOOK	36
FAZIT	56
LITERATURVERZEICHNIS	60

EXECUTIVE SUMMARY

Die Praktiken der Online-Hassrede stellen eine Herausforderung für die Demokratien dar, insbesondere weil sie Fragen rund um die Meinungsfreiheit, eine der Säulen der Demokratie, aufwerfen. Damit die Verhandlungsdynamiken zwischen Sicherheit der geschützten Gruppen und Gewährleistung der Meinungsfreiheit verstanden werden können, untersucht die vorliegende Studie die öffentliche Debatte rund um Hassrede und Zensur auf Twitter und Facebook. Dafür wird zunächst ein allgemeiner Überblick über die Bestimmungen und Gesetzgebungen, die über das Verständnis der Hassrede informieren, gegeben. Dabei werden die Spezifitäten des Online-Kontexts und die von den Plattformen vorgeschlagenen Bestimmungs- und Selbstregulierungsformen diskutiert. Anschließend wird die öffentliche Debatte rund um Hassrede und Zensur basierend auf der Datenerfassung beider sozialen Netzwerke analysiert. Dabei werden insbesondere die Höhepunkte der Diskussion über die Themen und die damit verbundenen Ereignisse berücksichtigt. Schließlich werden die Hauptakteure und wichtigsten Interaktionen, die den Höhepunkt der Debatte über Hassrede und Zensur erreichten, ausgehend von Visualisierungen, die einer strukturellen Analyse der sozialen Netzwerke dient, identifiziert. Die Studie verwendet bibliographische und dokumentarische Quellen sowie Daten, die auf Facebook und Twitter erfasst und anschließend sowohl qualitativ als auch quantitativ analysiert wurden.

Schlüsselwörter: Hassrede; digitale Plattformen; Zensur; Meinungsfreiheit; soziale Netzwerke.

ZUSAMMENFASSUNG DER ERGEBNISSE

- Die theoretische Überprüfung in Zusammenhang mit der Analyse der Plattformen zeigte, dass es zwar Richtlinien zur Bestrafung von Accounts, die beleidigende Inhalte veröffentlichen, gibt, es aber schwer und komplex ist, derartige Postings zu ermitteln.

- Die Forschung setzt den Fokus auf die Erfassung von Daten mit Bezug mit Hassrede und Zensur und umfasste zwischen November 2020 und Januar 2021 ca. 11 Mio. Postings auf Twitter und 34.000 auf Facebook;
- Hinsichtlich der Debatte über Hassrede auf Twitter und Facebook konzentrierten sich die Postings auf Themen wie Misogynie und Rassismus. Auf beiden Plattformen erreichten sie ihren Spitzenwert am 20. November 2020, Tag des Schwarzen Bewusstseins. Die Postings verstärkten sich durch die Diskussion um strukturellen Rassismus in Verbindung mit der Nachricht über den körperlichen Angriff mit Todesfolge am vorherigen Tag auf einen schwarzen Kunden im Supermarkt Carrefour in Porto Alegre (RS).
- Der Diskurs über Zensur und Meinungsfreiheit verschärfte sich mit den durch die Plattformen verhängten Sanktionen für Accounts wie diejenige des ehemaligen Präsidenten der Vereinigten Staaten, Donald Trump, deren Postings zur Infragestellung des Ergebnis der Präsidentschaftswahlen mit dem Sturm von Extremisten auf das Kapitol assoziiert werden. Sowohl auf Twitter als auch auf Facebook mobilisierte die Sperrung von Accounts die Debatte über die Zensur auf den Plattformen anhand von der Meinungsäußerung ihrer Nutzer.

VORSTELLUNG DES PROJEKTS

Die Aufrechterhaltung demokratischer Systeme hängt unter anderem von einem gesunden Gleichgewicht zwischen dem Recht auf freie Meinungsäußerung und der Bekämpfung der Intoleranz, vor allem gegen Minderheiten, ab. Genauso wichtig, wie die Möglichkeit zu freier Meinungsäußerung jedes Individuums, ist die Aufrechterhaltung einer demokratischen Umgebung, in der die Diskurse, die beispielsweise Hassreden oder Gewalt fördern und einen Angriff auf die Diversität darstellen, ausreichend unterdrückt werden. Die Möglichkeiten, die die sozialen Medien zur öffentlichen Meinungsäußerung bieten, stellen neue Herausforderungen für dieses Gleichgewicht dar.

Unterschiedliche juristische Auffassungen verstehen die Rolle des Staats bei der Aufrechterhaltung dieser Umgebung auf verschiedene Formen: Länder wie die Vereinigten Staaten schützen in erster Linie die Meinungsfreiheit, während beispielsweise Deutschland

den Inhalten, die geäußert werden dürfen, strengere Grenzen setzt. Im Online-Kontext sollen zudem die Positionierungen der digitalen Plattformen selbst, auf denen viele dieser Diskurse verbreitet werden, berücksichtigt werden. Sie setzen sich den staatlichen Regulierungen in der Regel entgegen und stehen für die Selbstregulierungen ein.

Angesichts des Bedarfs eines Ökosystems, in dem Meinungsfreiheit und Sicherheit diverser Gruppen koexistieren, ist die Problematik der Hassreden von großer Bedeutung, wenn es um digitale Demokratie geht. Die Berücksichtigung dieser Frage nimmt sowohl in Forschungen mit einem größeren Umfang, wie die Untersuchungen der Desinformation (BARBOSA et al. 2020), als auch in spezifischen Recherchen, die ihre konstitutiven Aspekte analysieren (NERIS et al. 2020), zu. Die vorliegende Studie gibt einen allgemeinen Überblick über die Bestimmungen und Gesetzgebungen bezüglich der Hassrede und diskutiert die Spezifitäten des Online-Kontexts sowie die von den Plattformen vorgeschlagenen Bestimmungs- und Selbstregulierungsformen. Anhand der Datenerfassung in sozialen Netzwerken, spezifisch auf Twitter und Facebook, wurde die öffentliche Debatte über Hassrede und Zensur unter Berücksichtigung vor allem der Spitzenwerte der Erwähnungen der Themen und der damit verbundenen Ereignisse analysiert. Zudem wurde eine Analyse der Hauptakteure und wichtigsten Interaktionen, die ausgehend von Momenten von größerem Engagement in der Debatte über Hassrede und Zensur und anhand von Visualisierungen der strukturellen Darstellung der sozialen Netzwerke mobilisiert wurden, durchgeführt.

Dieses ist das dritte Policy Paper, das vom Projekt **Digitalisierung und Demokratie in Brasilien**, eine Partnerschaft zwischen dem Vorstand für Evaluation öffentlicher Politiken der Fundação Getulio Vargas (FGV DAPP) und der Deutschen Botschaft Brasília, veröffentlicht wird. Das Projekt läuft bis 2022 und profitiert von der Erfahrung in der angewandten Forschung zu sozialen Netzwerken und der im letzten Jahrzehnt vom FGV DAPP aufgebauten Expertise über die Analyse öffentlicher Politiken.

Es wird die Entwicklung von Strategien zum Verständnis sowie zur Bekämpfung der neuen und der unmittelbar bevorstehenden Herausforderungen der brasilianischen Demokratie

beabsichtigt, die nun von einem Szenario digitales Extremismus und dessen gesellschaftlichen Folgen geprägt wird. In diesem Sinne liegt der Schwerpunkt in erster Linie auf der Stärkung der demokratischen Werte und auf der Schaffung von Mechanismen und Kenntnissen, die für eine konsequente und fruchtbare öffentliche Digitaldebatte erforderlich sind.

METHODEN

Für die vorliegende Arbeit wurden bibliographische und dokumentarische Quellen sowie Daten verwendet, die auf Facebook und Twitter erfasst wurden. Die bibliographische Forschung dient den ersten zwei Abschnitten dieses Dokuments, in denen der Stand der Technik bezüglich der Hassrede in digitalen Umgebungen vorgestellt wird. Die dokumentarische Forschung beschäftigt sich im ersten Abschnitt mit den in Brasilien geltenden gesetzlichen Richtlinien, und im zweiten Abschnitt mit den Positionierungen der jeweiligen Plattformen bezüglich der Hassrede, die durch sie als Medium verbreitet werden. Für die Zuordnung der Hassrede-Begriffe und der von Plattformen festgelegten Verhaltensrichtlinien, wurden die Nutzungsbedingungen und Gemeinschaftsrichtlinien von Twitter, Facebook, Instagram und YouTube konsultiert. Diese sind auf bestimmten Website-Tabs der jeweiligen Plattformen öffentlich zugreifbar und werden regelmäßig aktualisiert. Zusätzlich wurden gegebenenfalls weitere Links, die auf den Plattformen zu finden sind, verwendet. Die in dieser Arbeit vorliegenden Informationen beziehen sich auf die Nutzungsbedingungen und Gemeinschaftsrichtlinien, die im Februar 2021 für die jeweilige Plattform galten.

Bezüglich der Datenanalyse sozialer Netzwerke, welche diese Studie vervollständigt, griff der empirische Ansatz auf automatische Methoden für die Erfassung und Einordnung von Postings auf Social-Media-Plattformen zurück. Angesichts der auf Twitter zu erfassenden Daten wurde zunächst eine Liste von Begriffen erstellt, die den thematischen Rahmen

dieser Studie, Hassrede und Meinungsfreiheit, umfassen konnten. Ein Teil der Erfassung wurde mit dem Tracking-Dienst TrendsMap durchgeführt, welcher das Echtzeit-Monitoring der Postings mit großem Engagement sowohl bei dem Teilen als auch den Kommentaren und Antworten ermöglicht. Darüber hinaus wurde diese Liste von Begriffen verwendet, um die Postings aus dem analysierten Zeitraum - zwischen dem 16. November 2020 und dem 6. Februar 2021 - herauszufiltern. Diese wurden in einer Datenbank des Managementsystems PostgreSQL gespeichert.

Der anfänglichen Datenerfassung auf Twitter folgte die Klassifizierung der erhaltenen Daten basierend auf den sprachlichen Schemata, die mit Bezug auf das Thema aufgebaut wurden (RUEDIGER 2017). Diese Etappe bestand aus der Entwicklung von Suchsyntaxen, die sich an den beiden zentralen Themen dieser Studie orientieren: „Hassrede“ und „Meinungsfreiheit“ (einschließlich „Zensur“). Suchsyntaxen entsprechen einem Satz von sprachlichen Werten, welche mithilfe boolescher (oder logischer) Operatoren artikuliert werden. Dieser basiert auf semantisch-pragmatischen Elementen, mit denen ein Thema normalerweise instanziiert wird. Diese Suchsyntaxen wurden also mittels einer textuellen Suche in der Datenbank durchgeführt. Für die Erfassung und Klassifizierung der Daten auf Facebook wiederum wurden die Suchsyntaxen, die für die Klassifizierung der Daten von Twitter bezüglich der Themen „Hassrede“ und „Meinungsfreiheit“ ausgearbeitet worden waren, an CrowdTangle, ein Tool von Facebook für das Monitoring der Postings auf der Plattform, angepasst.

Trotz der Tatsache, dass die angewandte Methodik auf der Konstruktion von strukturierten sprachlichen Schemata basiert, ist das Zielthema dieser Forschung, die Debatte über Hassrede, äußerst komplex. Aus diesem Grund ließen sich Postings beobachten, die zwar einen vereinfachten Wortschatz aufzeigten, aber unterschiedliche Bedeutungen beinhalteten. Diese Herausforderung führte dazu, dass einige Filter nach den Erfassungen erstellt wurden, wobei der Fokus auf die Spitzenwerte der Erwähnungen gelegt wurde, um das Engagement der Debatte zu analysieren. Die qualitative Analyse der Momente, in denen die Diskussion sich verschärft, wird in diesem Sinne zu einer methodologischen

Strategie, die dem Umgang mit den Einschränkungen der Ergebnisse der Datenextraktion ausgehend von der Entwicklung von Suchsyntaxen dient.

THEORETISCHE DEFINITIONEN UND KONZEPTUALISIERUNGEN

1. Was ist Hassrede?

Die Definition von Hassrede - online oder offline - ist äußerst schwer festzulegen. Dieses Thema wird von Akademikern diverser Bereiche, Juristen und Gesetzgebern oft diskutiert, ohne einen Konsens zu erreichen. Es lässt sich behaupten, dass die Hassrede von Vorurteilen, Beleidigungen oder persönlich angreifenden Aussagen gegen eine bestimmte Person oder Gruppe aufgrund ihrer Charakteristika motiviert wird (COHEN-ALMAGOR 2011; FARIS et al. 2016). Allerdings weist Sellars (2016) darauf hin, dass trotz der Existenz umfassender Literatur über die Gründe und Auswirkungen derartiger Diskurse noch ein Mangel an Definition und Systematisierung des Begriffs besteht.

Es lässt sich feststellen, dass eine Vielzahl von Themen unter die Definition von Hassrede fallen kann, wie z.B. Verleumdungen und Beleidigungen, die leicht zu erkennen sind. Die Sprache hat jedoch viele Nuancen und nicht immer können die Inhalte von dem Sprecher oder der betroffenen Person zwangsläufig als Hassrede betrachtet werden. Ein weiterer hervorzuhebende Punkt beschäftigt sich mit der Verwendung von Codes, vor allem seitens Online-Communities, um Hass auszudrücken. Die Erkennung des Diskurses wird dadurch erschwert. Die Literatur zum Thema deckt ebenso die Existenz von sogenannten Gewaltdiskursen auf, die zur Gewalt in der Offline-Umgebung gegen schutzbedürftige Gruppen aufhetzen (SIEGEL 2020).

In diesem Sinne können die bestehenden Definitionen von Hassrede sehr weitreichend sein. So umfassen sie beispielsweise Inhalte, die sich aufgrund von körperlicher Eigenschaften gegen eine Person oder Gruppe richten, oder aber aufgrund von Haltungen,

die dem festgelegten, normativen Standard widersprechen (PAREKH et al. 2012). Weiterhin lässt sich Hassrede auch dadurch definieren, dass die Inhalte nachhaltigen Schaden anrichten. Restriktive Definitionen gehen von einer direkten Assoziation von Hassrede mit der Anstiftung zu Gewalt in Menschenmassen oder zur Körperverletzung gegen bestimmte Gruppen aus (BENESCH 2013). Diese Dualität der Definitionsversuche (weitreichend und restriktiv) spiegelt die Schwierigkeit wider, eine Definition festzulegen, die die Vielzahl an als Hassrede zu betrachtenden Phänomenen adäquat umschließt.

Trotz der dem Begriff innewohnenden Komplexität und der in Verbindung mit dem Thema stehenden Herausforderungen übernahm die vorliegende Studie die Definition aus dem *Guia para análise de discurso de ódio* (Handbuch zur Analyse der Hassrede). Laut diesem beinhalten Hassreden „Ausdrücke, die eine vulnerable Gruppe oder eine Person, die einer vulnerablen Gruppe angehört, negativ bewerten und darauf zielen, diese als weniger würdig darzustellen in Bezug auf Zugang zu Rechten, Möglichkeiten und Hilfsmitteln, darzustellen und somit Diskriminierung oder Gewalt zu legitimieren“ (LUCCAS/GOMES/SALVADOR 2020: 4). Nandi (2018) zufolge gehören die Gruppen, die dem Angriffsziel von Hassreden entsprechen, üblicherweise zu Minderheiten und befinden sich in einer Situation sozialer sowie finanzieller Vulnerabilität wie Schwarze, Indigene, Homosexuelle, Frauen und religiöse Minderheiten. Dem Autor zufolge hängt die Verbreitung von Hassrede mit der Einschüchterung der Opfer bei der Inanspruchnahme von Rechten zusammen. Das erklärt, warum die Hassrede sich verstärkt auf Individuen oder Gruppen richtet, wenn diese in akademischen, juristischen und legislativen Debatten in Erscheinung treten. Da der Angriff die Funktion erfüllt, individuelle Rechte zu delegitimieren, hat die Hassrede laut Nandi immer eine politische Tendenz.

Ein erwähnenswerter Punkt in dieser Debatte sind die speziellen Merkmale der Online-Hassrede, weil hierbei die Verfasser dessen, ihre unterschiedlichen Angriffsziele sowie diverse Hintergründe und Taktiken berücksichtigt werden müssen. Des Weiteren ist die Tatsache, dass die Verfasser von Hassrede im Netz ihre eigenen Anhänger haben und die Reichweite ihres Einflusses der Inhalte gegen bestimmte Individuen oder Gruppen leicht erweitern können. Des Weiteren ist die Tatsache, dass die Verfasser von Hassrede im

Netz ihre eigenen Anhänger haben und die Reichweite ihres Einflusses der Inhalte gegen bestimmte Individuen oder Gruppen leicht erweitern können. Dieser Aspekt beunruhigt Akademiker, die in der Debatte über Hassrede beteiligt sind, da die Anstiftung zu Gewalt seitens eines einzigen Individuums andere Auswirkungen hat, wenn sie als eine Art „Hasskampagne“ von einer digitalen Menschenmasse verwendet wird (SELLARS 2016).

Unter diesem Aspekt gehen die Diskussionen immer mit der Beziehung zwischen Meinungsfreiheit und Hassrede einher. Obwohl es sich hierbei um eine Debatte im Rechtsbereich handelt, lässt sich behaupten, dass die in internationalen Abkommen und nationalen Gesetzen vorgesehene Meinungsfreiheit in verschiedenen Bereichen zunehmend als ein Recht interpretiert wird, das nicht absolut, sondern durch die Gewährleistung anderer Grundrechte bedingt ist, wobei die individuelle Freiheit Rechenschaftspflichten und Einschränkungen mit sich bringt (LUNA/SANTOS 2014). Laut Antoine Buyse (2014) ist ein Konsens über die legalen und sozialen Grenzen der Meinungsfreiheit jedoch schwer zu erreichen, da diese selten in direkter Verbindung mit Gewalt steht.

