

# Online-Aktivitäten der „Identitären Bewegung“ auf Twitter

## Warum Kontensperrungen die Anzahl an Hassnachrichten nicht reduzieren

**Ruben Wienigk & Dennis Klinkhammer**

*Auf Twitter kann es in den sogenannten Echokammern zu einer Akkumulation radikaler über extreme bis hin zu verfassungsfeindlichen Meinungen kommen. Dahinter stehen oftmals bekannte Akteure/-innen mit einer hohen Reichweite. Social-Media-Plattformen wie Twitter reagieren auf dieses Phänomen mit Kontensperrungen und den Maßnahmen im neuen Gesetzespaket zur Bekämpfung von Hasskriminalität im Internet. Eine Längsschnittanalyse zu den Dynamiken von Echokammern und Hassnachrichten der „Identitären Bewegung“ lässt allerdings vermuten, dass Kontensperrungen das Phänomen weiter verschärfen. Der Beitrag erläutert die Mechanismen und gibt erste Anregungen, welche präventiven Einflussnahmen möglich wären.*

### Echokammern auf Twitter

Social-Media-Plattformen wie Twitter verzeichneten in den letzten Jahren einen kontinuierlichen Anstieg an aktiven Nutzern/-innen (vgl. Pereira-Kohatsu et al. 2019). Mit durchschnittlich 500 Millionen Kurznachrichten pro Tag eröffnet Twitter auf internationaler Ebene einen niedrigschwelligen Kommunikationsraum mit einer hohen Diversität an Meinungen (vgl. Koehler 2015). Dabei kann Twitter nicht als ein soziales Netzwerk interpretiert werden, sondern als mehrere soziale Teilnetzwerke, die sogenannten Echokammern (vgl. Bright 2017).

Diese Echokammern werden nicht nur über Kurznachrichten, sondern auch über Likes und dem Folgen der Accounts anderer Nutzer/-innen generiert und aufrechterhalten. Twitter als Social-Media-Plattform sieht dabei vor, dass der Wechsel von einer in eine andere Echokammer schnell und unkompliziert möglich sein soll (vgl. Prior 2005). Dies führt dazu, dass sich die Nutzer/-innen meist in den Echokammern aufhalten, die ihrer eigenen Meinung entsprechen – das so-

genannte Echo entsteht, welches die eigene Meinung bestätigen und bis zu einer sogenannten Bestätigungsverzerrung führen kann. Dabei konnte bereits nachgezeichnet werden, dass diese Bestätigungsverzerrungen insbesondere in Echokammern mit einschlägigen politischen Meinungen zu einer sukzessiven Akkumulation von radikalen über extreme bis hin zu verfassungsfeindlichen Meinungen beitragen kann (vgl. O'Hara & Stevens 2015). Eine Online-Aktivität der Nutzer/-innen in diesen Echokammern ist das kontinuierliche Diffamieren von Andersdenkenden mit dem Ziel der politischen Einflussnahme (vgl. Glaser & Pfeiffer 2017), weshalb diese nachfolgend auch als Akteure/-innen bezeichnet werden.

In den Jahren 2017 und 2020, dem Analysezeitraum für die hier dargestellten Befunde, zeigt sich insbesondere die „Identitäre Bewegung“ in Deutschland äußerst aktiv darin, eine entsprechende Aufmerksamkeit zu generieren. Vorgelagerte Analysen bestätigen diese Befunde und ließen bereits eine Zunahme entsprechender Online-Aktivitäten erahnen (vgl. Kreißel et al. 2018).

### Hintergrund „Identitäre Bewegung“

Neben einer Ablehnung der repräsentativen Demokratie zeichnet sich die „Identitäre Bewegung“ durch eine subtile Beeinflussung öffentlicher Diskurse aus (vgl. Bundesministerium des Innern, für Bau und Heimat 2018), um extreme und zum Teil verfassungsfeindliche Meinungen aus dem rechten Spektrum anschlussfähig erscheinen zu lassen. Im Verfassungsschutzbericht 2019 wird die „Identitäre Bewegung“ schließlich den parteiungebundenen Strukturen zugeordnet (vgl. Bundesministerium des Innern, für Bau und Heimat 2020), wodurch insbesondere auf Social-Media-Plattformen eine hohe Reichweite erzielt werden kann.

### Kontensperrungen als Analysemöglichkeit

Im Jahr 2020 hat Twitter zwischenzeitlich mit Kontensperrungen bei Vorlage entsprechender Inhalte reagiert, da die Diffamierung von Andersdenkenden sowie der Aufruf zu Straftaten gegen die Regeln der Social-Media-Plattform verstoßen. Fast zeitgleich ist das Gesetzespaket zur Bekämpfung von Hasskriminalität im Internet vom Bundesrat gebilligt worden, welches die Betreiber von Social-Media-Plattformen dazu verpflichtet, entsprechende Inhalte neben den Kontensperrungen auch dem Bundeskriminalamt zu melden (vgl. Bundesregierung 2020; vgl. Bundesministerium der Justiz und für Verbraucherschutz 2020).

Die Kontensperrungen betreffen dabei insbesondere in der Szene

bekannte Akteure/-innen mit einer hohen Reichweite. Da die Kontensperrungen während des Analysezeitraums stattfanden, lassen sich mit dem generierten Datenmaterial nicht nur Effekte auf Ebene der Echokammern nachzeichnen, sondern auch die Online-Aktivitäten auf Ebene der Akteure/-innen selbst. Insbesondere deren unterschiedliche Reichweite innerhalb der sozialen Teilnetzwerke wurde bereits in mehreren aktuellen Forschungsbeiträgen zur automatisierten Erkennung sogenannter Hassnachrichten auf Twitter aufgegriffen (vgl. Ross et al. 2017; Bretschneider & Peters 2017; Wiegand et al. 2018; Mandl et al. 2019 sowie Charitidis et al. 2020), konnte bisher aber nicht in eine Ursache-Wirkungs-Analyse hinsichtlich der Verbreitung radikaler, extremer oder verfassungsfeindlicher Meinungen überführt werden.

Die Hypothese, ob Kontensperrungen von Akteure/-innen mit hoher Reichweite zu einer Reduktion der Anzahl an Hassnachrichten beitragen können, konnte jedenfalls in den bisherigen Forschungsbeiträgen aufgrund des fundamentalen Evaluationsproblems, also der Nichtbeobachtbarkeit einer Kontensperrung vor der tatsächlichen Realisation einer Kontensperrung, nicht empirisch untersucht werden.

Die dahinterstehende Annahme lautet, dass **IA** Kontensperrungen von bekannten Akteure/-innen mit großer Reichweite eine Unterbrechung der Kurznachrichtenkette mit einschlägigen Meinungen bewirken und somit zu einer **IB** Freisetzung von Akteure/-innen innerhalb der Echokammern führen können (siehe Ab-

bildung 1). Akteure/-innen innerhalb einer Echokammer lassen sich dabei stilisiert über einen Kreis zusammenfassen:

### Hassnachrichten auf Social-Media-Plattformen erkennen

Um erstmals die Bedeutung der Reichweite der Akteure/-innen für die Verbreitung radikaler, extremer oder verfassungsfeindlicher Meinungen im Rahmen einer Ursache-Wirkungs-Analyse nachzeichnen zu können, wurde in einem ersten Arbeitsschritt auf Grundlage der zuvor genannten Forschungsbeiträge ein Modell zur automatisierten Erkennung von Hassnachrichten auf Twitter entwickelt. Die Hassnachrichten gelten dabei als Ausdrucksform radikaler, extremer oder verfassungsfeindlicher Meinungen auf Twitter und können unter Rückgriff auf bereits in diesem Kontext erprobte und kontinuierlich erweiterte Textkorpora erkannt werden.