Die Debatte über Hassrede muss allerdings von dem Diskurs über Angst unterschieden werden. Während Individuen bestimmter Gruppen durch Hassworte angegriffen werden, richtet sich der Angstdiskurs in der Regel an Angehörige derselben sozialen Gruppe des Angreifers. Das Ziel dabei ist es, ein internes Angstgefühl in dieser Gruppe gegenüber externen Individuen zu erzeugen. Mögliche folgende Gewalttaten werden durch diesen Mechanismus der Angst legitimiert. Die Entstehung von Gewaltdiskursen werde also vom innerhalb sozialer Gruppen gehegten Angstgefühl aktiviert. Buyse, Professor für Menschenrechte an der Universität Utrecht, erinnert daran, dass in Kontexten, in denen es strenge Gesetze gegen Hassrede gibt, die Individuen dazu neigen, direkte Ausdrücke der Anstiftung zur Gewalt gegen bestimmte Gruppen zu vermeiden und stattdessen indirekte Formulierungen zur Förderung von Feindseligkeit gegen bestimmte Gruppen zu verwenden.

Bemerkenswert ist auch, dass die durch den Gewaltdiskurs entstehende Angst sozialer Minderheiten und bedrohter Gruppen oft größer ist als die Gewalt (Waffengewalt, körperliche Gewalt) selbst. Er wird als ein Mittel präventiver Gewalt verwendet, eine Art Bestimmung der Verhältnisse zwischen denjenigen, die diesen Diskurs führen, und denjenigen, die davon bedroht werden.

Hinsichtlich der Angstdiskurse ist es wichtig, auf die Auswirkungen, die sie auf eine Gruppe oder die ganze Bevölkerung haben, aufmerksam zu machen. Nachrichten über verübte Verbrechen in brasilianischen Städten oder über von Terrororganisationen begangenen Anschläge in Europa, die täglich in Medienkanälen (im Fernsehen oder in der gedruckten oder digitalen Presse) zu sehen sind, führen dazu, dass der Diskurs über die Vernichtung dieser Gruppen - die verbreitete Idee, dass „nur ein toter Verbrecher ein guter Verbrecher ist“ - oft von der Bevölkerung befürwortet wird und somit mit der Idee von Gerechtigkeit und demokratischem Haftsystem kollidiert. Anna Simons und John Mueller (2001) zufolge ist dies auf die Vereinnahmung des Angstdiskurses einer Gruppe durch eine andere Gruppe zurückzuführen, was extreme Reaktionen auslösen kann, indem die Gewalttaten zugunsten vom Gemeinwohl akzeptiert werden¹.

Eine weitere relevante Begriffsunterscheidung ist das Verständnis dessen, was Hassrede ist, und dessen, was den extremistischen Diskurs darstellt. Laut einem von der UNESCO veröffentlichten Bericht (ALAVA/FRAU-MEIGS/HASSAN 2017) wird die Aufforderung zur Anwendung extremer Gewalt zunehmend im Internet festgestellt, welches sich als ein Mittel für die Ausbreitung der Radikalisierung extremistischer Diskurse und Ideologien bezeichnen lässt. Allerdings behauptet derselbe Bericht, es gäbe kein Ursache-Wirkung-Verhältnis zwischen radikalen Online-Diskursen und echten extremen Gewalttaten trotz der Verbreitung derartiger Diskurse. Die Frage, inwiefern die öffentlichen oder privaten Postings und Nachrichten in digitaler Umgebung alltägliche soziale Praktiken bestimmter Gruppen beeinflussen, führt zur Frage, ob die Social-Media-Unternehmen

¹ Die erwähnten Beispiele sprechen Fragen an, die die ganze Gesellschaft anbelangen, aber jede Gruppe, die sich durch die Aktionen einer anderen Gruppe verfolgt, unterdrückt oder betroffen fühlt, kann den Angstdiskurs nutzen oder aufnehmen und die Anwendung von physischer und Waffengewalt akzeptieren.

Verantwortung für die Bekämpfung entsprechender Diskurse mittels der Inhaltsmoderation tragen, wodurch Druck auf gerichtliche Instanzen für die Verabschiedung von Gesetzen zur Regulierung der Online-Aktivitäten der Nutzer ausgeübt werden würde, oder ob eine gemeinsame Strategie zur Regulierung der Inhalte entwickelt werden soll.

Schließlich ist die Hassrede kein einheitliches Phänomen. Diskriminierungspraktiken und -diskurse gegen Minderheiten stehen in direkter Verbindung mit den sozialen und politischen Kontexten, in denen sie zum Ausdruck kommen. Die Vielzahl an Formen, die die Hassrede in unterschiedlichen Kontexten annimmt, ist ein Faktor, der die Entwicklung von Strategien zur Bekämpfung der Hassverbreitung in Online- oder Offline-Umgebungen deutlich erschwert. Aus diesem Grund ist es angemessener, das Phänomen als „Hassreden“, also in der Pluralform, zu bezeichnen, da ihre Praktiken in sozialen Kontexten durch die Vielfalt und den engen Zusammenhang mit diversen Kontexten bestimmt werden.

2. Juristische Ansätze und Rahmenbestimmungen bezüglich der Hassrede weltweit und in Brasilien

Der immer komplexer werdende Begriff der Hassreden, einschließlich seine Verwendung im Plural, hängt in hohem Maße mit den Bemühungen im juristischen Bereich zusammen. Die internationalen Abkommen, vor allem diejenigen, die nach beiden Weltkriegen entstanden, fördern Diskussionen und Debatten, die zunehmend rechtliche Mechanismen von den Staaten verlangen, um Handlungen, deren Motivation auf den Wunsch nach Disqualifizierung, Schweigen, Invisibilisierung und Dehumanisierung vulnerabler Gruppen zurückzuführen ist, vorzubeugen, zu verhindern und zu bestrafen. Obwohl die internationalen Abkommen keine Definitionen von Hassrede bieten, legen sie grundlegende Rahmenbedingungen für ihre Identifizierung, und dementsprechend für die Bekämpfung ihrer direkten und indirekten Auswirkungen fest.

Diese Chartas, Pakte und Erklärungen basieren auf dem Prinzip der Menschenwürde, welches nach den „unsäglichen Leiden“ des Kriegs vehement gefordert wurde, wie die Präambel der Charta der Vereinten Nationen erkennen lässt. Mit dem Dokument entstand die Organisation der Vereinten Nationen (UN 1945). Der Text verteidigt die „Achtung vor den Menschenrechten und Grundfreiheiten für alle ohne Unterscheidungen aufgrund der „Rasse“, des Geschlechts, der Sprache oder der Religion“. Auf ebendieser Idee, dass das Menschsein allein die Aufrechterhaltung rechtlicher Gleichheit und des Rechts auf Freiheit begründet, basiert das gerichtliche Verständnis von der Bekämpfung von Hassreden. Die Allgemeine Erklärung der Menschenrechte (UN 1948) erweitert die Bekämpfung der Diskriminierung bezüglich „Rasse, Hautfarbe, Geschlecht, Sprache, Religion, politischer oder sonstiger Überzeugung, nationaler oder sozialer Herkunft, Vermögen, Geburt oder sonstigem Stand“ (Art. 2) und geht „gegen jede Aufhetzung zu einer derartigen Diskriminierung“ (Art. 7) oder gegen Angriffe auf „Ehre und Ruf“ (Art. 12) vor. In diesem Sinne werden diskriminierende Diskurse widergelegt.

In der zweiten Hälfte des 20. Jahrhunderts verstärkte sich der Widerstand gegen Diskriminierung und Aufhetzungsmechanismen. Das 1965 von der UN besiegelte Internationale Übereinkommen zur Beseitigung jeder Form von Rassendiskriminierung² führte den Begriff „Rassenhass“ in die Debatte ein, welcher sich kurz danach auf die nationalen und religiösen Sphären Brasiliens im Rahmen des 1966 von der UN übernommenen Internationalen Pakts über bürgerliche und politische Rechte³ und der 1969 von der Organisation Amerikanischer Staaten unterzeichneten Amerikanischen Menschenrechtskonvention⁴ ausdehnte. Alle drei Dokumente standen dafür ein, dass die

² Verordnung Nr. 65.810 vom 8. Dezember 1969. Sie verkündet das Internationale Übereinkommen zur Beseitigung jeder Form von Rassendiskriminierung. Verfügbar unter: http://www.planalto.gov.br/ccivil_03/decreto/1950-1969/D65810.html. Letzter Zugriff am: 26.02.2021.

³ Verordnung Nr. 592 vom 6. Juli 1992. Sie verkündet die Internationalen Akten. Internationaler Pakt über bürgerliche und politische Rechte. Verfügbar unter: http://www.planalto.gov.br/ccivil_03/decreto/1990-1994/d0592.htm. Letzter Zugriff am: 26.02.2021.

⁴ Verordnung Nr. 678 vom 6. November de 1992. Sie verkündet die Amerikanische Menschenrechtskonvention vom 22. November 1969. Verfügbar unter: http://www.planalto.gov.br/ccivil_03/decreto/d0678.htm. Letzter Zugriff am: 26.02.2021.

Anstiftung zum Hass bzgl. „Rasse“, Nation und Religion zu einer gerichtlich strafbaren Handlung werden sollte. Im Jahr 1992 promulgierte Brasilien sowohl den Internationalen Pakt über bürgerliche und politische Rechte als auch die Amerikanische Menschenrechtskonvention, d.h. es wurde ihnen Gesetzeskraft im Lande verliehen.

In der Erklärung und dem Aktionsprogramm der Dritten Weltkonferenz gegen Rassismus (Weltkonferenz gegen Rassismus, rassistische Diskriminierung, Fremdenfeindlichkeit und damit zusammenhängende Intoleranz, 2001)⁵ wurden die Staaten aufgefordert, weniger Strafmaßnahmen und mehr Präventions- und Schutzmaßnahmen zu ergreifen. Zudem wurde der Bekämpfung des „Hasses“ der Bedarf einer Vielzahl von Akteuren anerkannt sowie dessen diverse Erscheinungsformen und Hauptziele aufgelistet. Hervorgehoben wurde die Rolle unterschiedlicher Medien, vor allem der digitalen, einerseits bei der Verbreitung von Ideologien, die den Hass gegenüber vulnerablen Gruppen schüren und andererseits bei der Förderung von Aktionen zur Bekämpfung dessen. Außerdem verlangte der Artikel Nr. 147 des Aktionsprogramms den Staaten ab, „soweit möglich geeignete Rechtsvorschriften zur strafrechtlichen Verfolgung derjenigen zu erlassen und anzuwenden, die für die Aufstachelung zum Rassenhass oder zu rassistisch motivierter Gewalt mittels der neuen Informations- und Kommunikationstechnologien, einschließlich des Internets“ unter Gewährleistung der Meinungsfreiheit und des Rechts der freien Meinungsäußerung einzuhalten und umzusetzen.

Die Meinungsfreiheit wird von internationalen Abkommen - wie auch von nationalen Rechtsvorschriften wie der brasilianischen Verfassung von 1988 - als ein Grundrecht für die volle Entwicklung demokratischer Systeme anerkannt. Seit Beginn des 21. Jahrhunderts verstärken sich jedoch durch die zunehmenden Möglichkeiten der Verbreitung von Hassreden in digitaler Umgebung die Diskussionen über die Rolle des Staates, nicht nur bei der Gewährleistung der Meinungsfreiheit, sondern auch bei der Einschränkung dieser im Falle von Verletzung bestimmter Prinzipien, wie der Menschenwürde. Diejenigen, die

⁵ Erklärung und Aktionsprogramm der Dritten Weltkonferenz gegen Rassismus, Rassendiskriminierung, Fremdenfeindlichkeit und damit zusammenhängende Intoleranz. Brasília, DF: Bundeskulturministerium 2001. Verfügbar unter: http://www.unfpa.org.br/Arquivos/declaracao_durban.pdf. Letzter Zugriff am: 26.02.2021.

sich diesem Staatseingriff entgegensetzen, glauben, dass derartige Eingriffe in die freie Meinungsäußerung, einschließlich Fällen von Hassrede, das Recht auf Freiheit verletzen (DWORKIN 2006, 2009) und somit die Säulen der Demokratie entkräften. Die Befürworter eines Staats, der der Meinungsfreiheit Grenzen setzt, argumentieren hingegen, dass dies notwendig für die Bekämpfung der Intoleranz ist (WALDRON 2012). Dieser Meinungskonflikt stellt dementsprechend ein Hindernis für die Durchsetzung der Menschenwürde und für das demokratische Bewusstsein selbst dar. Die Auffassung der ersten Gruppe wird in der Regel eher von US-amerikanischen Theoretikern vertreten und basiert üblicherweise auf dem Prinzip der Neutralität des Staates. Die europäischen Staaten, in denen zwei Weltkriege und der Holocaust geschahen, neigen zu einem strengeren Ansatz bzgl. der Grenzen der Meinungsfreiheit (BRUGGER 2007, CONSANI 2015).

In diesem Sinne ist es nicht einfach, Rechtsnormen zu schaffen, die sowohl die sozialen Dynamiken, die Hassreden einbeziehen, als auch die Kontexte, aus denen die Hassreden entstehen, berücksichtigen. Die sogenannten *Soft Laws* fungieren als Unterstützung für die Überwindung der in den internationalen Abkommen erkannten Einschränkungen. Sie sind Leitlinien, die im Kontext des internationalen Rechts darauf zielen, Konfliktfällen, in denen z.B. die Forderung nach Gewährleistung der Meinungsfreiheit zur Verbreitung von Hassreden missbraucht wird, entgegenzuwirken. Der Posten eines Sonderberichterstatters zur Meinungsfreiheit beispielsweise wurde von der Organisation Amerikanischer Staaten geschaffen, um Anweisungen ausgehend von der Amerikanischen Menschenrechtskonvention zu geben.

Die brasilianische Rechtsordnung lehnt sich in diesem Thema eher an die europäische Auffassung an, in der das Recht auf freie Meinungsäußerung der Gleichberechtigung und der Menschenwürde unterliegt. In Brasilien ist die Diskriminierung aufgrund von „Rasse“, Hautfarbe, Ethnie, Religion und Herkunft ein Straftat gemäß dem Gesetz Nr. 7.716⁶, welches Handlungen, die den Nationalsozialismus fördern, ebenfalls unter Strafe stellt.

⁶ Gesetz Nr. 7.716 vom 5. Januar 1989. Es bestimmt die aus Vorurteilen aufgrund von Rasse oder Hautfarbe resultierenden Straftaten. Verfügbar unter: http://www.planalto.gov.br/ccivil_03/leis/l7716.htm. Letzter Zugriff am: 26.02.2021.

[Der Gesetzentwurf Nr. 7582/2014](#), der Hass- und Intoleranzverbrechen zum Gegenstand hat, erweitert den Umfang der Zielgruppen um soziale Klasse und Herkunft, Migranten-, Flüchtlings- oder Binnenvertriebenenstatus, sexuelle Orientierung, Geschlechtsidentität und -ausdruck, Alter, Obdachlosigkeit und Behinderung. Bis diese Kriterien in die Rechtsordnung aufgenommen werden, werden die entsprechenden Fälle von Hassreden normalerweise basierend auf Prinzipien der brasilianischen Bundesverfassung⁷, wie Menschenwürde (Art. 1) und Rechtsgleichheit (Art. 5), die u.a. die Geschlechtergleichheit und den menschenwürdigen Umgang einschließt, verhandelt.

Es soll nicht ungeachtet bleiben, dass das brasilianische Strafgesetzbuch⁸ Beleidigung als Straftat einordnet, während der „Bezug auf Rasse, Hautfarbe, Ethnie, Religion, Herkunft, Alter oder Behinderung“ gemäß dem Gesetz Nr. 10.741⁹ einen erschwerenden Umstand darstellt. Eine weitere wichtige Rechtsvorschrift für die Bekämpfung der Hassreden in Brasilien ist das Gesetz Nr. 12.735¹⁰. Es erweitert die Rechtsordnung, die die Strafbarkeit von Diskriminierungen bestimmt, indem es „Handlungen mittels der Verwendung von elektronischem, digitalem oder ähnlichem System“ einbezieht und somit die Bemühung um die Inkraftsetzung der funktionierenden Rechtsordnung im digitalen Raum aufweist.

Hinsichtlich der Bekämpfung diskriminierender Inhalte in der virtuellen Umgebung ist das *Zivilgesellschaftliche Rahmenabkommen des Internets*, welches mit der Verabschiedung

⁷ Verfassung der Föderativen Republik Brasilien. Brasília 1988. Verfügbar unter: http://www.planalto.gov.br/ccivil_03/constituicao/constituicaocompilado.htm. Letzter Zugriff am: 25.02.2021.

⁸ Verordnung Nr. 2.848 vom 7. Dezember 1940: Strafgesetzbuch. Verfügbar unter: http://www.planalto.gov.br/ccivil_03/decreto-lei/del2848compilado.htm. Letzter Zugriff am: 26.02.2021.

⁹ Gesetz Nr. 10.741 vom 1. Oktober 2003. Es entscheidet über das Statut der Ältere und erteilt weitere Vorschriften. Verfügbar unter: http://www.planalto.gov.br/ccivil_03/leis/2003/l10.741.htm. Letzter Zugriff am: 26.02.2021.

¹⁰ Gesetz Nr. 12.735 vom 30. November 2012. Es ändert die Verordnung Nr. 2.848 vom 7. Dezember 1940 - Strafgesetzbuch, die Verordnung Nr. 1.001 vom 21. Oktober 1969 - Militärstrafgesetzbuch und das Gesetz Nr. 7.716 vom 5. Januar 1989, um Handlungen mittels der Verwendung von elektronischem, digitalem oder ähnlichem System, die gegen elektronische und ähnliche Systeme vorgenommen werden, anzuordnen, und erteilt weitere Vorschriften. Verfügbar unter: http://www.planalto.gov.br/ccivil_03/ato2011-2014/2012/lei/l12735.htm. Letzter Zugriff am: 26.02.2021.

des Gesetzes Nr. 12.965¹¹ in Kraft trat, die wichtigste brasilianische Rechtsvorschrift. Obwohl der Artikel Nr. 2 des Gesetzestexts besagt, dass „Die Regulierung der Internetnutzung in Brasilien auf dem Respekt für die freie Meinungsäußerung basiert“, stellt der Einzelparagraph des Artikels Nr. 3 klar, dass „die in diesem Gesetz ausdrücklich angeführten Prinzipien diejenigen, die im anderen nationalen Gesetzrahmen oder in internationalen Abkommen, an denen Brasilien beteiligt ist, vorgesehen sind, nicht ausschließen“. In diesem Sinne vertritt Brasilien eine großzügigere sozialgerichtliche Auffassung, nach der die Gewährleistung der Meinungsfreiheit keine Grundlage für das freie Halten von Hassreden bildet.

In jüngster Vergangenheit begann im Kongress das Verfahren zum [Gesetzentwurf Nr. 2.630 vom 3. Juli 2020](#), welcher das brasilianische Gesetz für Freiheit, Verantwortung und Transparenz im Internet, das sogenannte „Fake-News-Gesetz“, einbringt. Der vom Senator Alessandro Vieira (Cidadania-SE) verfasste Gesetzentwurf wurde 2020 vom Bundessenat verabschiedet und wird momentan in der Abgeordnetenversammlung diskutiert. Einige Vorschläge zur Änderung des Originaltextes gehen auf das Thema der Hassrede ein, so z.B. der Vorschlag, der die Plattformen zur Veröffentlichung vierteljährlicher, öffentlich zugänglicher Berichte über die gelöschten Accounts und Inhalte verpflichtet und die Löschung von Inhalten in Fällen von Anstiftung zur Gewalt und unmittelbarem Schaden, der schwer zu kompensieren ist, ohne Warnung für die Nutzer anfordert. Derzeit wird das Vorhaben insbesondere aufgrund der Durchsetzung von Kontrollmechanismen und der möglichen Auswirkungen für das Online-Ökosystem kritisiert, wie vom [Kollektiv Intervozes](#) und der [Zeitschrift Piauí](#).

Die obengenannten juristischen Ansätze und Rahmenbestimmungen entsprechen also Elementen, die wertvoll für die Bekämpfung der Entstehung und Verbreitung von Hassreden in den Online-Interaktionen im brasilianischen Kontext sind. Die bereits existierenden sowie vorgeschlagenen rechtlichen Mechanismen bevorzugen die

¹¹ Gesetz Nr. 12.965 vom 23. April 2014. Es bestimmt Prinzipien, Garantien, Rechte und Pflichten der Internetnutzung in Brasilien. Verfügbar unter: http://www.planalto.gov.br/ccivil_03/_ato2011-2014/2014/lei/l12965.htm. Letzter Zugriff am: 26.02.2021.

Herangehensweise, dass die freie Meinungsäußerung kein absolutes Recht ist und somit in Einklang mit Grundrechten, welche durch Verfassungsprinzipien wie die Menschenwürde unterstützt werden, gewährleistet werden muss.

3. Hassrede in digitalen Umgebungen

Das Thema Hassrede ist viel älter als das Internet, doch die Ausbreitung der digitalen Medien und die intensive Nutzung der digitalen sozialen Netzwerke bringen neue, spezifische Herausforderungen bei der Bekämpfung von Hassrede mit sich. Noch bis vor kurzem wurden die Praktiken der Online-Hassrede als eine Nischenaktivität betrachtet. In den letzten Jahren jedoch machte ihre hervortretende, häufige Präsenz in den Mainstream-Internetplattformen¹² dieses Thema immer sichtbarer. Durch die Auswirkungen koordinierter Aktivitäten im Bereich der digitalen Hassrede, die zu realen Angriffen führen, verschärft sich die Situation auf Offline-Ebene (SIEGEL 2020). Der [Fall in Suzano](#) in der Metropolregion São Paulo veranschaulicht dies¹³. In diesem Abschnitt wird diskutiert, welche neuen Nuancen die Hassrede annimmt, wenn sie sich in digitalen Umgebungen ausdrückt. Dazu werden zunächst auf einige Merkmale der digitalen Umgebungen, die zur Entstehung und Verbreitung von Hassreden beitragen können, unter Berücksichtigung der Spezifitäten der digitalen Medien und der davon ausgehenden

¹² Die Bezeichnung „Mainstream-Internetplattformen“ betrifft die digitalen Plattformen, die am meisten von Brasilianern benutzt werden und weder Nischen für sehr spezifische Interessen noch digitale Räume wie die Imageboards für die Organisation von Hassgruppen darstellen. Laut einer von der Agentur We are Social in Zusammenarbeit mit dem Unternehmen Hootsuite durchgeführten Forschung waren Facebook, WhatsApp, YouTube, Instagram, Facebook Messenger, LinkedIn, Pinterest, Twitter, TikTok und Snapchat, in absteigender Reihenfolge, die zehn im Jahr 2020 von Brasilianern am meisten benutzten sozialen Netzwerke.

¹³ Amoklauf am 13.03.2019 an der Schule Professor Raul Brasil in der Stadt Suzano, Metropolregion São Paulo. Zwei ehemalige Schüler der betroffenen Schule, Guilherme Monteiro und Luiz Henrique de Castro, erschossen fünf Schüler, zwei Angestellte und Monteiros Onkel. Anschließend erschoss Monteiro seinen Komplizen, bevor er Suizid beging. Beide waren Nutzer eines Imageboards, eines anonymen Forums, der umgangssprachlich Chan genannt wird, in dem sie den Amoklauf planten und Anweisungen dafür erhielten. Die Beteiligung von Hass-Communities an den Chans und deren Auswirkungen auf reale Situationen wurden in der brasilianischen Presse breit diskutiert.

gesellschaftlichen Organisationsformen hingewiesen. Die Strukturen der sozialen Netzwerke erschweren die Einschränkung der Verbreitung derartiger Narrative, was eine Herausforderung sowohl für die nationalen Gesetze als auch für die Selbstregulierungen der Plattformen darstellt, wie im folgenden Abschnitt diskutiert.

Um die Thematik angemessen anzugehen, müssen zunächst die Unterschiede zwischen Online- und Offline-Hassrede nachvollzogen werden. Systematische Übersichtsarbeiten deuten darauf hin, dass einige Merkmale digitaler Medien wichtige Bestandteile für deren Verbreitung sind: Sie ermöglichen Anonymität, Unsichtbarkeit, Schaffung geographisch barrierefreier Affinity-Communities (einschließlich Hasses), geringen zeitlichen und finanziellen Aufwand für die Verbreitung derartiger Diskurse und Unmittelbarkeit (BROWN 2018).

Die Kombination dieser Faktoren bietet Hassgruppen neuartige Möglichkeiten und verstärkt dadurch die Häufigkeit von Diskriminierung im Internet, was das Auftreten solcher Diskurse teilweise erklärt. Die Anonymität der Nutzer, wenn auch nur teilweise (Deep Web ausgenommen), drückt sich insbesondere in der Beseitigung von Mitteln für die Rechenschaftspflicht bzgl. Hasskriminalität im Online-Kontext und in der geringen Wahrscheinlichkeit einer Reaktion seitens des Opfers oder einer körperlich gewaltvollen Auseinandersetzung zwischen diesem und dem Aggressor aus. Die Unsichtbarkeit, die durch die visuelle Abwesenheit des Aggressors und des Opfers entsteht, erleichtert die Durchführung der Angriffe, da die Auswirkungen ebendieser auf das Opfer nicht sichtbar für den Aggressor sind. Dementsprechend *kann* die Online-Hassrede aufgrund ihrer teilweisen Unsichtbarkeit und Anonymität weniger real und gegebenenfalls folgenlos *erscheinen* (BROWN 2018).

Die Rolle des Internets bei der Schaffung von Affinity-Communities, die auf Angehörigkeitsgefühlen gegenüber gemeinsamen Merkmalen oder Interessen seitens Menschen unterschiedlicher Herkunftsorten basieren, hat auch Folgen auf die Verbreitung von Hassrede. Der einfache Zugang zu digitalen Kommunikationsressourcen, der sowohl bei der Organisation als auch der Verbreitung von Hassreden zeitlich und finanziell

ressourcenschonend ist, ist ebenfalls von Bedeutung (BROWN 2018). Hassgruppen verwenden das Internet als ein systematisches Mittel für die Anwerbung von Mitstreitern durch Gemeinschaften und Werte, indem sie Gemeinschaften und Werte schaffen, die im digitalen Kontext identitätsstiftend und -stärkend wirken (WEAVER 2013). In diesen Räumen ist es möglich, ein Gemeinschaftsgefühl unter den Nutzern zu erzeugen und zu hegen, welches die Verbundenheit und die Identitäten mit einer potenziellen Zunahme der Hassgruppen durch das Engagement ihrer Mitglieder stärkt (BOWMAN-GRIEVE 2009). Dieses Engagement lässt sich in den bereits erwähnten Hasskampagnen in sozialen Netzwerken beobachten, wie bei der Koordinierung von gleichzeitigen Bemühungen seitens der Nutzer um Äußerungen zu einem selben Thema oder Angriffe auf einen bestimmten Akteur.

Schließlich besteht auch die Möglichkeit impulsiver Aktionen, die gelegentlich aufgrund mangelnder Zeit für Reflexion erzeugt werden. In der traditionellen Presse wird diese Zeit üblicherweise durch die Transaktionskosten der Produktion und Verbreitung derartiger Diskurse, wie z.B. die Herstellung, Druck und Verteilung eines Pamphlets, bestimmt. Derzeit reicht es aus, das Handy zu nehmen, die Twitter-App zu öffnen, einen hasserfüllten Tweet zu schreiben und „Senden“ zu drücken. Somit fördert die Funktionsweise des Internets das spontane Wesen von Hassrede (BROWN 2018).

Ein anderes Merkmal der digitalen Plattformen, das zur Verbreitung von Hassreden beitragen kann, ist die auf Algorithmen basierende Funktionsweise, deren Verwaltung der Sichtbarkeit von Inhalten in direkter Verbindung mit den Nutzereinstellungen steht (GILLESPIE 2018), was einige Autoren Echokammern (echo-chambers) nennen¹⁴. Durch die Bevorzugung von der Meinung des Nutzers entsprechenden Inhalten auf dessen *Timeline* können die Algorithmen auch eine verzerrte Wahrnehmung der sozialen Szenarien, neben der Polarisierung aufgrund der Entstehung ideologischer Blasen, verursachen. Die Funktionsweisen der Algorithmen, die diese personalisierten Sichtbarkeitseinstellungen von Inhalten erzeugen, werden nicht vollständig von den

¹⁴ Eine gegensätzliche Auffassung vertreten Autoren, die die Idee der selektiven Exposition und der unbeabsichtigten Exposition verteidigen. D.h. Individuen werden Inhalten, die ihren Meinungen entgegensetzen, ausgesetzt. Für mehr Informationen über diese Debatte siehe Barberá (2015).

digitalen Plattformen erklärt und stellen undurchschaubare Systeme für die meisten Nutzer dar (JURNO/D'ANDREA 2017).

Durch das Zusammenbringen von Menschen mit ähnlichen Gedanken, Haltungen und Meinungen kann die Algorithmenlogik der sozialen Netzwerke demnach Nutzer zu einem tendenziösen Verstand des in der Gesellschaft vorherrschenden Meinungsklimas bewegen: „Wenn ich in meinen sozialen Netzwerken sehe, dass die meisten Menschen bestimmte soziale Verhaltensweisen akzeptieren und gelegentlich gutheißen, neige ich dazu, meine Verhaltensweise als eine sozial angemessene Norm zu betrachten.“ Ausgehend von diesem Gedanken entwickelt sich das Risiko, dass bestimmte Communities Hassreden als sozial akzeptable Inhalte ansehen, wenn sie auf ihren Timelines damit konfrontiert werden. Folglich tendieren die Menschen dazu, selber Hassreden zu verbreiten, da sie durch derartige Verhaltensweisen in digitalen Netzwerken soziale Belohnungen (Likes, Teilungen, Interaktion usw.) erhalten.

Im Gegensatz zur Inhaltskontrolle in Massenmedien, in denen die Überprüfung vor der Veröffentlichung stattfindet, werden die Diskurse der Nutzer in digitalen Plattformen erst nach der Veröffentlichung und in der Regel nach Meldung anderer Nutzer kontrolliert. Dies ist auch auf Besonderheiten des digitalen Kontexts zurückzuführen: Der Umfang drastischer Inhalte, die von Tausenden Nutzern täglich verbreitet werden, verhindert die Vorabkontrolle. Darüber hinaus bleibt die Idee der Meinungsfreiheit der bedeutsamste Wert digitaler Plattformen, wie im folgenden Abschnitt erläutert. Daraus folgt, dass die Hassrede in Mainstream-Online-Plattformen erst zum Ausdruck kommt und nach ihrer Verbreitung bekämpft wird. Im Offline-Kontext hingegen wird die Hassrede meistens ausgeschlossen von der etablierten Presse und eher in randständigen Kreisen verbreitet.

Ein weiterer bedeutender Unterschied in den Verbreitungsformen zwischen den digitalen Plattformen und den Kanälen der etablierten Presse liegt darin, dass die Unternehmen im ersten Fall, im Gegensatz zu denjenigen im zweiten Fall, selbst keine Inhalte erstellen. Aus diesem Grund reicht es nicht aus, die Formen von Funktionsweisen, Kontrolle und Bekämpfung der Hassrede in beiden Kontexten lediglich zu vergleichen. In diesem Sinne

unterscheiden sich ebenfalls die Rechenschaftspflichten in beiden Fällen. Obwohl die digitalen Plattformen keine diskursiven Inhalte produzieren, schaffen sie eine Umgebung, in der diese Diskurse verbreitet werden, und vor allem profitieren sie finanziell von der Interaktion und Aufmerksamkeit der Nutzer. Insofern sind die digitalen Plattformen auch nicht mit üblichen öffentlichen Räumen vergleichbar. Sie sind Unternehmen und daher private Orte, die in gewissem Maße Verantwortung für das, was dort geschieht, tragen. Dieser Status, der weder der Inhaltserstellung der traditionellen Medien noch dem komplett öffentlichen Diskussionsraum entspricht, macht die Lösungen für die Bekämpfung der Hassrede (und anderer Problematiken) komplexer.

Vorherige Studien zeigen, dass die Anzahl von Hassreden auf den Mainstream-Plattformen im Verhältnis zur Gesamtheit der Inhalte klein ist. Doch die Strategien zur Verbreitung dieser Diskurse erzeugen Sichtbarkeit, sodass ein großer Teil der Personen im Internet bereits derartigen Inhalten ausgesetzt war. Communities, die sich für auf Hassrede basierte Angriffe organisieren, zielen vor allem auf öffentliche Personen wie Künstler, Politiker und Journalisten ab, agieren in koordinierter Form und erweitern somit die Reichweite ihrer Aktionen (SIEGEL 2020).

Aktuell weisen die systematischen Übersichtsarbeiten jedoch darauf hin, dass die meisten Studien zum Thema Online-Hassrede den Schwerpunkt auf Communities, die sich in Nischenplattformen organisieren, legen. Deshalb ist der Wissenstand über die Akteure, die Hassrede auf informaler Art und Weise - insbesondere in den meistbenutzten Plattformen - verbreiten, noch begrenzt. Mit dem Ziel, diese Wissenslücke zu füllen, stellt der vorliegende Bericht eine zeitliche Analyse der Hassrede auf den Plattformen Twitter und Facebook im brasilianischen Kontext aus. Laut Daten einer 2020 [von We are social und Hootsuite durchgeführten Forschung](#) gibt es in Brasilien 130 Mio. Facebook-Nutzer und 16,6 Mio. Twitter-Nutzer. Zusammen mit YouTube und Instagram bilden sie die wichtigsten Räume der brasilianischen digitalen Umgebung für die offenen Meinungsäußerungen und die öffentliche Debatte.

Es ist wichtig, die Definition von Hassrede von der Kategorisierung der Arten von Hassrede abzugrenzen. Die Definition nähert sich dem Konzept an, ist daher abstrakt und kann unterschiedliche spezifische Kontexte umfassen. Die Kategorisierung hingegen ist der Prozess zur Bestimmung dessen, was eine Hassrede ist, und dessen, was keine Hassrede ist. Er geht von einem realen Ereignis aus und basiert auf einer konzeptuellen Definition (juristisch, akademisch oder auf die Funktionsweise der Plattformen bezogen). Demzufolge kann eine Definition von Hassrede, die konzeptuell klar zu sein scheint, unklar in konkreten Situationen werden, da diese von Kontext, Sprachgebräuchen, Handlungsweisen bestimmter Kulturen und sogar sprachlichen Aneignungen und Bedeutungen jeder Community abhängen.

Diese situationsbedingte Dynamik ist es, die die Erkennung und Bekämpfung der Hassrede auf den Plattformen deutlich schwieriger als die Konzeptualisierung macht. Auf digitalen Plattformen wie Twitter und Facebook wird die Hassrede vorwiegend auf Personen anhand ihrer spezifischen Eigenschaften wie sexueller Orientierung, sozialer Klasse, Geschlecht, Ethnie und körperlicher Merkmale abgezielt. Einige dieser Eigenschaften werden von den Plattformen als schutzbedürftig verstanden und sind in Einklang mit rechtlichen Vorschriften, obwohl sie nicht immer auf beiden Ebenen genau miteinander übereinstimmen.

4. Richtlinien der digitalen Plattformen bezüglich Hassrede

Im Online-Kontext weist die Hassrede Spezifitäten auf, wie diejenigen, die im vorherigen Abschnitt betrachtet wurden. Außer der Besonderheiten, die sich aus dem Kommunikationsmittel selbst ergeben, sind die Interaktionen auf jeder Plattform auch von den jeweiligen Nutzungsbedingungen geprägt. Die Nutzungsbedingungen eines Online-Diensts sind Dokumente, die über die Funktionsweisen der Plattformen und die Regeln für die Nutzer informieren. Wenn die Menschen einen Dienst nutzen, sind sie

diesem unterstellt und mit dessen Nutzungsbedingungen einverstanden, selbst wenn die Nutzer diesen Dienstvertrag zumeist nicht lesen¹⁵. Social-Media-Plattformen weisen normalerweise in den Nutzungsbedingungen darauf hin, dass die Nutzer den Gemeinschaftsrichtlinien folgen müssen und dass bei Nichteinhaltung ebendieser einige Strafen gegen sie verhängt werden können. Aufgrund dessen werden darauffolgend die Gemeinschaftsrichtlinien der in Brasilien meistbenutzten digitalen Plattformen, in denen die Definitionen von Hassrede je nach Plattform zu finden sind, beschrieben, analysiert und miteinander verglichen.