In einem zweiten Arbeitsschritt geht es um die Dynamiken von Hassnachrichten. Dynamiken entstehen, wenn sich Hassnachrichten innerhalb einer Echokammer gegenseitig verstärken. Daher werden zunächst die Dynamiken innerhalb einer Echokammer nachgezeichnet, um im Anschluss die Übertragungswahrscheinlichkeit auf eine andere Echokammer ermitteln zu können. Vereinfacht ausgedrückt geht es bei der Ursache-Wirkungs-Analyse darum, die Dynamiken von Hassnachrichten in einer Echokammer nachzuzeichnen, bevor sie über die Reichweite der Akteure/-innen in andere Echokammern über-

führt werden, um dort gegebenenfalls neue Dynamiken zu entfalten, die ebenfalls nachgezeichnet werden können.

### Quantitative Analyse des Datenmaterials

Seit Beginn des Analysezeitraums konnten Daten von 21 325 Akteure/-innen mit Bezug zur „Identitären Bewegung“ gesammelt werden. Zum Abschluss des Analysezeitraums im Jahr 2020 lagen somit insgesamt 5 055 468 Tweets von bis dahin noch 15 677 aktiven Akteurinnen und Akteuren vor. Die Differenz von 5 648 Akteure/-innen basiert zum einen auf den Kontensperrungen seitens Twitter sowie den damit einhergehenden Unterbrechungen der Kurznachrichtenketten und zum anderen auf der Verwendung von privaten und somit nicht öffentlich einsehbaren Accounts. Ein Teil der relevanten Akteure/-innen ist zudem als Reaktion auf die Kontensperrungen auf andere Social-Media-Plattformen wie Telegram migriert.

Auf Basis von insgesamt 616 724 Tweets lassen sich sogenannte Netzwerkbeziehungen von Akteure/-innen nachzeichnen, welche über ihre einschlägigen Online-Aktivitäten die bereits genannten Echokammern begründen. Der Anteil der als Hassnachrichten erkannten Tweets liegt bei 11,4 %, was im generierten Datenmaterial insgesamt 70 198 Tweets entspricht.

Mithilfe der ebenfalls im generierten Datenmaterial enthaltenen Informationen, wann welche Akteure/-innen wem eine Hassnachricht

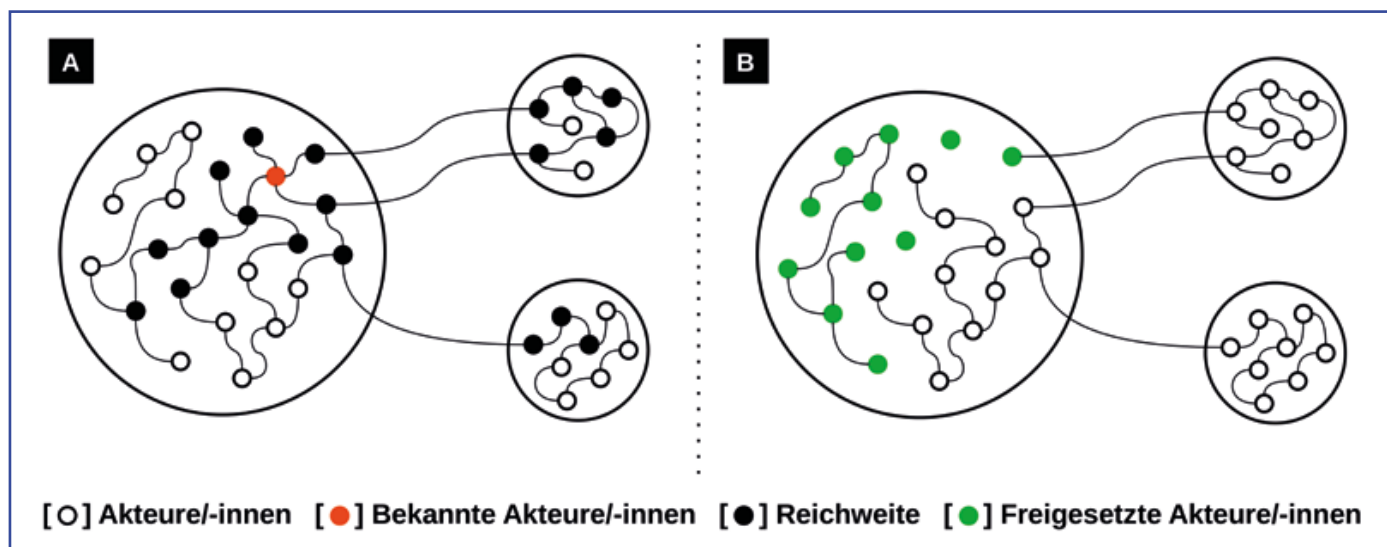


Abbildung 1: Annahme über die Auswirkungen der Kontensperrungen

(Quelle: Eigene Darstellung)

geschrieben haben, konnten zusätzlich Zeitintervalle modelliert, ansteigende Dynamiken von Hassnachrichten identifiziert und kausal den jeweiligen Echokammern zugeordnet werden. Ein Regression Discontinuity Design misst dabei die in der Statistik als Sprungstellen bekannten Variationen im zeitlichen Verlauf (vgl. Klinkhammer/Spermann 2020). Kommt es im Rahmen der Analyse zu einem signifikanten Anstieg in der Anzahl an Hassnachrichten, kann dies als empirischer Befund zu den Dynamiken von Hassnachrichten gesehen werden: Der Zuwachs zeichnet sich dabei sprichwörtlich als Sprungstelle ab. Insgesamt wurden im Rahmen des Analysezeitraums 8 727 Zeitintervalle auf derartige Sprungstellen untersucht, sodass Zeiträume vor und nach den Kontensperrungen abbildbar sind.

## Echokammern transportieren Hassnachrichten dynamisch

Die Kontensperrungen von in der Szene bekannten Akteure/-innen mit hoher Reichweite führen in den entsprechenden Zeitintervallen zwar zunächst zu einer signifikanten Reduktion des Anteils der Netzwerkbeziehungen in den ursprünglich stark frequentierten Echokammern, allerdings wird die Anzahl an Hassnachrichten durch diese Maßnahme insgesamt nur temporär reduziert. Insbesondere die weniger bekannten Akteure/-innen mit geringer Reichweite greifen die Dynamiken von Hassnachrichten aus den ursprünglich stark frequentierten Echokammern auf, setzen die

se in peripheren sowie thematisch nahestehenden Echokammern fort und reichern diese zunächst hinsichtlich der Anzahl an Akteure/-innen an. Mit anderen Worten: Kontensperrungen von in der Szene bekannten Akteure/-innen entfernen zwar in einem ersten Schritt die Redensführer/-innen, bringen in einem zweiten Schritt aber neue Redensführer/-innen an anderer Stelle hervor.

Folglich [C] schrumpfen die ursprünglich stark frequentierten Echokammern bereits kurzfristig, während die peripheren und thematisch nahestehenden Echokammern mittelfristig [D] wachsen (siehe Abbildung 2). Damit reagieren die Echokammern dynamisch auf die Kontensperrungen. Die neuerliche Zunahme der Hassnachrichten in den peripheren sowie thematisch nahestehenden Echokammern lässt zudem vermuten, dass die freigesetzten Akteure/-innen hier in einem weiteren Schritt neue Netzwerkbeziehungen eingehen und somit diese Echokammern langfristig stabilisieren. Dieser Aspekt konnte mit den vorliegenden Zeitintervallen allerdings nicht abschließend analysiert werden, da hierfür weitere Zeitintervalle erforderlich wären.