Die Nutzungsbedingungen und die Gemeinschaftsrichtlinien - ein Regelwerk, das die von den Nutzern erwartete Verhaltensweise bestimmt - sind Dokumente, die bei den verschiedenen Plattformen Unterschiede aufweisen. In Bezug auf Hassrede, schädliche oder gewaltsame Diskurse erläutern diese Dokumente das jeweilige Verständnis des sozialen Netzwerks von jedem dieser Begriffe, geben Auskunft über die Toleranz im Falle möglicher Verstöße und über die Art und Weise, wie die Erwartungen der Nutzer bzgl. der freien Meinungsäußerung sowie zu ihrer Sicherheit, vor allem derjenigen, die zu sogenannten geschützten Kategorien gehören. Auch, wenn die von den Plattformen entwickelten Maßnahmen gelegentlich auf wissenschaftlichen Forschungen basieren, wie die von Facebook und Twitter, gibt es kein gemeinsames, standardisiertes Verständnis unter den sozialen Netzwerken.

Trotz der spezifischen Bemühungen um die Erzeugung automatisierter Mittel für die Erkennung von Hassreden in sozialen Netzwerken findet die von den analysierten Plattformen bevorzugte Funktionsweise durch die Meldungen anderer Nutzer statt. Das heißt, wenn ein Nutzer einem Posting begegnet, das gegen irgendeinen Aspekt der Gemeinschaftsrichtlinien verstößt, kann er den Inhalt melden, der anschließend von Moderatoren analysiert wird. Seit 2013 sind Twitter, Facebook und YouTube Unterzeichner

¹⁵ Die Zustimmung in digitalen Umgebungen basiert auf dem „Notice-and-Choice“-Prinzip, d. h. die Dienste informieren die Nutzer über die potenziellen Risiken, die mit der Nutzung der Plattform verbunden sind, und die Nutzer drücken ihre Zustimmung und ihr Einverständnis aus, indem sie auf ein „Ich stimme zu“ klicken. Mehr zu diesem Prinzip und seinen Auswirkungen auf die Debatte über Privatsphäre in digitalen Umgebungen, siehe Sloan und Warner (2014).

des Abkommens zur Bekämpfung von Hassrede, das von der amerikanischen Non-Profit-Organisation *Anti-Defamation League* ins Leben gerufen wurde. Ausgehend vom Dokument [Best Practices for Responding to Cyberhate](#) verpflichten sich die drei digitalen Plattformen zu einigen Praktiken, darunter: Meldungen und Berichte über Hassrede verbindlich und rechtzeitig zu analysieren; Den Nutzern auf eine verständliche Art und Weise zu erklären, wie die Inhaltsmoderation durchgeführt wird, und vorgesehene Sanktionen konsequent und gerecht einzuleiten; vereinfachte Formen für die Meldung von hasserfüllten Inhalten zu bieten (SILVA et al. 2019).

Dieses Abkommen ist eine der Triebfedern der 2015 von den digitalen Plattformen eingeführten Änderungen und Verbesserungen bei der Bekämpfung hasserfüllter Inhalte. Deutlichere Nutzungsbedingungen und Gemeinschaftsrichtlinien, Einführung von Moderationsberichten und Entwicklung automatisierter proaktiver Techniken zur Ermittlung von Hassreden sind einige der Änderungen, die im Lauf der Zeit beim Umgang seitens der digitalen Plattformen mit diesem Thema zu beobachten sind. Weitere Faktoren trugen ebenso zu diesen Verbesserungen bei, wie beispielsweise die neuen Gesetze, die Online-Hassinhalte (wie in Deutschland), Leaking von Daten, die dann öffentlich werden, durch Online-Gruppen geplante Terroranschläge und Regierungsinterventionen regulieren (SILVA et al. 2019).

Im Folgenden wird eine Zusammenfassung der aktuellen Gemeinschaftsrichtlinien von [Facebook](#), [Twitter](#), [Instagram](#) und [YouTube](#) vorgestellt, um die Konzeptualisierungen und Herangehensweisen bzgl. der Hassrede und deren Variablen im Online-Kontext nachvollzuziehen. Trotz der legislativen Bemühungen, sowohl auf nationaler als auch internationaler Ebene, und der Gemeinschaftsrichtlinien und Nutzungsbedingungen der Plattformen sind die Verhältnisse zwischen dem Recht auf freie Meinungsäußerung jedes Individuums und den Hassreden nicht immer eindeutig. Wie bereits weiter oben erläutert, unterscheiden sich die Auffassungen diesbezüglich teilweise erheblich. Das Gleichgewicht dieses Verhältnisses ist eine der Hauptfragen, die sich die Plattformen bei der Herangehensweise an das Thema stellen müssen: Wie ist die Sicherheit der zu

schützenden Gruppen mit möglichst geringem Eingriff in die Meinungsfreiheit der Nutzer zu gewährleisten?

Darüber hinaus stellen die Schwierigkeiten, die durch die Vielzahl an Kontexten, in denen Hassrede stattfindet, entstehen, eine Herausforderung dar, die Twitter, Facebook, Instagram und YouTube auf ihren Websites formulieren: Ausdrücke, denen historisch eine erniedrigende Bedeutung innewohnt, können durch andere soziale Gruppen umgedeutet werden, wobei neue Formen von verbalem Angriff oder Aufhetzung in entschlüsselten Formaten entstehen können. Folglich ist die Kultur ein Schlüsselement bei der Identifizierung derartiger Diskurse.

Allgemeiner Überblick

Facebook, Twitter, Instagram und YouTube gehören zu den zehn digitalen Plattformen, die von Brasilianern für die Erstellung und den Konsum von Inhalten am meisten benutzt werden. Die Untersuchung der Nutzungsbedingungen und Gemeinschaftsrichtlinien der vier Plattformen ließ erkennen, dass ihre Richtlinien eine gewisse Einschränkung oder Ablehnung gegenüber der Hassrede aufweisen. In einigen Fällen werden auch Richtlinien gegen gewaltsame, extremistische oder gefährliche Diskurse mit einbezogen. Der Detaillierungsgrad, die Konzeptualisierung der Begriffe und die Erläuterung der angegebenen Richtlinien variieren jedoch zwischen den jeweiligen Plattformen.

In diesem Sinne präsentiert sich Instagram als die Plattform, die diese Aspekte am wenigsten detailliert darstellt und ihre [Gemeinschaftsrichtlinien](#) prägnanter und allgemeiner formuliert. Allerdings informiert sie die Nutzer darüber, dass sie auch den Politiken von Facebook unterliegen, da beide Plattformen zum selben Unternehmen gehören. Der Mangel an Ausführlichkeit ist ein Hinweis auf eine noch größere Schwierigkeit bei der Moderation, die aufgrund der Herausforderungen des Kontexts und der Meinungsfreiheit besonders komplex ist.

Ein weiterer Aspekt, der bei der Analyse der Plattformen von Bedeutung ist, ist die Erwähnung der zu schützenden Gruppen aufgrund spezifischer Merkmale, die aufgelistet sein können oder nicht. Auf Instagram sind sie nicht definiert. Im Gegensatz dazu stellen Facebook, Twitter und YouTube eine Liste zur Verfügung, auf der die als geschützt geltenden Kategorien oder Gruppen einbezogen sind, mit Angaben zu Alter, Geschlecht, sexueller Orientierung, Ethnie, „Rasse“, Religion und Migrationshintergrund. Einige davon erscheinen aber nicht auf allen Plattformen, sondern ausschließlich auf einer oder zwei Plattformen. Facebook und Twitter beziehen körperliche Merkmale oder Krankheiten mit ein, Twitter deutet auf marginalisierte, historisch unterrepräsentierte Gemeinschaften und YouTube betrachtet auch soziale Klasse und Kriegsveteranen als geschützte Kategorien und Gruppen. Es lässt sich beobachten, dass drei der aufgeführten Diskriminierungsmerkmale, die von allen Plattformen konsensual den zu schützenden Gruppen zugeordnet werden, mit den Gruppen assoziiert werden, die laut Literatur am häufigsten Opfer von Hassreden werden: sexuelle Orientierung, Geschlecht und Ethnie (SILVA et al. 2016). Jedoch werden zwei der Diskriminierungsmerkmale, die in der Literatur als vulnerabel aufgelistet sind, lediglich von einer der vier Plattformen, YouTube, miteinbezogen: körperliche Eigenschaften und soziale Klasse (SILVA et al. 2016).

Weiterhin ist YouTube die einzige der vier Plattformen, die in ihren [Gemeinschaftsrichtlinien](#) darüber informiert, dass sie, neben den Meldungen von ungeeignetem Inhalt seitens der Nutzer, eine eigene Überwachung mithilfe eines Systems automatischer Signalisierung für die Erkennung von Inhalten, die gegen die Gemeinschaftsrichtlinien verstoßen, durchführt. Möglicherweise steht diese größere Sorgfalt von YouTube bei der Überprüfung der geposteten Inhalten in Verbindung mit der Monetarisierung von Videos, weil Videos mit beleidigendem Inhalt auf der Plattform nicht monetarisiert werden dürfen. Die anderen Plattformen informieren lediglich darüber, dass sie die von Nutzern gemeldeten Inhalte nur in Bezug auf den Kontext analysieren, ohne auf weitere Initiativen für die Kontrolle und Monitoring der Umgebung in den Gemeinschaftsrichtlinien hinzudeuten. Es ist jedoch bekannt, dass sowohl Facebook als

auch Twitter automatisierte Mechanismen für die Erkennung von Inhalten, die gegen ihre Gemeinschaftsrichtlinien verstoßen, verwenden (SILVA und al. 2019).

Im Folgenden werden die Gemeinschaftsrichtlinien von Facebook und Twitter bezüglich der Hassrede und derartiger Themen ausführlicher betrachtet, da die Erfassung der Daten, die die Analysen des vorliegenden Dokuments erläutern, in den Umgebungen dieser beiden Plattformen realisiert wurde.

	Facebook
<p>In den Gemeinschaftsrichtlinien hebt Facebook die „Verpflichtung zur Meinungsfreiheit“ als „höchste Priorität“ hervor, was in Einklang mit der amerikanischen Perspektive der Wahrung der Meinungsfreiheit steht. Ausgehend vom Argument der Meinungsfreiheit rechtfertigt die Plattform ihre Richtlinien gegenüber der Hassrede, um „sicherzustellen, dass die Meinung aller Seiten gehört wird“, einschließlich der sogenannten „geschützten Kategorien“.</p> <p>Facebook definiert Hassrede als „direkten Angriff auf Personen aufgrund geschützter Eigenschaften: <i>ethnische Zugehörigkeit, nationale Herkunft, Behinderung, religiöse Zugehörigkeit, Kaste, sexuelle Orientierung, Geschlecht, Geschlechtsidentität und ernsthafte Erkrankung</i>“. Daraufgehend bezeichnet der Text Angriffe „als gewalttätige oder menschenverachtende Sprache, schädliche Stereotypisierung, Aussagen über Minderwertigkeit, Ausdrücke der Verachtung, der Abscheu oder Ablehnung, Beschimpfungen oder Aufrufe, Personen auszugrenzen oder zu isolieren.“ In diesem Sinne gelingt es der Plattform, ein klares Verständnis von Hassrede zu vermitteln, auch wenn in der Praxis die Einschätzung dieses Verständnisses durch die Moderatoren von kontextuellen und möglicherweise sogar persönlichen Faktoren abhängt. Durch den Hinweis auf die genannten Merkmale wird auch auf die Gruppen, die von derartigen</p>	

Angriffen häufiger betroffen sind, und auf den Schwerpunkt der Moderation hingewiesen.

Die Richtlinien drei verschiedene Ebenen der Hassrede auf: Die erste Ebene umschließt den gewaltsamen Diskurs, erniedrigende Vergleiche, Spott und Negationismus historischer Situationen bezüglich der zu schützenden Kategorien; Die zweite Ebene handelt von Aussagen über Minderwertigkeit und Ausdrücke der Verachtung der zu schützenden Kategorien; Die dritte Ebene betrifft die Verhaltensweisen, die zur Ausgrenzung oder Isolierung der geschützten Kategorien aufrufen. Auf einem [separaten Tab](#) wird zudem angegeben, dass Facebook weder Inhalte, die Gewalt verherrlichen und preisen, noch explizite Inhalte in jeglicher Form zulässt.



Twitter

Twitter informiert über die unzulässigen Verhaltensweisen auf der Plattform im Abschnitt „Regeln und Richtlinien“ und erstellt drei für die vorliegende Studie interessante Kategorien: „Richtlinie zur Gewaltverherrlichung“, „Richtlinie zu Hass schürendem Verhalten“ und „Richtlinie zu Gewaltandrohungen“. Von den vier sozialen Netzwerken ist Twitter dasjenige, das das ausführlichste Dokument darüber anbietet, was erlaubt oder nicht erlaubt ist und wie es mit dem Thema umgeht.

Die [Richtlinie zur Gewaltverherrlichung](#) erklärt, dass dem Nutzer „nicht erlaubt ist, Gewaltverbrechen sowie Gewalttätigkeiten, die sich gegen Personen aufgrund ihrer Zugehörigkeit zu einer vulnerablen Gruppe richten, oder die Gewalttäter selbst, zu verherrlichen, zu bejubeln, zu loben oder zu billigen“, gefolgt von einer Definition davon, was unter Verherrlichung zu verstehen ist, und einer Auflistung von Beispielen für solche Handlungen. Diese Kategorie der Plattform ist direkt mit dem, was hier über gefährliche Diskurse diskutiert wurde, assoziiert.

Die [Richtlinie zu Hass schürendem Verhalten](#) legt den Schwerpunkt auf Gruppen, die online verstärkt Missbrauch ausgesetzt sind, insbesondere wenn intersektionale Eigenschaften bestehen: „Frauen, Farbige, Homosexuelle beider Geschlechter, Bisexuelle, Transgender-, Queer- und Intersex-Personen, asexuelle Personen, ausgegrenzte und historisch unterrepräsentierte Gemeinschaften“. Die Richtlinie besagt, dass der Nutzer aufgrund der genannten Eigenschaften „keine Gewalt gegen andere Personen fördern, sie direkt angreifen oder ihnen drohen“ darf. Außerdem verbietet sie die Verwendung von Hass schürenden Bildern wie denjenigen, die historisch mit Hassgruppen verknüpft sind, sowie von Inhalten, die zu Gewalt gegen vulnerablen Gruppen aufstacheln.

Die [Richtlinie zu Gewaltandrohungen](#) erläutert, dass der Nutzer „einer Person oder einer Personengruppe keine Gewalt androhen“ darf, insbesondere dann nicht, wenn eine explizite Absicht, beispielsweise vom sexuellen Missbrauch, geäußert wird. Gemäß den Dokumenten wertet die Plattform einige dieser Ausdrücke, wie verschiedene Varianten von „ich bringe dich um“, in Bezug auf den Kontext aus, um rhetorische Figuren auszuschließen.

Öffentliches Interesse

Die vier Plattformen erkennen an, dass sie, unter besonderen Umständen des öffentlichen Interesses oder der Rolle bei der Bewusstseinsbildung und Aufklärung über das Thema, Inhalte, die sich den Gemeinschaftsstandards bzgl. Hassreden entgegensetzen, erlauben können, auch wenn sie Meldungen erhielten. Beispiele dafür sind Postings, die die Bildung vom öffentlichen Bewusstsein für ein Thema und die öffentliche Diskussion darüber beabsichtigen, sofern die Absicht des Nutzers in dieser Hinsicht deutlich ist.

Sanktionen bei Verstoß gegen die Richtlinien

Keine der vier betrachteten Plattformen spezifiziert auf eine klare oder transparente Art und Weise die Verfahren rund um die Sanktionen, die gegen Nutzer, die Inhalte von Hassrede o.ä. veröffentlichen, verhängt werden. Sie beschreiben nur vage, was mit denjenigen, die gegen die Nutzungsrichtlinien verstoßen, passiert: Instagram ist die Plattform, die am wenigsten Details zu Sanktionen und Verfahren angibt und YouTube ist diejenige, die in dieser Hinsicht tendenziell die meisten Informationen anbietet. Diese Transparenz ist selbstverständlich nur teilweise gegeben, da die Plattform diejenige ist, die uneingeschränkten Zugang zu den Daten sowohl über ihre Entscheidungen als auch ihre Berichte bietet.

Facebook gibt an, dass es den Schweregrad und das Verhalten des Nutzers auf der Plattform analysiert, um eine Entscheidung zu treffen, die von Verwarnungen, Nutzungseinschränkungen, bis zur vollständigen Deaktivierung des Profils reichen kann. Die Absicht des Benutzers spielt eine Rolle bei der Bewertung, sowohl für das Entfernen des Inhalts als auch für die Sanktion. Im Fall von Gefahr für Leib und Leben einer Person oder für die öffentliche Sicherheit werden die Strafverfolgungsbehörden informiert.

Twitter behauptet, nur das Verhalten der Nutzer auf der Plattform zu berücksichtigen, wobei ein erster Verstoß gegen die Richtlinien zur Löschung des Inhalts und potenziell zur vorübergehenden Sperrung des Accounts führen kann. Wiederholte Verstöße führen zur dauerhaften Sperrung des Accounts - allerdings gibt Twitter die Anzahl der nötigen Verstöße nicht an.

Instagram informiert, dass Verletzungen der Richtlinien zur Löschung von Inhalten und gegebenenfalls zur Kündigung der Konten führen, ohne irgendwelche Kriterien, die als Grundlage für diese Entscheidung dienen, zu nennen. YouTube legt fest, dass einem ersten Verstoß gegen die Community-Richtlinien keine Sanktion, sondern nur eine Warnung folgt.

Wiederkehrende Verstöße führen zur Löschung des Inhalts (Video oder Kommentar), sowie zu einer Verwarnung, welche [die Einschränkung der Aktionen](#) des Nutzers auf der Plattform für eine oder zwei Wochen vorsieht. Nach Erhaltung von drei Verwarnungen innerhalb von 90 Tagen werden der Kanal und seine Videos endgültig entfernt.

Obwohl die Plattform nicht erklärt, was bei der Analyse der Meldungen berücksichtigt wird, geht aus der Beschreibung der Verfahren hervor, dass sie den Schweregrad und die Häufigkeit der Verstöße beachtet, um Maßnahmen zu ergreifen. Außerdem stellt YouTube einen leicht zugänglichen [Transparenzbericht](#) über die Sanktionen bezüglich der Verletzungen der Gemeinschaftsrichtlinien zur Verfügung. Laut dem Bericht wurden zwischen Oktober 2020 und Dezember 2020 2.055.515 Kanäle gelöscht. Dabei stellte die Kategorie „Hasserfüllter oder beleidigender Inhalt“ den zweithäufigsten Entfernungsgrund dar, der 8,3% betrug. Der häufigste Entfernungsgrund war „Spam, irreführende oder betrügerische Inhalte“, der über 75% der Gesamtheit entsprach. Die folgende Tabelle systematisiert die wichtigsten Ergebnisse der vergleichenden Untersuchung der Gemeinschaftsrichtlinien von Facebook, Twitter, Instagram und YouTube.

Tabelle 1 - Vergleichsrahmen der Gemeinschaftsrichtlinien von Facebook, Twitter, Instagram und YouTube

Vergleichsrahmen - Gemeinschaftsrichtlinien				
	Facebook	Twitter	Instagram	YouTube
Richtlinien über Hassrede	Ja	Ja	Ja	Ja
Definition von Hassrede	„[...] direkter Angriff auf Personen aufgrund geschützter Eigenschaften.“	aufgrund geschützter Eigenschaften „Gewalt gegen andere Personen fördern, sie direkt angreifen oder ihnen drohen [...]“.	Nicht angegeben.	Inhalte, die basierend auf den geschützten Eigenschaften „[...] zu Gewalt oder Hass gegen Einzelpersonen oder Gruppen aufrufen“.

Geschützte Kategorien und Gruppen	ethnische Zugehörigkeit, nationale Herkunft, Behinderung, religiöse Zugehörigkeit, Kaste, sexuelle Orientierung, Geschlecht, Geschlechtsidentität und ernsthafte Erkrankung.	Kategorien: Abstammung, ethnische Zugehörigkeit, nationale Herkunft, Kastenzugehörigkeit, sexuelle Orientierung, Geschlecht, Geschlechtsidentität, religiöse Zugehörigkeit, Alter, Behinderung oder eine schwere Erkrankung. Personengruppen: Frauen, Farbige, Homosexuelle beider Geschlechter, Bisexuelle, Transgender-, Queer- und Intersex-Personen, asexuelle Personen, ausgegrenzte und historisch unterrepräsentierte Gemeinschaften.	Keine Kategorien oder Gruppen werden als geschützt beschrieben, aber „Angriffe gegen bestimmte Personen aufgrund ihrer Rasse, ethnischen Herkunft, Nationalität, ihres Geschlechts oder ihrer geschlechtlichen Identität, ihrer sexuellen Orientierung, religiösen Zugehörigkeit oder aufgrund von Behinderung oder Krankheit“ werden „keinesfalls“ zugelassen.	Alter, Gesellschaftsklasse, Behinderung, Ethnische Herkunft, Geschlechtsidentität und -ausdruck, Nationalität, Rasse, Einwanderungsstatus, Religion, Geschlecht, sexuelle Orientierung, Opfer von schwerwiegenden Gewaltereignissen und deren Angehörige, Veteranenstatus.
Bewertungskriterien der Meldungen	Schweregrad, Absicht und Verhalten des Nutzers auf der Plattform.	Verhalten des Nutzers auf der Plattform.	Nicht angegeben.	Schweregrad und Häufigkeit der Verstöße.
Sanktionen aufgrund von Verstößen	Reicht von Verwarnungen, Nutzungseinschränkungen bis zur vollständigen Deaktivierung des Profils.	Entfernung des Inhalts und mögliche temporäre Sperrung des Accounts. Dauerhafte Sperrung des Accounts erst nach wiederholten Verstößen.	Entfernung des Inhalts und mögliche Deaktivierung des Kontos.	Erster Verstoß: Warnung ohne Sanktionen. Darauf folgende Verstöße: Löschung des Inhalts (Videos oder Kommentare), Verwarnungen und eingeschränkter Nutzung der

				Plattform für einen bestimmten Zeitraum. Nach Erhaltung von drei Verwarnungen innerhalb von 90 Tagen werden der Kanal und seine Videos endgültig entfernt.
Öffentliches Interesse	Inhalte, die gegen die Richtlinien verstoßen, dürfen zum Zwecke der Bewusstseinsbildung, Bekämpfung oder Aufklärung über das Thema beibehalten werden.	Inhalte, die gegen die Richtlinien verstoßen, dürfen zum Zwecke der Bewusstseinsbildung, Bekämpfung oder Aufklärung über das Thema beibehalten werden.	Inhalte, die gegen die Richtlinien verstoßen, dürfen zum Zwecke der Bewusstseinsbildung, Bekämpfung oder Aufklärung über das Thema beibehalten werden.	Inhalte, die gegen die Richtlinien verstoßen, dürfen zum Zwecke der Bewusstseinsbildung, Bekämpfung oder Aufklärung über das Thema beibehalten werden.

Quelle: Gemeinschaftsrichtlinien der Plattformen | Gestaltung: FGV DAPP

ANALYSE DER DEBATTE ÜBER HASSREDEN UND ZENSUR AUF TWITTER UND FACEBOOK

In diesem Abschnitt werden die Analysen der Debatte über Hassreden in den Netzwerken vorgestellt. Aufgrund der Tatsache, dass sich ein bedeutender Anteil der Social-Media-Nutzer auf Twitter und Facebook konzentriert, und sie gemeinsame Punkte bezüglich erlaubter Posting-Inhalte und Formen von Engagement aufweisen, wurden Postings dieser beiden Plattformen herangezogen.

Die auf Twitter und Facebook erfassten Daten wurden in zwei Gruppen eingestuft: **Debatte über Hassrede** und **Debatte über Zensur** auf der jeweiligen Plattform. Die Debatte über

Hassrede erreichte ihren Höhepunkt am 20. November 2020 sowohl auf Twitter als auch auf Facebook. Hinsichtlich der Debatte über Hassrede wurde beobachtet, dass die Nutzer der analysierten sozialen Netzwerke dazu neigen, darunter beleidigende Äußerungen, in Online- sowie Offline-Kontexten, gegenüber gesellschaftlichen Minderheiten zu verstehen. Aus diesem Grund werden immer wieder Debatten über Rassismus und Misogynie festgestellt. Basierend auf den erfassten Daten ist interessant hervorzuheben, dass die Debatte in den Netzwerken über dieses Thema immer durch Ereignisse, die im Offline-Kontext geschahen und die Identifizierung des Opfers bzw. des Aggressors ermöglichen, vorangebracht wird.

Der Untersuchung der gesammelten Daten ist zu entnehmen, dass die Debatte über Hassrede kontinuierlich mit der Diskussion über die Zensurpraxis oder den Angriff auf die Meinungsfreiheit einhergeht, wie die Literatur zum Thema zeigt. Unter den analysierten Kommentaren und Veröffentlichungen taucht immer wieder die Idee auf, dass die Sperrung von Konten wegen als beleidigend verstandener Postings die Einschränkung der freien Meinungsäußerung von Individuen bedeute. Daran, dass der Fokus der Debatte auf den Maßnahmen zur Löschung von Postings und zur Sperrung oder Deaktivierung von Konten liegt, wie im Fall vom ehemaligen US-Präsidenten Donald Trump liegt, lässt sich deutlich erkennen, dass die Sanktionen, die die Plattformen gegen die Nutzer der sozialen Netzwerke verhängen, Kern des Diskurses über freie Meinungsäußerung und Zensur ist.

Die Datenextraktion wurde anhand von diesem Überblick über Hassrede und Meinungsfreiheit, der von den Postings in den sozialen Netzwerken hervorgehoben wird, durchgeführt. Basierend auf diesen Daten wurde die Entscheidung getroffen, den Spitzenwert der Erwähnungen und die Tage des untersuchten Zeitraums, in denen die Anzahl von Postings zu den Themen besonders hoch war, zu analysieren. Es handelt sich also um eine qualitativ-quantitative Datenanalyse, da zunächst alle Postings zum Thema gesammelt wurden und anschließend der Fokus auf die Höhepunkte dieser Angaben gelegt wurde. Das Ziel bestand darin, das Engagement nachzuvollziehen bzw. zu veranschaulichen, welche Postings, Nachrichten oder Nutzer zur Intensivierung der Debatte beitrugen.

Der Blick auf das Engagement konzentrierte sich auf die erreichte Reichweite der Postings. Es wurden sämtliche Möglichkeiten betrachtet, die ein Posting bietet, Facebook- und Twitter-Nutzer für Themen zu engagieren, nicht nur die sogenannten „Vanity-Metriken“. Diese definiert Rogers (2019) als Metriken, die nur Likes und die Gesamtzahl der Follower eines bestimmten Profils enthalten. Die vorliegende Studie führt eine kritische Analyse der sozialen Netzwerke durch und setzt sich somit den Ergebnissen derartig einfacher Metriken von Postings oder Profilseiten entgegen (SANCHES 2020). Basierend auf dieser Perspektive wurde die qualitative Analyse ausgehend von Fällen, die durch die quantitative Datenerhebung mit Bezug auf Interaktionen wie Likes, Teilungen und Kommentare bestimmter Accounts identifiziert wurden, durchgeführt. Somit konnten die Reichweite der Postings und die zentrale Rolle des Verfassers oder dessen Posting bei den Diskussionen auf der Plattform verstanden werden. Im Folgenden werden die Analyseergebnisse der Debatten auf Twitter und Facebook diskutiert. Dabei werden zunächst die Spitzenwerte der Postings über Hassrede und anschließend jene über Zensur untersucht.

Debatte über Hassrede und Zensur auf Twitter

Zwischen dem 16. November 2020 und dem 6. Februar 2021 gab es ca. 11,6 Mio. Postings auf Twitter, die Hassrede thematisieren. Die Debatte wurde durch die Feierlichkeiten des Tages des Schwarzen Bewusstseins, die Ermordung eines schwarzen Kunden durch einen Sicherheitsmann im Carrefour-Supermarkt in Porto Alegre (RS) sowie durch den Sturm auf das Kapitol in den USA und die darauffolgende Sperrung der Social-Media-Accounts des ehemaligen US-Präsidenten Donald Trump angefacht. Demzufolge waren Hassrede (Debatte über strukturellen Rassismus) und Meinungsfreiheit (Debatte über Zensur) die Themen, die die Debatten auf Twitter am meisten anregten. Wie die Literatur zum Thema zeigt, geht die Debatte über Online-Delikte immer mit Fragen bezüglich der Einschränkung

individueller Freiheiten und der Zensur durch die Plattformen gegenüber Meinungsäußerungen der Nutzer einher.

Im analysierten Zeitraum trugen weitere Ereignisse deutlich zu dieser Debatte bei. Hinsichtlich der Hassrede wurde [der Fall von Belästigung der Abgeordneten Isa Penna \(PSOL-SP\) von einer Diskussion über die Vulnerabilität von Frauen und Misogynie in unterschiedlichen Umgebungen durchdrungen](#)¹⁶. In Bezug auf die Meinungsfreiheit lässt sich auf die Veröffentlichung des Gesundheitsministeriums über die präventive Behandlung gegen Covid-19 hinweisen, die von Twitter als irreführenden Inhalt eingestuft wurde.

Wie bereits erwähnt, geht die Erfassung von Daten über Online-Hassrede oder Beleidigungen immer mit Postings über das Recht auf freie Meinungsäußerung seitens der Nutzer einher. Unter diesem Aspekt wurde die Debatte über die Bekanntgabe der Sperrung von Konten oder der Löschung von Inhalten auf Social-Media-Plattformen wie Facebook, Instagram und Twitter wegen Verstoßes gegen deren Richtlinien im analysierten Zeitraum intensiv geführt. Einige Postings - von Accounts von parteiideologisch oder identitär orientierten Politikern oder Bloggern - stellen die Entscheidung der Plattformen infrage. Dabei wird argumentiert, dass die gesperrten Inhalte keine Gemeinschaftsstandards verletzen, oder, dass die Begründung die Sanktionen nicht rechtfertigen würde, sondern es sich um Diffamierungskampagnen seitens der Oppositionsgruppen handle. Teilweise wird die Sperrung mit anderen Fällen verglichen, in denen Nutzer ebenfalls Postings mit sensiblen Inhalten meldeten, dies jedoch zu keiner Löschung des Beitrags oder Sperrung von Accounts durch die Plattformen führte. Dazu gehören beispielsweise Beleidigungen und Bedrohungen gegenüber bestimmten Politikern oder parteiischen Gruppen sowie die Verbreitung von Falschnachrichten und Informationen.

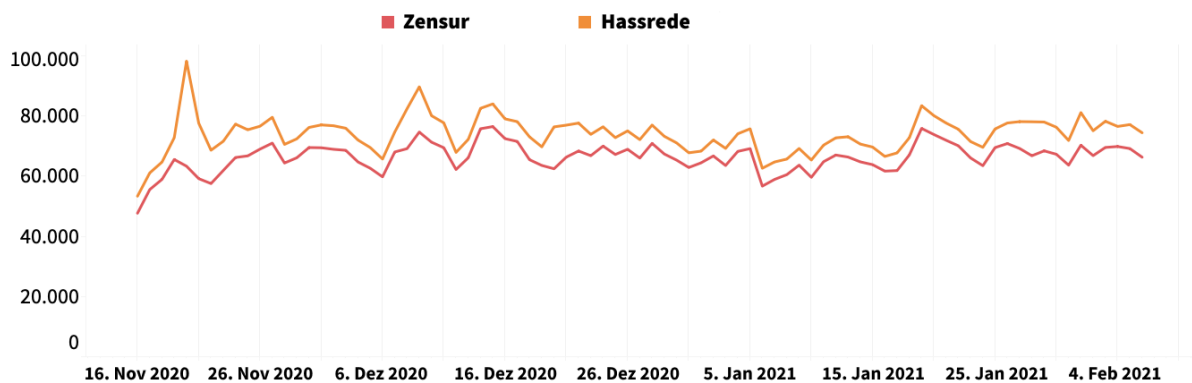
Das folgende Diagramm zeigt die Entwicklung der Debatten auf Twitter. Ausgehend von den dargestellten Spitzenwerten am 20. November (über Hassrede) und am 15. Dezember

¹⁶ Während einer Sitzung der Legislativversammlung des Bundesstaates São Paulo am 16. Dezember 2020 wurde die Abgeordnete Isa Penna (PSOL-SP) vom Abgeordneten Fernando Cury (Cidadania-SP) sexuell belästigt. Die Belästigung betrifft ungewolltes Anfassen intimer Körperteile der Abgeordneten.

2020 sowie am 19. Januar 2021 (über Meinungsfreiheit) wird eine qualitative Analyse basierend auf dem Engagement für die Themen durchgeführt. Hierzu werden die Teilungen sowie die Anzahl von Likes und von Followern jedes Accounts berücksichtigt, um die Postings, die die Debatte anfachten, zu identifizieren.

Diagramm 1 - Umfang der Interaktionen bezüglich Hassrede und Zensur auf Twitter

Zeitraum: vom 1. November 2020 bis zum 6. Februar 2021



Quelle: Twitter | Gestaltung: FGV DAPP

Hinsichtlich der Debatte über Hassrede wurde eine Zentralität bzgl. des in der brasilianischen Gesellschaft präsenten strukturellen Rassismus beobachtet. Diese Diskussion wurde durch zwei Gründe, die in Verbindung miteinander gebracht wurden, ausgelöst: Den Tag des Schwarzen Bewusstseins am 20. November und [die Ermordung João Alberto Freitas durch Sicherheitskräfte des Carrefour-Supermarkts in Bundesland Rio Grande do Sul](#). Die Frage, ob der Tod Freitas ein rassistisch motiviertes Verbrechen oder eine Folge der Aggressivität der Sicherheitskraft darstellte, bestimmte die Diskussion. Die Reichweite dieser Debatte ist auf die Berichterstattung über den Fall seitens diverser Nachrichtenkanäle zurückzuführen und verstärkte sich durch die Demonstration, die als Reaktion auf die Ermordung am 20. November 2020 an der Avenida Paulista in São Paulo stattfand. Postings von Kanälen wie *Mídia Ninja* und *Quebrando o Tabu*, die über den Protest berichteten, fachten die Debatte über Rassismus in den sozialen Netzwerken an.

Was die Diskussionen über Zensur und Meinungsfreiheit bezogen auf die Plattformen angeht, ließ sich eine bedeutsame Anzahl von Tweets am 15. Dezember 2020 beobachten, welche 76.467 Postings betrug. Bei der Identifizierung der Debatte wurden zwei Postings, die viel Aufmerksamkeit erzeugten, festgestellt: eines von Jair Bolsonaro und eines von der Abgeordneten Carla Zambelli (PSL-SP). Das Posting der Abgeordneten bezog sich explizit auf den [vom Abgeordneten Alessandro Vieira \(Cidadania\) verfassten Gesetzentwurf Nr. 2630/2020](#), der im zweiten Abschnitt bereits erwähnt wurde. Laut ihrem Beitrag fand im Bundessenat eine Konsultation statt, die zwar noch nicht abgeschlossen war, aber entgegen des Wunsches der Abgeordneten und einiger ihrer Unterstützer dazu tendierte, den Gesetzesentwurf zu verabschieden. Unter anderen Maßnahmen zielt das Gesetz darauf hin, dass die Plattformen Sanktionen gegen falsche und beleidigende Postings regulieren und durchsetzen. Diese Absicht wurde als Zensur und Nichtachtung der individuellen Meinungsfreiheit dargestellt. Ausgehend von diesen Postings entstanden weitere Beiträge, die auf Momente hindeuteten, in denen Twitter keine Likes oder Kommentare von Inhalten zuließ, die die Plattform als unvereinbar mit ihren Nutzungsbedingungen oder Gemeinschaftsrichtlinien verstand.