## Hassnachrichten lassen sich nur temporär reduzieren

Demnach scheint die Reichweite einzelner Akteure/-innen keinen signifikanten Einfluss auf die Übertragungswahrscheinlichkeit von Hassnachrichten zu haben. Vielmehr erfüllt sich die von Twitter als Social-Media-Plattform angebotene Möglich-

keit eines schnellen und unkomplizierten Wechsels von einer in eine andere Echokammer, wovon insbesondere die durch die Kontensperrungen freigesetzten Akteure/-innen Gebrauch machen. Gerade unter den Angeboten von Parteien, einzelnen Politiker/-innen sowie öffentlich-rechtlichen Medien und deren Journalisten/-innen auf Twitter ist nach den Kontensperrungen ein Anstieg in den Dynamiken von Hassnachrichten zu verzeichnen.

Entsprechend ist auch die Anzahl an Hassnachrichten bereits mittelfristig nach den Kontensperrungen, wenn die freigesetzten Akteure/-innen den peripheren und nahestehenden Echokammern beigetreten sind, vergleichbar mit der Anzahl vor den Kontensperrungen. In Einzelfällen scheinen die Kontensperrungen dabei als Katalysatoren für die weniger bekannten Akteure/-innen mit geringer Reichweite gewirkt zu haben, sodass die neu entstandenen Dynamiken von Hassnachrichten als unmittelbare Reaktion auf die Kontensperrungen interpretiert werden können.

Zudem ist aus methodischer Perspektive festzuhalten, dass nach den Kontensperrungen von in der Szene bekannten Akteure/-innen mit hoher Reichweite diese nicht mehr als Zugangspunkt für entsprechende Analysen herangezogen werden können, sondern der Fokus auf weniger bekannte Akteure/-innen mit geringer Reichweite in einer größeren Anzahl an Echokammern auszuweiten ist. Deren Online-Aktivitäten sind zum gegenwärtigen Zeitpunkt weniger systematisch zu erfassen als die in den ursprünglich stark frequentierten Echokammern. Dies erschwert nicht

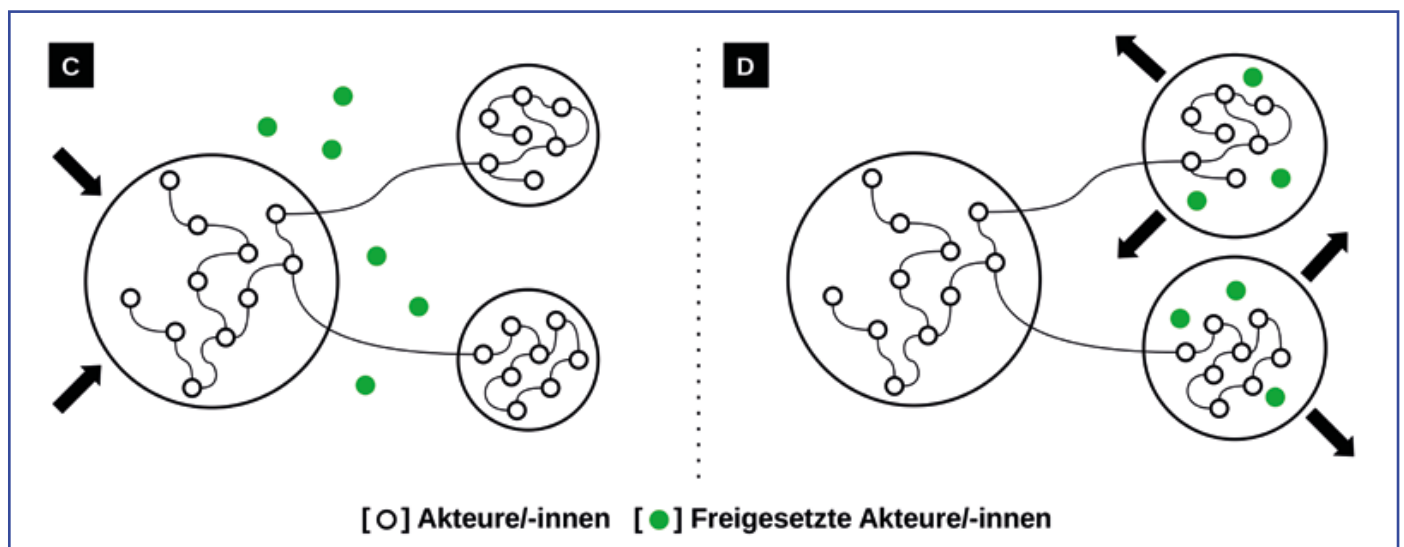


Abbildung 2: Befunde zu den Auswirkungen der Kontensperrungen

(Quelle: Eigene Darstellung)