Ein deutliches Beispiel für diese Debatte war die große Mobilisierung rund um die Entscheidung von Twitter, das offizielle Konto Donald Trumps zu sperren, in den ersten Wochen vom Januar 2021. Dies geschah nachdem der ehemalige Präsident ein Video postete, auf dem er, laut den Kommentaren, seine Anhänger zum Sturm auf den US-Kongress anfeuerte, um gegen die Ergebnisse der von Joe Biden gewonnenen US-Wahl 2020 zu protestieren. Die Entscheidung der Plattform basierte auf der Behauptung, dass das Video zu Gewalttaten unter den Demonstranten aufhetzen könnte. Anschließend kündigten andere soziale Netzwerke - wie Facebook, Instagram und Snapchat - die Sperrung seines Kontos an.

Die Sperrung von den Konten des ehemaligen amerikanischen Präsidenten auf verschiedenen Plattformen löste eine Diskussion über das konfliktreiche Dilemma von Hassrede und Meinungsfreiheit in den sozialen Netzwerken aus. Einige Profilseiten, die den Protesten am US-Kapitol kritisch gegenüberstanden, befürworteten die Entscheidung

von Twitter, verurteilten die von Trump verbreitete “Hassbotschaft” und verteidigten die Auffassung, dass der Meinungsfreiheit – die als Motto von Unterstützern des ehemaligen Präsidenten angeführt wird – Grenzen gesetzt werden müssen. Es soll nicht ungeachtet bleiben, dass durch die Berichterstattung diverser Presseorganisationen zu sehen war, dass viele der Demonstranten, die den US-Kongress stürmten, Kostüme oder Symbole trugen, die bekannterweise mit der weißen Vorherrschaft assoziiert werden. Dieses Ereignis bekräftigte das Argument derjenigen, die Einschränkungen der freien Meinungsäußerungen befürworten.

Allerdings ist die vor allem durch konservative Politiker, Blogger und Influencer vertretene Meinung, die in dieser Debatte herrschte, dass die Entscheidung von Twitter willkürlich war. Mit dem Argument, dass Trump verfolgt würde, wurde das Handeln der Plattform als eine Strategie linker Gruppen zur Machtergreifung in den Vereinigten Staaten dargestellt. Die Postings deuteten auf die Inkohärenz bei der Entscheidung der Plattform hin, die Konten „diktatorischer“ Regierungsbehörden im sozialen Netzwerk zuzulassen - wie die Profileiten von Nicolás Maduro, dem Präsidenten Venezuelas, und von Hassan Rouhani, dem Präsidenten des Irans. Dabei wurden die Kommentare, die die Botschaft Trumps kritisierten, als heuchlerisch bezeichnet, indem ihnen vorgeworfen wurde, einen Akt von Zensur gegen den ehemaligen Präsidenten zu preisen, obwohl sie sich als Verteidiger der Demokratie präsentieren. Einige Nutzer, [beeinflusst vom Posting des Influencers Allan dos Santos](#), werteten die Entscheidung als absurd, diktatorisch und als Versuch, bestimmte Akteure zum Schweigen zu bringen, was die Debatte anfachte.

Ein weiterer Punkt, der die Debatte über Zensur und Meinungsfreiheit anregte, war der vom brasilianischen Gesundheitsministerium gepostete und als irreführend gekennzeichnete Inhalt über die präventive Behandlung gegen COVID-19. Das Engagement erfolgte basierend auf einer [vom Kanal 247 veröffentlichten Nachricht](#) über die Ermittlung des Falls, die von der Behörde der Generalstaatsanwaltschaft des Bundeslandes Goiás eingeleitet wurde. Obwohl die Debatte unter diesem Aspekt von Postings über COVID-19 und dessen Gefährlichkeit durchdrungen ist, wurde sie auch durch den Eingriff seitens der

Plattform in ein offizielles, behördliches Konto geprägt, was auf die Risiken dieser Maßnahme für die Meinungsfreiheit hindeutete.

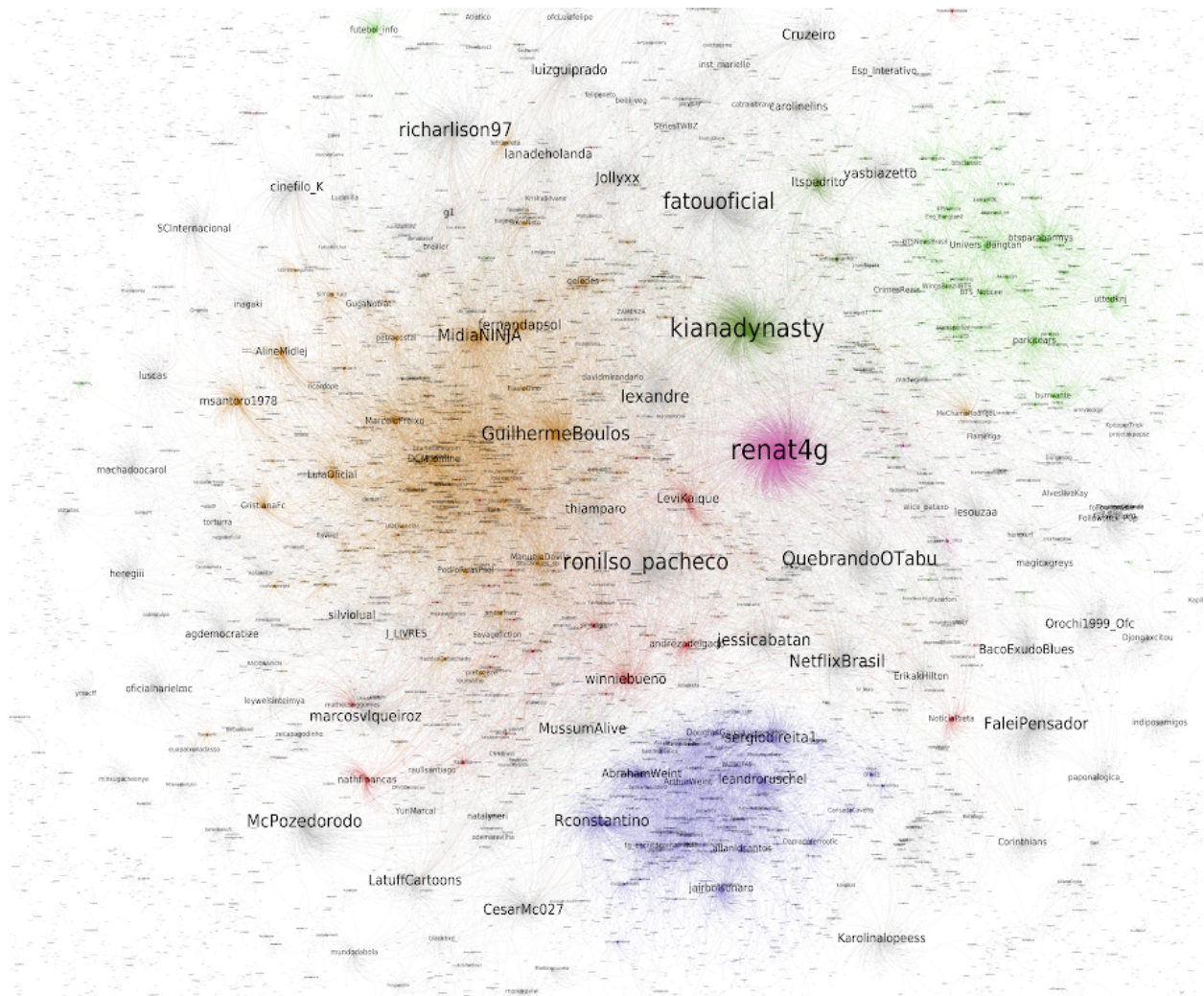
Der folgende Abschnitt erläutert, wie diese Profilseiten miteinander interagierten und wer und was eine zentrale Rolle bei der Debatte spielte. Im Allgemeinen lässt sich feststellen, dass die Debatte über Zensur im ganzen Analysezeitraum geführt und von auf nationaler Ebene relevanten Personen des politischen Lebens bekräftigt wurde.

Analyse der Akteure auf Twitter

Am 20. November 2020 beherrschten die Gespräche über die Ermordung João Alberto Silveira de Freitas am vorherigen Tag in einem Supermarkt in Porto Alegre die Debatte über Hassrede auf Twitter. Die Verbreitung von persönlichen Meinungen und Interpretationen des Falls ermöglichte die Identifizierung spezifischer Gruppen in diesem Kontext.

Visualisierung 1 - Interaktionen in der Debatte über Hassrede auf Twitter

Zeitpunkt: 20. November 2020



Quelle: Twitter | Gestaltung: FGV DAPP

Orange – 11,21% der Interaktionen

Die aus linken Politikern und Medienkanälen bestehende Gruppe bedauert den Mord an João Alberto Silveira de Freitas und weist auf den rassistischen Charakter des Verbrechens hin. Einige Beiträge rufen zur Mobilisierung sozialer Bewegungen auf, um gegen das Geschehene zu demonstrieren.

Blau – 6,63% der Interaktionen

Beeinflusst von rechten Politikern, Bloggern und bekannten Persönlichkeiten konservativer Orientierung hält die Gruppe die Bezeichnung der Ermordung Freitas als rassistisches Verbrechen für unüberlegt. Die Postings berichten über die angebliche Erklärung einer Polizeikommissarin, die an der Ermittlung beteiligt ist und Indizien für eine rassistische Motivation ausschließt. Einige Profilseiten rufen ein Video von 2021 ab, in dem der amerikanische Schauspieler Morgan Freeman während eines Interviews den *Black History Month* infrage stellt und fragt, wann der *White History Month* wäre. Neben dem Fall von Freitas thematisieren die Postings in dieser Gruppe zwei weitere Ereignisse: die Ankündigung von gerichtlichen Verfahren, die von der ehemaligen Athletin Ana Paula Henkel wegen Beleidigungen und Angriffe auf sie in sozialen Netzwerken eingeleitet wurden, und die Ankündigung vom Gegendarstellungsanspruch, den der Influencer Olavo de Carvalho nach Äußerungen der Journalistin Thaís Oyama gegen ihn erhielt.

Hellgrün – 4,98% der Interaktionen

Bezugnehmend auf die Ermordung Freitas veröffentlicht die Gruppe der Entertainment-Kanäle weitere Fälle von Rassismus, die schwarze Berufstätige und Influencer in Brasilien erleben. In diesem Kontext sticht fällt der Tod des Angestellten eines Supermarkts in Recife (PE) im August 2020 besonders heraus. Seine Leiche wurde unter Sonnenschirmen versteckt, während das Geschäft weiterhin geöffnet blieb. Weitere in diesem Zuge diskutierte Fälle sind der der schwarzen Youtuberin @badgallore, die in einer Bankfiliale beschämt worden sei, nachdem sie über den hohen Geldbetrag auf ihrem Konto befragt wurde sowie der Fall des Managers eines Kaufhauses in Minas Gerais, dessen Position im Unternehmen von einer Kundin angezweifelt wurde, weil er schwarz ist.

Rot – 3,17% der Interaktionen

Ausgehend von Reaktionen auf den Mord an Freitas verurteilten Profilseiten schwarzer Aktivisten die mögliche Relativierung des Rassismus und die Schuldzuweisung an die Gewaltopfer. Die Postings stellen die angebliche Aussage eines für die Ermittlung

verantwortlichen Polizeikommissars, dass es keine Beweise für Rassismus im Fall gäbe, sowie die brasilianische Justiz infrage und verglichen die Situation mit der angeblichen Aussage eines Richters über „fahrlässige Vergewaltigung“ beim im September 2020 gefällten Urteil des Sexualverbrechens gegen die Bloggerin Mariana Ferrer.

Lila – 2,69% der Interaktionen

Die von Profilseiten schwarzer Aktivisten und Influencer mobilisierte Gruppe äußert Zorn und Empörung über den körperlichen Angriff mit Todesfolge auf Freitas und besonders über die Tatsache, dass dieser am Vortag des brasilianischen Tages des Schwarzen Bewusstseins geschah. Mit dem Argument, dass es am betroffenen Datum nichts zu feiern gibt, veröffentlichen viele Profilseiten Daten über die Realität der schwarzen Bevölkerung im Land, um hervorzuheben, dass das Verbrechen keinen Einzelfall darstellt, sondern vielmehr einem Abbild der Situation der Schwarzen in Brasilien entspricht.

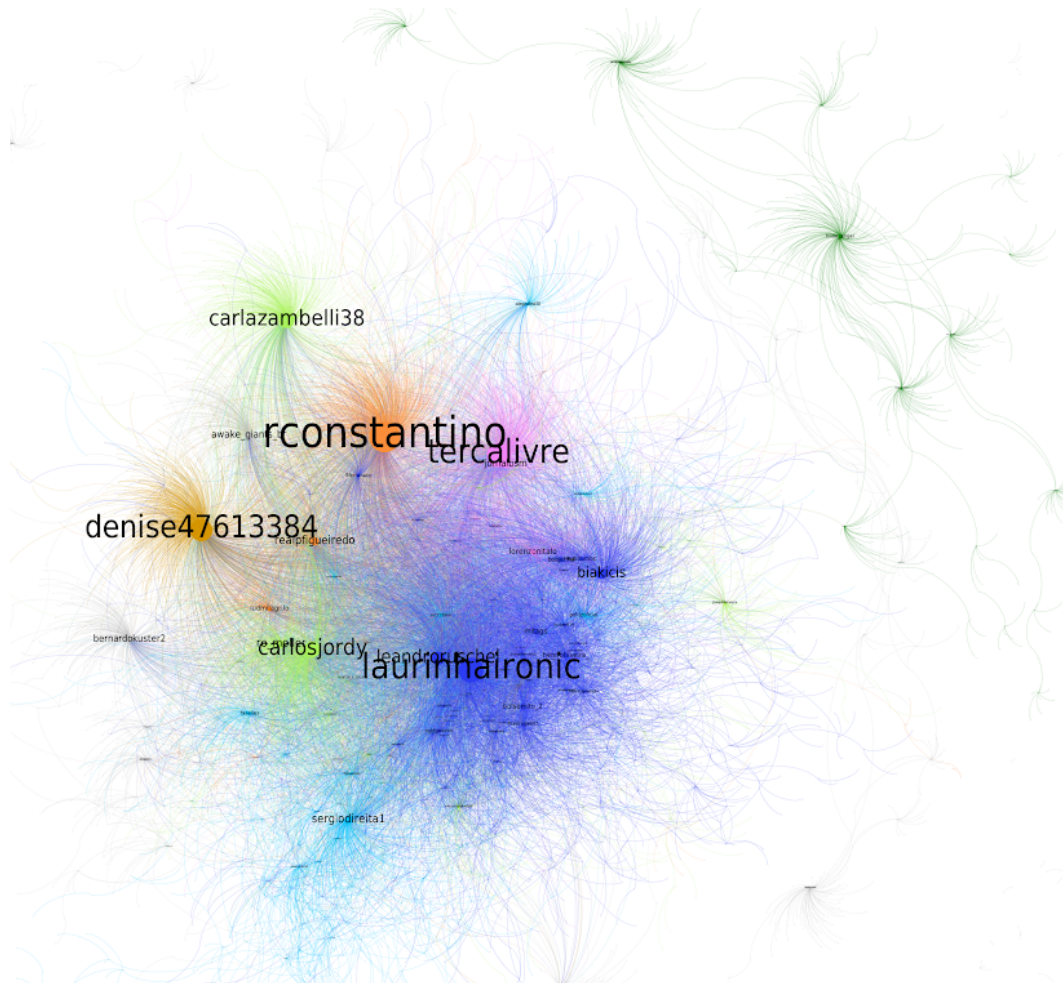
Dunkelgrün – 2,50% der Interaktionen

Die aus digitalen Influencern bestehende Gruppe greift die Supermarktkette an, indem sie betont, dass durch den Mord an Freitas nicht zum ersten Mal eine Filiale des Unternehmens zum Schauplatz grässlicher Taten wurde, und besteht auf der Exponierung der Sicherheitskräfte, die das Verbrechen begangen, statt der Exponierung des Opfers.

Hinsichtlich der Meinungsfreiheit - und möglicherweise der Zensur - legte die Debatte auf Twitter den Fokus auf unterschiedliche Ereignisse mit dem Hauptziel, die Entscheidungen, die verschiedene soziale Netzwerke für die Moderation der auf ihren Plattformen verbreiteten Inhalte trafen, infrage zu stellen.

Visualisierung 2- Interaktionen in der Debatte über Zensur auf Twitter

Zeitraum: vom 14. bis zum 15. Januar 2021



Fonte: Twitter | Elaboração: FGV DAPP

Dunkelblau – 39.23% der Interaktionen

Die Gruppe, die von rechten Politikern und Profilseiten rechtsorientierter Humoristen mobilisiert wird, ruft ihre Anhänger dazu auf, sich der Kampagne **#silenceday** anzuschließen, um gegen eine vermeintliche Einschränkung der Meinungsfreiheit durch soziale Netzwerke wie Facebook, Twitter, Instagram und YouTube zu protestieren. Als Reaktion auf die Moderation einer Reihe von Inhalten, die hauptsächlich von konservativen Profilen veröffentlicht wurden, weist die Kampagne die Nutzer an, einen Tag lang nichts auf den Plattformen zu posten, zu liken oder zu teilen.

Hellgrün – 11,17% der Interaktionen

Angesichts der Gesundheitskrise, die die Stadt Manaus (AM) im Januar aufgrund der COVID-19-Pandemie schwer traf, wendet sich die aus rechten Politikern, Bloggern und digitalen Influencern konservativer Auffassung zusammengesetzte Gruppe gegen einen vermeintlichen Versuch von traditionellen Medien und linken Politikern, die Bundesregierung für die Episode verantwortlich zu machen. Die Profileseiten werfen der brasilianischen Presse die Verbreitung von Fake News vor und argumentieren, dass der Gouverneur von Amazonas die Verantwortung für den Zusammenbruch des Gesundheitssystems der Hauptstadt des Bundesstaates trägt.

Rosa – 9,03% der Interaktionen

In Zusammenhang mit rechtsorientierten, alternativen Medienkanälen kritisiert die Gruppe Entscheidungen einiger sozialen Netzwerke wie Twitter und Facebook wegen der Moderation offizieller Accounts des damaligen amerikanischen Präsidenten Donald Trump sowie der entsprechenden Inhalte über die US-Präsidentschaftswahlen 2020. Mit der Begründung, dass die Entscheidungen der Plattformen Zensur darstellen würden, mobilisierten die Profileseiten ebenfalls für die **#silenceday**-Kampagne.

Orange – 8,89% der Interaktionen

Mit dem Aufruf zur **#silenceday**-Kampagne greift die aus Profileseiten konservativer Journalisten bestehende Gruppe die Gerichtsverfahren an, die versuchen, die Richterin Ludmila Lins Grilo daran zu hindern, auf ihren sozialen Netzwerken Inhalte zu veröffentlichen, die die Lockerung der sozialen Isolation und der Maskenpflicht im Rahmen der COVID-19-Pandemie befürworten. Dabei wird argumentiert, dass die Gerichtsverfahren das Recht auf freie Meinungsäußerung der Magistrate verletzt.

Dunkelgrün – 3,11% der Interaktionen

Mit der Unterstützung von Journalisten, politischen Kommentatoren und digitalen Influencern ironisierte die Gruppe die **#silenceday**-Kampagne und bestreitet die Behauptung, dass die Moderation von Postings und Kanälen des ehemaligen US-Präsidenten Donald Trump in den sozialen Medien gegen das Prinzip der Meinungsfreiheit verstößt. In den Postings werden die Aussagen von mit der Bundesregierung verbundenen Politikern, die die COVID-19-Pandemie verharmlosen und sich den Empfehlungen der Gesundheitsbehörden entgegensetzen, kritisiert, und die Social-Media-Plattformen dazu aufgefordert, Maßnahmen wie Deaktivierung von Konten und Löschung der Inhalte zu ergreifen.

Hellblau – 12,34% der Interaktionen

Die Gruppe, die durch Profilseiten rechter Politiker sowie konservativer Informationskanäle und digitaler Influencer mobilisiert wird, weigert sich, der **#silenceday**-Kampagne beizutreten, mit der Begründung, dass die Initiative das Recht auf freie Meinungsäußerung nicht garantiert. Stattdessen halten die Profilseiten es für wichtig, dass Menschen die sozialen Netzwerke nutzen, um den Diskurs der traditionellen Medien, welcher der Bundesregierung die Gesundheitskrise in Manaus vorwirft, zu hinterfragen. Im Zuge der Mobilisierung bestehen die Postings auf der Verantwortung des Gouverneurs von Amazonas für den Zusammenbruch des Gesundheitssystems in der Region.

Gelb – 4,16% der Interaktionen

Profilseiten durchschnittlicher Nutzer geben die Nachricht wieder, dass die russische Regierung die von Social-Media-Plattformen durchgeführte Sperrung von Postings und Konten des ehemaligen US-Präsidenten Donald Trump kritisierte und dass Wladimir Putin bereit wäre, Versuche seitens der Plattformen, die Meinungsfreiheit in seinem Land einzuschränken, zu bekämpfen. Die Gruppe teilt auch die Information, dass Websites wie

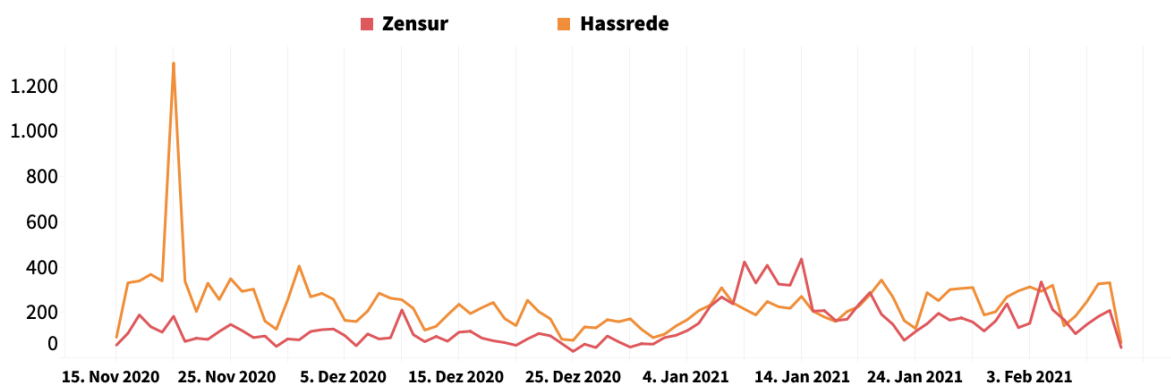
Twitter und Facebook wegen der angeblichen Zensur gegen Trump finanzielle Verluste erleiden.

Debatte über Hassrede und Zensur auf Facebook

Dieser Abschnitt stellt die Ergebnisse der Analyse vor, die zwischen dem 15. November 2020 und dem 11. Februar 2021 auf Facebook durchgeführt wurde. Insgesamt wurden 21.149 Postings über Hassrede und 12.773 über Meinungsfreiheit oder Zensur gesammelt. Die Entwicklung beider Debatten wurde separat beobachtet, unter Berücksichtigung ihrer direkten Beziehung zueinander, wie bereits diskutiert. Die Auseinandersetzung mit beiden Fragen spielt eine wichtige Rolle bei dieser Studie, da die Forderung nach Mechanismen zur Eindämmung der Hassreden und parallel zur Setzung von den Grenzen der Meinungsfreiheit zunimmt.

Diagramm 2 - Entwicklung der Debatte über Hassrede und Zensur auf Facebook

Zeitraum: vom 15. November 2020 bis zum 11. Februar 2021



Quelle: Facebook | Gestaltung: FGV DAPP

Beginnend mit den auf Hassrede bezogenen Postings zeigte der Korpus einige spezifische Debatten, die im analysierten Zeitraum hervorstachen. Einige Profilseiten hielten das Thema für sehr wichtig und wiesen auf den Bedarf hin, diskriminierende Diskurse im Netz

zu bekämpfen, wie [das Engagement dafür seitens einer gewählten Stadträtin in Jarú \(RO\)](#) oder die Einladung einer Kirche in Osasco (SP), die vorschlug, die [Hasskultur](#) in einer ihrer Feiern zu thematisieren. Die Hassreden wurden auch von Seiten erwähnt, die das Thema als eine Art von Bevorzugung der [von Beleidigungen geprägten](#) Sozialbeziehungen zum Nachteil der anderen betrachtet. Die Daten deckten das Engagement für Diskussionen auf, die spezifisch mit Rassismus, Misogynie und Homophobie zusammenhängen.

In Anbetracht dessen, dass der Zeitraum der Analyse die zwei Runden der brasilianischen Kommunalwahlen umfasste, beeinflussten diese die Postings, vor allem in den Monaten November und Dezember. Ein Beispiel, welches viel diskutiert wurde, handelt von der [ersten schwarzen Frau, die als Stadträtin in Joinville \(SC\) gewählt wurde](#) und daraufhin in sozialen Netzwerken rassistisch beleidigt und mit dem Tod bedroht wurde. Zudem gibt es auch Postings, die ein Video von der Komikergruppe *Porta dos Fundos* kritisieren, auf dem der Wahlerfolg der Stadträtin von Curitiba (PR), die mit den meisten Stimmen gewählt wurde, [durch sexuelle Konnotation](#) erklärt wird.

Wie auch auf Twitter zu beobachten war, erreichte die Debatte über das Thema auf Facebook ihren Höhepunkt am 20. November, Tag des Schwarzen Bewusstseins, mit 1.301 Postings über Hassrede auf der Plattform, was dem fünffachen Wert des Durchschnitts des analysierten Zeitraums (237,6 Postings/Tag) entspricht. Das Datum allein fördert normalerweise die Produktion von Inhalten, vor allem seitens öffentlicher Persönlichkeiten, über den Bedarf, Rassismus zu bekämpfen. Doch der Totschlag vom schwarzen Kunden des Carrefour-Supermarkts von Porto Alegre (RS), João Alberto Silveira de Freitas, einen Tag vor dem Tag des Schwarzen Bewusstseins auf dem Parkplatz des Geschäfts, steigerte das Niveau des Engagements.

Im Lauf des Tages hoben die Postings einerseits die Betrachtung des Falls als ein Beispiel vom strukturellen Rassismus hervor, andererseits die Bekämpfung der Idee, dass das Verbrechen rassistisch motiviert war. Obwohl die meisten Postings das Ereignis, unter Berücksichtigung des Feiertages, mit dem schwachen Verständnis der Gleichberechtigung im Land assoziierten, erzeugte das Posting auf der [Profilseite Jair Bolsonaro](#) die meisten

Interaktionen. Im seinem 41.787 mal geteilten Beitrag behauptete der Präsident, dass der Hass von denjenigen ausgeht, die das brasilianische Volk durch Ressentiments spalten wollen, wodurch die Menschen anfällig für Kontrollmechanismen werden.

Die Behauptung wurde von denjenigen begrüßt, die der Meinung sind, dass die Diskussion über Rassismus in Brasilien eine Strategie ist, Dissens zwischen Brasilianern zu erzeugen und ein bestimmtes linksorientiertes Chaosprojekt zu fördern. Die Datenerfassung identifizierte jedoch auch Postings von Profalseiten, die nicht konservativ sind, laut denen die Betrachtung der Tötung von Freitas als Rassismus [den Klassenkampf](#) unsichtbar macht und die Möglichkeiten zur Bekämpfung der Ungleichheiten reduziert. Einerseits wird die Debatte über Rassismus als schädlich gesehen, da sie die sozialen Beziehungen beeinträchtigt und somit eine angebliche nationale Einheit schwäche. Andererseits wird sie als simplifizierend dargestellt, weil sie die Widerstandsmöglichkeiten auf einen punktuellen Aktivismus reduziere, welcher die Strukturen nicht ändern könne.

Sieben der zehn an jenem Tag meistgeteilten Beiträge beschäftigten sich spezifisch mit der direkten Verbindung zwischen der Tötung von Freitas und dem Rassismus. Unter den Argumenten dafür tauchten statistische Daten auf, die die höhere Anzahl von ermordeten schwarzen Brasilianern gegenüber weißen Brasilianern aufzeigen. In fast allen Postings wurde darauf hingewiesen, wie Carrefour gesetzlich bestraft werden könnte bzw. müsste, aber auch wie die Menschen ihre Empörung mittels des Boykotts gegen die Supermarktkette zeigen könnten. Die öffentlichen Äußerungen im Internet gegen den Supermarkt wurden zu großem Teil als beschönigend dargestellt und es wurde über andere Fälle sowohl von Rassismus als auch von Tod, an denen das Unternehmen beteiligt war, berichtet. Im Allgemeinen wurden die Menschen dazu aufgerufen, sich der Bekämpfung des Rassismus individuell und kollektiv anzuschließen.

Insgesamt gab es sechs Postings, die wiederholt veröffentlicht wurden. Die Analyse ergab jedoch verschiedene Werte bezüglich der unterschiedlichen Kriterien (Anzahl von Likes, Teilungen und Kommentaren) und erreichten somit unterschiedliche Niveaus von Engagement. Vier davon wurden von Personen des öffentlichen Lebens veröffentlicht:

Außer Jair Bolsonaro bestärkten die Debatte am Tag, an dem die Postings ihren Höhepunkt erreichten, der Musiker Nando Reis, die Sängerin Maria Rita und der Priester Reginaldo Manzotti. Bei getrennter Betrachtung ermöglichen die unterschiedlichen Kriterien des Engagements die Identifizierung von Profalseiten anderer Personen des öffentlichen Lebens. Beispielsweise wurde das Posting aus der Stichprobe, das die zweitmeisten Likes erhielt, vom brasilianischen Fußballspieler Fred (Manchester United) veröffentlicht. Wenn das Kriterium der Anzahl von Kommentaren berücksichtigt wird, taucht der Beitrag der Food-Bloggerin Rita Lobo unter den zehn Postings mit dem größtem Engagement auf. Mit Ausnahme des Präsidenten betrachteten alle den Tag des Schwarzen Bewusstseins als einen Tag, an dem die Leiden, die durch den Rassismus verursacht werden, reflektiert werden sollen.

Die Debatte über Hassrede im analysierten Zeitraum wurde also am Tag des Schwarzen Bewusstseins, dem 20. November, angefacht und anschließend mit den Postings über den Mord an Freitas in Bundesland Rio Grande do Sul in Verbindung gebracht. Im Allgemeinen verstanden die Akteure den Totschlag als eine rassistisch motivierte Straftat, nicht nur weil das Opfer schwarz war, sondern auch aufgrund der Brutalität der Aggression und der Tatsache, dass es sich nicht verteidigen konnte. Wie die umfassende Literatur zum Thema verdeutlicht, basiert die Definition von Hassrede auf der Idee von Vorherrschaft, der zufolge bestimmte Gruppen nicht verdienen, gleichberechtigt behandelt zu werden. Es ist die soziale Vulnerabilität dieser Gruppen, die auf bestimmten Eigenschaften basiert, die sie zum Ziel von Angriffen und Hassreden macht.

In Bezug auf die Debatte über Zensur auf Facebook fallen neben Beiträgen, die das Thema in spezifischen Kontexten wie religiösen und künstlerischen ansprechen, zwei Arten von Postings im analysierten Zeitintervall auf: Diejenigen, die den Plattformen, vor allem Twitter und Facebook, die Verletzung der Meinungsfreiheit im Netz vorwerfen, und diejenigen, die gerichtliche Entscheidungen, oft mit Aussagen wie „Sieg der Meinungsfreiheit“ bezeichnet, preisen.

In der ersten Gruppe befinden sich Postings, in denen behauptet wird, es gäbe eine koordinierte Aktion seitens fortschrittlicher, linksorientierter Gruppen in den Kommunikationsmedien, Plattformen und öffentlichen Institutionen, vor allem in den gerichtlichen Institutionen. Das angebliche Ziel dieser Aktion sei es, die konservativen, rechtsorientierten Bewegung zu demontieren. Eines der meistkommentierten Themen in dieser Diskussion war, dass es nicht möglich war, den Link zu einem Artikel der *New York Post* zu veröffentlichen, obwohl dieser nicht geprüft wurde. In dem Text wurden dem Sohn Joe Bidens, derzeit Kandidat der US-Demokraten für die Präsidentschaft, Vorwürfe gegenüber geäußert. Die analysierten Postings warfen Facebook und Twitter vor, eine „Pro-Demokraten-Moderation“ des Textinhalts gemacht und somit Zensur ausgeübt zu haben.

Die zweite auffällige Gruppe beinhaltet Publikationen, die sich unter anderem auf Gerichtsverfahren bezüglich Anklagen gegen Hassreden und Desinformationsinhalte bezogen. Diese Gruppe enthält eine Reihe von Postings, die die Entscheidung der Sportjustiz zugunsten der Angeklagten [Carol Solber, Volleyballspielerin, die „#ForaBolsonaro“ \(#BolsonaroRaus\) während eines Interviews schrie](#), feierten. Die Beiträge, die die Tatsache lobten, dass das Oberste Bundesgericht die Postings des Bundesabgeordneten [Eduardo Bolsonaro \(PSL-SP\) gegen die Bewegung der obdachlosen Arbeiter \(Movimento dos Trabalhadores Sem Teto\)](#) nicht als illegal urteilte, erzeugten ebenfalls Engagement.

Die Postings über Zensur erreichten einen Durchschnitt von 145,5/Tag, was einem kleineren Wert entspricht als der der Postings, die sich auf Hassrede beziehen. An den Tagen 9., 11. und 14. Januar jedoch überschritt die Anzahl von Postings über Zensur diejenige über Hassrede. Dies ist auf die Nachrichten zurückzuführen, dass die Accounts des damaligen US-Präsidenten Donald Trump von Twitter dauerhaft gelöscht und von Facebook und Instagram für unbegrenzte Zeit deaktiviert wurden. An diesen drei Tagen erreichten die auf die Debatte über Zensur bezogenen Postings ihren Spitzenwert, mit insgesamt 423 Postings am 9.1.2021, 408 am 11.1. und 436 am 14.1.

Drei Ereignisse prägten die Diskussionen über Zensur. Das erste betrifft die Sperrung von Trumps Konten auf den betrachteten Plattformen und das Verbot der Plattform Parler, auf der die Profilseite des ehemaligen Präsidenten beibehalten wurde, in den App-Stores von Google, Apple und Amazon. Die Maßnahmen wurden als Zensur gegen konservative Bewegungen und als Angriff auf einen legitimen Vertreter der extremen Rechten interpretiert. Der zweite für die Diskussion relevante Inhalt mit hohem Engagement in den Postings war die Werbung für alternative Plattformen: Parler (vor deren Verbot), Telegram und Paatria. Die Postings vom 14.1.2021, immer noch von den Nachrichten vom 9.1. und 11.1. beeinflusst, wurden durch die **#silenceDay**-Kampagne, die als dritte Komponente die Diskussion prägte, angestachelt. Die Kampagne zielte darauf, den ganzen 15. Januar ohne Postings, Likes, Tags oder weiteren Interaktionen auf Twitter, Instagram, YouTube und WhatsApp zu verbringen, um gegen das Verhalten dieser Plattformen zu protestieren.

In dieser Reihe von Postings war der Beitrag von der [Bundesabgeordneten Carla Zambeli \(PSL-SP\)](#) derjenige, der das größte Engagement erzeugte. Darin befürwortet sie die Ankündigung, dass Trump in Zusammenarbeit „mit anderen Unternehmen zur Bekämpfung der Zensur gegen die Rechte“ eine neue Plattform schaffen würde. Ihre Veröffentlichung wurde 29.524 mal geteilt und erhielt 13.825 Kommentare. Außer Zambelli wurden Profilseiten anderer Personen des politischen Lebens in der Stichprobe als Verfasser von den Inhalten, die die Debatte am meisten anfachten, identifiziert. Darunter sind der Präsident Jair Bolsonaro und die Bundesabgeordneten Bia Kicis (PSL-DF), Caroline de Toni (PSL-SC), Filipe Barros (PSL-PR) und Carlos Jordy (PSL-RJ) besonders hervorzuheben. Die Profilseiten *República de Curitiba* und *Jovem Pan News* veröffentlichten ebenfalls Postings, die die Diskussionen belebten. Im Vergleich zur Debatte über Hassrede war die Zensur-Agenda durch eine größere Einheitlichkeit zwischen den Akteuren gekennzeichnet, was zeigt, dass die Debatte im analysierten Zeitraum größtenteils von rechtskonservativen Gruppen des politischen Feldes artikuliert wurde.