nur die automatisierte Erkennung von Hassnachrichten für die Anbieter von Social-Media-Plattformen selbst, sondern auch für die Radikalisierungsforschung und Strafverfolgung bei gleichzeitiger Persistenz des Phänomens.

## Alternative Handlungsmöglichkeiten

Eine Alternative zu den Kontensperrungen könnte die gezielte Löschung einzelner Hassnachrichten darstellen. In moderierten Online-Foren und den ebenfalls moderierten Kommentarsektionen auf Online-Nachrichtenseiten wirkt dieses Vorgehen einem Anstieg der Anzahl an Hassnachrichten effektiv entgegen. Die Wirksamkeit dieses Vorgehens gilt es in weiteren Analysen nicht nur quantitativ, sondern insbesondere auch qualitativ auf Social-Media-Plattformen zu überprüfen. Derartige Analysen erfordern allerdings eine aktive Umsetzungsbereitschaft seitens der Betreiber von Social-Media-Plattformen, da die gezielte Löschung einzelner Hassnachrichten gegenüber den Kontensperrungen mit einem Mehraufwand verbunden ist.

Darüber hinaus ist anzumerken, dass es die zugrunde liegenden Algorithmen der Social-Media-Plattformen sind, die Akteure/-innen mit ähnlichen Meinungen überhaupt erst gezielt zusammenführen. Da diese Algorithmen öffentlich nicht einseh-

bar sind, können zum gegenwärtigen Zeitpunkt diesbezüglich keine alternativen Handlungsmöglichkeiten konkretisiert werden. Allerdings können bereits heute Hassnachrichten von allen Nutzern/-innen nicht nur an die Betreiber von Social-Media-Plattformen zur Überprüfung gemeldet werden, sondern auch an eine entsprechende vom Bundesministerium der Justiz und für Verbraucherschutz empfohlene Initiative ([www.hassmel.de](http://www.hassmel.de)). Als aktive Handlung bedeutet dies nicht nur ein Empowerment der Nutzer/-innen von Social-Media-Plattformen gegenüber den durch die Kontensperrungen freigesetzten Akteure/-innen, sondern eröffnet darüber hinaus auch die Möglichkeit von Öffentlichkeitswirksamen Counter Narratives als Input für die Echokammern.

Ruben Wienigk, M. Sc. ist Data Scientist mit Schwerpunkt im Natural Language Processing bei der Deutsche Rückversicherung AG in Düsseldorf

Kontakt: [ruben.wienigk@koeln.de](mailto:ruben.wienigk@koeln.de)

Prof. Dr. Dennis Klinkhammer ist Fachhochschullehrer für Empirische Sozialforschung an der FOM Hochschule in Köln und unterrichtet neben der angewandten Statistik insbesondere die Grundlagen der Programmiersprache R und des maschinellen Lernens

Kontakt: [dennis.klinkhammer@fom.de](mailto:dennis.klinkhammer@fom.de)

### Literatur

Charitidis, P., Doropoulos, S., Vologiannidis, S., Papatogiou, I., & Karakeva, S. (2020). *Towards countering hate speech against journalists on social media*. *Online Social Networks and Media*, 17.

Bretschneider, U., & Peters, R. (2017). *Detecting Offensive Statements towards Foreigners in Social Media*. Online-Artikel: <http://hdl.handle.net/10125/41423> (Stand: April 2021).