Die Analyse der Debatten über Hassrede und Zensur auf Facebook deutet auf die wichtige Rolle einer öffentlichen Agenda und der Nachrichtenkanäle bei der Förderung öffentlicher Debatten hin. Die Kombination eines Gedenktages, der eine Reihe von Aktionen und Feiern

von Regierungen und der Zivilgesellschaft umfasst, mit dem tragischen Tod von Freitas und den internationalen Nachwirkungen der Löschung von Trumps Konten ermutigte die Internetnutzer dazu, sich online mitzuteilen. Beide Debatten gewannen an Bedeutung durch die Beteiligung von Personen des öffentlichen Lebens aus den Bereichen der Politik, Musik, Sport, Religion und des Internets. Während in der Debatte über Hassrede die Persönlichkeiten, die aus dem Internet bekannt sind, hervorzuheben sind, fallen die im politischen Feld tätigen Personen in der Debatte über Zensur auf. Das deutet auf die Relevanz der Beteiligung bekannter Menschen in großen Netzwerken hin, die dazu führt, dass eine Diskussion wächst und unterschiedliche Instanzen der Gesellschaft erreicht. Dabei wird ebenso offensichtlich, dass die Autoritätspersonen, die in öffentlichen Ämtern tätig sind, Verantwortung für die Entwicklung der Debatten tragen.

FAZIT

Obwohl die in Hassreden enthaltenen Beleidigungen, Bedrohungen und Diskriminierungen, unter Berücksichtigung der großen seelischen Schmerzen, die sie verursachen, subjektive Folgen haben, kann eine eingehende Untersuchung nicht auf die politischen Auswirkungen verzichten. Die Unsichtbarkeit vulnerabler Gruppen, die zum Schweigen gebracht werden, schwächen ihre Forderungen und Ansprüche und schließen sie somit von den demokratischen Entscheidungsprozessen aus. Deswegen sind die Identifizierung und Bekämpfung der Hassreden wichtig für die Demokratie.

Die vorliegende Arbeit gab eine Literaturübersicht über das Thema mit dem Zweck, die wichtigsten Bedeutungen von Hassreden im akademischen und juristischen Bereich bedeuten, deren Spezifitäten in der digitalen Umgebung und deren Definitionen nach den Selbstregulierungsprotokollen der digitalen Plattformen aufzulisten. Die bibliographische und dokumentarische Recherche zeigte, dass Hassreden einerseits theoretisch konzeptualisiert werden können und es andererseits es besonders komplex ist, sie in konkreten Situationen zu definieren. Diese Herausforderung verstärkt sich auf

Internetseiten, die Memes, Videos, organisierte Communities und differenzierte Lexika unter anderen Aspekten beinhalten.

Neben der Einführung internationaler Abkommen zur Bekämpfung unterschiedlicher Diskriminierungsformen deutete das Material auf die Notwendigkeit hin, dass die Staaten bei der Gesetzgebung ihre sozialen, historischen und vor allem kulturellen Kontexte mit den Spezifitäten des Online-Universums in Verbindung bringen. In diesem Sinne präsentierte sich die Herausforderung für die digitalen Plattformen nicht als schwächer. Bei der Formulierung ihrer Nutzungsbedingungen und Gemeinschaftsrichtlinien geben sie an - manche ausführlicher, manche weniger ausführlich -, wie sie die diversen Äußerungen von Hassreden verstehen. Dabei ist jedoch eine Reihe von Informationslücken bzgl. der Art und Weise, wie die Plattformen mit den Nutzern umgehen und die Verbreitung dieser Diskurse verhindern, zu erkennen. Eine mögliche Lösung diesbezüglich wäre, dass globale Plattformen die soziokulturellen Besonderheiten des jeweiligen Kontexts ihrer Nutzer erkunden, um praktikable und produktivere Maßnahmen zur Bekämpfung der Hassreden zu entwickeln.

Die empirische Analyse der Studie bestand in der Lokalisierung der Debatte über Hassreden und deren Beziehung zur Meinungsfreiheit auf Twitter und Facebook im Zeitraum zwischen November 2020 und Februar 2021. Auch wenn die beiden Plattformen digitale soziale Netzwerke mit einer hohen Anzahl von Nutzern, einschließlich des brasilianischen Kontexts, sind, wurde an dieser Stelle festgestellt, dass sie hinsichtlich Repräsentativität Defizite aufweisen, sowohl bezüglich der Anzahl von digitalen Interaktionen als auch der Diversität der beteiligten Akteure. Demzufolge bietet die vorliegende Analyse einen beschränkten Überblick über das komplexe Ökosystem der Erzeugung, Verbreitung und Beibehaltung der Online-Hassreden.

Im Rahmen der methodischen Belange bezieht sich die Herausforderung der Ausarbeitung von Suchsyntaxen für die Erfassung und Klassifizierung der zu analysierenden Daten, auf die Feinheiten, die die Konzeptualisierung und den semantisch-pragmatischen Umfang des Themas kennzeichnen, auf das sich die Studie konzentriert, also das Thema

"Hassrede". Da eine große Schwierigkeit besteht, schlüssig und unbestreitbar zu bestimmen, was tatsächlich Hassrede bedeutet und vor allem, wie dieser Diskurs in der öffentlichen Debatte in sozialen Netzwerken instanziiert wird, ist die Identifizierung der sprachlichen und diskursiven Strategien bzgl. dieses Phänomens eine Aufgabe mit zweifellosen Unsicherheiten.

Basierend auf der Datenerfassung und durchgeführten Analyse ließ sich feststellen, dass die Entwicklung der Debatte über Hassrede und Zensur sich während eines Gedenktages und aufgrund der Nachwirkung von Nachrichten im nationalen und internationalen Kontext verstärkte. Dies lässt vermuten, dass das Vorhandensein einer öffentlichen Agenda und die Vorgehensweise der Nachrichtenmedien die auf den untersuchten Plattformen laufenden Diskussionen anheizten. Neben diesen Elementen erzeugten die Postings zum Thema seitens Personen des öffentlichen Lebens, die Teil eines großen Netzwerks sind, eine hohe Anzahl von Interaktionen, was die politisch-strategische Rolle von diesen Personen bei der Produktion und Reichweite der Debatte signalisiert.

Unter Berücksichtigung der Entwicklung der Online-Debatte im analysierten Zeitraum zeigte die Studie eine Spannung zwischen dem strukturellen und dem persönlichen Verständnis von Diskriminierung auf, was bedeutet, dass diese für einen Teil der Akteure von zentraler Bedeutung für die Fortentwicklung der Demokratie ist und im öffentlichen Geschehen Raum einnehmen sollte; allerdings ist sie für den anderen Teil ein Diskurs, der die politische Ebene destabilisiert, ohne jedoch wirksame Veränderungen hervorzubringen. Schließlich intensivierten die von den Plattformen verhängten Sanktionen, die in den Ablauf der Veröffentlichungen und die Aktivitäten bestimmter Akteure eingriffen, die Debatte über Meinungsfreiheit und erreichten damit eher die Problematisierung ihrer Grenzen, anstatt deren Legitimität.

Durch die Identifizierung der Argumente und Akteure, die diese Debatte verkörpern, schafft die vorliegende Arbeit die Grundlagen für neue Initiativen zur Erforschung der Verbreitung von Hassreden im Netz. Dazu gehören u.a. die Entstehung neuer Plattformen als Reaktion auf die verhängten Sanktionen, das Wachstum das Wachstum der lexikalischen Vielfalt, die

mit den vielfältigen Möglichkeiten der Sprache einhergeht, das Verhältnis der Selbstregulierungsverfahren der Plattformen zu den soziokulturellen Kontexten, in denen sie genutzt werden, und die Strategien sozialer Akteure zur Dekonstruktion der Online-Diskriminierungen.

LITERATURVERZEICHNIS

ALAVA, S./FRAU-MEIGS, D./HASSAN, G.: **Youth and violent extremism on social media**. Mapping the research. Organisation der Vereinten Nationen für Bildung, Wissenschaft und Kultur. Paris 2017.

BARBERÁ, P.: How social media reduces mass political polarization. Evidence from Germany, Spain, and the US. In: AMERICAN POLITICAL SCIENCE ASSOCIATION ANNUAL MEETING (Hrsg.): **Proceedings**. American Political Science Association. San Francisco 2015. Verfügbar unter: http://pablobarbera.com/static/barbera_polarization_APSA.pdf. Letzter Zugriff am: 26.02.2021.

BARBOSA, B./MARTINS, H./VALENTE, J.: **Fake News**. Como as plataformas enfrentam a desinformação. São Paulo 2020.

BENESCH, S.: Dangerous speech. A proposal to prevent group violence. In: **Dangerous Speech Project proposal paper**. February 23, 2013. Verfügbar unter: <https://dangerousspeech.org/wp-content/uploads/2018/01/Dangerous-Speech-Guidelines-2013.pdf>. Letzter Zugriff am: 01.03.2021.

BOWMAN-GRIEVE, L.: Exploring Stormfront. A virtual community of the radical right. In: **Studies in Conflict and Terrorism**, 2009, Bd. 11, N. 31, S. 989-1007.

BROWN, A.: What is so special about online (as compared to offline) hate speech?. In: **Ethnicities**, Bd. 18, Nr. 3, 2018, S. 297-326.

BRUGGER, W.: Proibição ou proteção do discurso do ódio?. Algumas observações sobre o direito alemão e o americano. In: **Revista de Direito Público**, Nr. 15, 2017, S. 117-136.

BUYSE, A.: Words of violence. "Fear speech," or how violent conflict escalation relates to the freedom of expression. In: **Human Rights Quarterly**, Bd. 36, Nr. 4, 2014, S. 779-797.

COHEN-ALMAGOR, R.: Fighting hate and bigotry on the Internet. In: **Policy and Internet**, Bd. 3, Nr. 3, 2011, S. 1-26.

CONSANI, C. F.: Democracia e os discursos de ódio religioso. O debate entre Dworkin e Waldron sobre os limites da tolerância. In: **Etic@**, Bd. 14, Nr. 2, 2015, S. 174-197.

DWORKIN, R.: **O direito da liberdade**. São Paulo 2006.

DWORKIN, R.: Foreword to Extreme Speech and Democracy. In: HARE, I./WEINSTEIN, J. (Hrsg.): **Extreme Speech and Democracy**. New York 2009, S. v-ix.

FARIS, R./ASHAR, A./GASSER, U./JOO, D.: Understanding harmful speech online. In: **Berkman Klein Center Research Publication**, Nr. 2016-21, 2016. Verfügbar unter: <https://papers.ssrn.com/sol3/papers.cfm?abstract%5Fid=2882824>. Letzter Zugriff am: 01.03.2021.

FERNANDEZ-MATAMORO, A./FARKAS, J.: Racism, hate speech, and social media. A systematic review and critique. In: **Television & New Media**, Bd. 22, Nr. 2, 2021, S. 205-224.

FORTUNA, P./SOLER-COMPANY, J./WANNER, L.: Toxic, hateful, offensive or abusive? What are we really classifying? An empirical analysis of hate speech datasets. In: CONFERENCE ON LANGUAGE RESOURCES AND EVALUATION, 12, 2020: **Proceedings**. European Language Resources Association. Marseille 2020. S. 6786-6794.

FUNDAÇÃO GETULIO VARGAS: Vorstand für Evaluation öffentlicher Politiken. **Monitor de temas**. Rio de Janeiro 2015. Verfügbar unter: <http://bibliotecadigital.fgv.br/dspace/handle/10438/15262>. Letzter Zugriff am: 26.02.2021.

GILLESPIE, T.: A relevância dos algoritmos. In **Parágrafo**, Bd. 6, Nr. 1, 2018, S. 95-121.

JURNO, A. C./D'ANDRÉA, C.: (In)visibilidade algorítmica no “feed de notícias” do Facebook. In: **Revista Contemporânea**, Bd. 15, Nr. 2, 2017, S. 463-484.

LUCCAS, V. N./GOMES, F. V./SALVADOR, J. P. F.: **Guia de análise de discurso de ódio**. Rio de Janeiro 2020. Verfügbar unter: <https://www.conib.org.br/wp-content/uploads/2019/11/Guia-de-An%C3%A1lise-de-Discurso-de-%C3%93dio.pdf>. Letzter Zugriff am: 26.02.2021.

LUNA, N./SANTOS, G.: Liberdade de expressão e discurso do ódio no Brasil. In: **Revista Direito e Liberdade**, Bd. 16, Nr. 3, 2014, S. 227-255. Verfügbar unter: http://ww2.esmarn.tjrn.jus.br/revistas/index.php/revista_direito_e_liberdade/article/view/780. Letzter Zugriff am: 01.03.2021.

NANDI, J.: **O combate ao discurso de ódio nas redes sociais**. Bachelorarbeit. Bundesuniversität von Santa Catarina (UFSC), Campus Araranguá, Bachelorabschluss: Informations- und Kommunikationstechnologien. Araranguá, 2018. Verfügbar unter: <https://repositorio.ufsc.br/handle/123456789/187510>. Letzter Zugriff am: 01.03.2021.

NERIS, N. (Hrsg.)/VALENTE, M./CRUZ, F./OLIVA, T.: **Outras vozes**. Gênero, raça, classe e sexualidade nas eleições de 2018. InternetLab. São Paulo 2019. Verfügbar unter: https://www.internetlab.org.br/wp-content/uploads/2019/10/OutrasVozes_2018.pdf. Letzter Zugriff am: 10.03.2021.

ORGANISATION DER VEREINTEN NATIONEN: **Charta der Vereinten Nationen**. San Francisco 1945. Verfügbar unter: <https://www.un.org/en/charte-united-nations/index.html>. Letzter Zugriff am: 26.02.2021.

ORGANISATION DER VEREINTEN NATIONEN: **Allgemeine Erklärung der Menschenrechte**. Paris 1948. Verfügbar unter: <https://www.un.org/en/universal-declaration-human-rights/>. Letzter Zugriff am: 26.02.2021.

PAREKH, B.: Is there a case for banning hate speech? In: HERZ, M./MOLNAR, P. (Hrsg.): **The Content and Context of Hate Speech**. Rethinking Regulation and Responses. Cambridge 2012, S. 37-56.

PEARSON, J./CHILD, J./DEWITT, L./KAHL, D.: Facing the fear. An analysis of speech-anxiety content in public-speaking textbooks. In: **Communication Research Reports**, Bd. 24, 2007, S. 159-168.

ROGERS, R.: Engajados de outra maneira. As Mídias Sociais - Das Métricas de Vaidade à Análise Crítica. In: OMENA, J. J. (Hrsg.): **Métodos Digitais**. Teoria e Prática. Lissabon 2019, S. 73-96.

RUEDIGER, M. (Hrsg.): **Nem tão #simples assim**. O desafio de monitorar políticas públicas nas redes sociais. Rio de Janeiro 2017. Verfügbar unter: <http://hdl.handle.net/10438/19436>. Letzter Zugriff am: 01.03.2021.

SANCHES, D.: Construindo grafos de conhecimento (knowledge graphs) para análise de redes sociais. Uma análise sobre discurso de ódio em Portugal. In: WORKSHOP ON MEDIA, INFORMATION AND DATA SCIENCE, 1, 2020. **Annalen** der Bundesuniversität von Goiás. Goiânia 2020.

ELLARS, A.: Defining Hate Speech. In: **Berkman Klein Center Research Publication**, Nr. 2016-20, Boston Univ. School of Law, Public Law Research Paper, 2016, S. 16-48. Verfügbar unter: <http://dx.doi.org/10.2139/ssrn.2882244>. Letzter Zugriff am: 01.03.2021.

SHEPHERD, A.: Extremism, free speech and the rule of law. Evaluating the compliance of legislation restricting extremist expressions with article 19 ICCPR. **Utrecht Journal of International and European Law**, Bd. 33, 2017, S. 62-83.

SIEGEL, A.: Online hate speech. In: PERSILY, N./TUCKER, J. (Hrsg.): **Social media and democracy**. Cambridge 2020, S. 56-88.

SILVA, L. R./BOTELHO-FRANCISCO, R. E./OLIVEIRA, A. A./PONTES, V. R.: A gestão do discurso de ódio nas plataformas de redes sociais digitais. Um comparativo entre Facebook, Twitter e Youtube. In: **Revista Ibero-americana de Ciência da Informação**, Bd. 12, N. 2, 2019, S. 470-492.

SILVIA, L./MONDAL, M./CORREA, D./BENEVENUTO, F./WEBER, I.: Analyzing the targets of hate in online social media. In: **Proceedings of the Tenth International AAAI conference on Web and Social Media**, 2016. Verfügbar unter: <https://arxiv.org/abs/1603.07709v1>.

SIMONS, A./MUELLER, J.: The Dynamics of Internal Conflict. In: **International Security**, Bd. 4, Nr. 4, 2001, S. 187-192.

SLOAN, R. H./WARNER, R.: Beyond notice and choice. Privacy, norms, and consent. In: **Journal of High Technology Law**, Bd. 14, 2014, S. 370.

WALDRON, J.: **The harm in hate speech**. Cambridge 2012.

WEAVER, S.: A rhetorical discourse analysis of online anti-Muslim and anti-Semitic jokes. In: **Ethnic and Radical Studies**, Bd. 3, N. 36, 2013, S. 483-499.

WEINSTEIN, J./HARE, I.: General introduction. Free speech, democracy, and the suppression of extreme speech past and present. In: WEINSTEIN, J./HARE, I. (Hrsg.): **Extreme speech and democracy**. New York 2009, S. v-ix.