Bright, J. (2017). *Explaining the emergence of echo chambers on social media: the role of ideology and extremism*. Online-Artikel: <https://arxiv.org/abs/1609.05003> (Stand: April 2021).

Bundesregierung (2020). *Gesetzespaket gegen Hasskriminalität*. Online-Artikel: <https://www.bundesregierung.de/breg-de/aktuelles/gesetz-gegen-hass-kriminalitaet-1722896> (Stand: April 2021).

Bundesministerium des Innern, für Bau und Heimat (2019). *Verfassungsschutzbericht 2018*. Bexbach: Kern Verlag.

Bundesministerium des Innern, für Bau und Heimat (2020). *Verfassungsschutzbericht 2019*. Bexbach: Kern Verlag.

Bundesministerium für Justiz und Verbraucherschutz (2020). *Gesetzespaket gegen Rechtsextremismus und Hasskriminalität*. Online-Artikel: [https://www.bmiv.de/SharedDocs/Artikel/DE/2020/021920\\_Kabinett\\_Bekaempfung\\_Rechtsextremismus\\_Hass\\_kriminalitaet.html](https://www.bmiv.de/SharedDocs/Artikel/DE/2020/021920_Kabinett_Bekaempfung_Rechtsextremismus_Hass_kriminalitaet.html) (Stand: April 2021).

Glaser, S. & Pfeiffer, Th. (2017). *Erlebniswelt Rechtsextremismus*. Schwalbach am Taunus: Wochenschau Verlag.

Klinkhammer, D., & Spermann, A. (2020). *Einführung in die empirische Kausalanalyse und Machine Learning mit R*. Stuttgart: UTB Verlag.

Koehler, D. (2015). *The Radical Online. Individual Radicalization Processes and the Role of the Internet*. *Journal for Deradicalization*, 15 (1), 116–134.

Kreißel, P., Ebner, J., Urban, A., & Guhl, J. (2018). *Hass auf Knopfdruck. Rechtsextreme Trollfabriken und das Ökosystem koordinierter Hasskampagnen im Netz*. London: Institute for Strategic Dialogue. Online-Artikel: [https://www.isdglobal.org/wp-content/uploads/2018/07/ISD\\_Ich\\_Bin\\_Hier\\_2.pdf](https://www.isdglobal.org/wp-content/uploads/2018/07/ISD_Ich_Bin_Hier_2.pdf) (Stand: April 2021).

Mandl, T., Modha, S., Majumder, P., & Patel, D. (2019). *Overview of the HASOC track at FIRE 2019: Hate Speech and Offensive Content Identification in Indo-European Languages*. New York: Association for Computing Machinery.

O'Hara, K., & Stevens, D. (2015). *Echo Chambers and Online Radicalism: Assessing the Internet's Complicity in Violent Extremism*. *Policy & Internet*, 7 (4), 401–422.

Pereira-Kohatsu, J., Quijano-Sánchez, L., Liberatore, F., & Camacho-Collados, M. (2019). *Detecting and Monitoring Hate Speech in Twitter*. *Sensors*, 19 (21).

Prior, M. (2005). *News vs. Entertainment: How Increasing Media Choice Widens Gaps in Political Knowledge and Turnout*. *American Journal of Political Science*, 49 (3), 577–592.

Ross, B., Rist, M., Carbonell, G., Cabrera, B., Kurowsky, N., & Wojatzki, M. (2017). *Measuring the Reliability of Hate Speech Annotations: The Case of the European Refugee Crisis*. Online-Artikel: <https://arxiv.org/abs/1701.08118> (Stand: April 2021).

Wiegand, M., Siegel, M., & Ruppenhofer, J. (2018). *Overview of the GermEval 2018 Shared Task on the Identification of Offensive Language*. Online-Artikel: [https://epub.oeaw.ac.at/0xc1aa5576\\_0x003a10d2.pdf](https://epub.oeaw.ac.at/0xc1aa5576_0x003a10d2.pdf) (Stand: April 2021).