

ISD

Powering new
generations
against extremism

ichbinhier 

HASS AUF KNOPFDRUCK

Rechtsextreme
Trollfabriken und das
Ökosystem koordinierter
Hasskampagnen im Netz

Philip Kreißel

Julia Ebner

Alexander Urban

Jakob Guhl



Organisationen:

Das **Institute for Strategic Dialogue (ISD)** ist ein unabhängiger „Think and Do Tank“, der mit Führungspersönlichkeiten in Politik, Wirtschaft, Zivilgesellschaft und Wissenschaft zusammenarbeitet, um länderübergreifende Antworten auf die geostrategischen, sozialen und sicherheitsrelevanten Herausforderungen unserer Zeit zu finden. Das Ziel des ISD ist es, Extremismus weltweit zu bekämpfen und interkommunale Gräben zu überbrücken.

Die Facebook-Aktionsgruppe **#ichbinhier** wurde im Dezember 2016 von Hannes Ley nach dem schwedischen Vorbild #jagärhär ins Leben gerufen. Ziel der Gruppe ist es, sich gegen die in den Kommentarspalten der sozialen Medien immer mehr um sich greifende Feindseligkeit und für eine bessere Diskussionskultur einzusetzen. Dies tun die Gruppenmitglieder, indem sie sich zu gemeinsamer Counter Speech (Gegenrede) auf den Facebook-Seiten der großen deutschen Medien verabreden. Ende Juli 2017 wurde der ichbinhier e.V. gegründet, der sich neben der Unterstützung der Aktionsgruppe Bildungsprojekten und der Medienarbeit widmet.

Danksagungen

Die Autoren möchten besonderen Dank richten an Prof. Dr. Martin Emmer vom Institut für Publizistik- und Kommunikationswissenschaft der Freien Universität Berlin, Prof. Dr. Nejdli (Forschungszentrum L3S), Dr. Gerrit Eichner (Justus-Liebig-University Giessen) und Alex Krasodonski-Jones (Center for the Analysis of Social Media) für kritische Blicke auf Methodik und Schlussfolgerungen. Außerdem gilt ihr Dank den Teams von ISD, der Online Civil Courage Initiative (OCCI) und #ichbinhier.

© ISD, 2018

London | Washington DC | Amman | Beirut | Toronto

This material is offered free of charge for personal and non-commercial use, provided the source is acknowledged. For commercial or any other use, prior written permission must be obtained from ISD.

In no case may this material be altered, sold or rented. ISD does not generally take positions on policy issues. The views expressed in this publication are those of the authors and do not necessarily reflect the views of the organisation.

Designed by forster.co.uk. Typeset by Lookandfeelstudio.

Schlüsselerkenntnisse

Zunahme von koordiniertem Hass: Die Analyse von über 1,6 Millionen rechtsextremen Posts in sozialen Medien (Twitter und öffentliche Facebook-Seiten) im Zeitraum Februar 2017 – Februar 2018 zeigt, dass explizit rassistische, antimuslimische und antisemitische Posts seit dem Inkrafttreten des NetzDG im Oktober 2017 zwar abgenommen haben, aber koordinierte rechtsextreme Online-Hasskampagnen seit Dezember im Schnitt mehr als dreimal so weit verbreitet sind (ca. 300.000 Posts/Monat) wie in den zehn vorangegangenen Monaten (ca. 90.000 Posts/Monat).

Lautstarke Minderheit: Die Auswertung von 700 Posts, 16.830 Kommentaren und 1,2 Millionen Likes zeigt, dass 5.500 Accounts (5% aller bei Hateful Speech aktiven Accounts im Januar) für 50% der Likes bei Hass in den Kommentarspalten verantwortlich sind. Etwa 1% der Accounts ist für 25% der Likes verantwortlich.

Koordinierte rechtsextreme Accounts dominieren:

Der Diskurs in vielen Kommentarspalten auf Facebook ist kein Abbild der Gesellschaft, sondern wird von Sympathisanten extremistischer und verfassungsfeindlicher Organisationen bestimmt. Es handelt sich in vielen Fällen um gezielte Versuche, Politiker, Journalisten und Aktivisten unter Druck zu setzen und einzuschüchtern.

Einflussreiche Kampagnen: In den letzten Monaten identifizierten wir drei große Hasskampagnen, die von Rechtsextremen initiiert und koordiniert wurden: #Kikagate, #Kandelistueberall und #120dB. Die pro-AfD-Wahlkampagne mit den Hashtags #AfD, #TraudichDeutschland und #Merkelmussweg im August-September 2017 war die längste und einflussreichste Kampagne der rechtsextremen Troll-accounts.

Koordinierung in rechtsextremen Chats: Diese Kampagnen werden von Reconquista Germanica und verwandten rechtsextremen Online-Netzwerken aus identitären Kreisen in den sozialen Medien zu koordinierten Uhrzeiten und Hashtags verbreitet, was dazu führt, dass sie teilweise wochenlang in den Top-Trends liegen und den Online-Diskurs bestimmen.

Mainstreaming durch Politik und Medien: Die Hashtags werden außerdem von AfD-Accounts und von russischen Medien wie RT und Sputnik aufgegriffen und geraten so in den medialen Mainstream, der in diesem Zusammenhang mit dem Begriff der „Massenmedien“ gleichgesetzt wird. RT und Sputnik waren unter den ersten Medien, die von den Kampagnen berichteten. Auch Mainstream-Medien griffen die Themen der Identitären, aber nicht deren Kampagnen auf.

Inhaltsverzeichnis

Organisationen	2
Danksagungen	2
Schlüsselerkenntnisse	2
Inhaltsverzeichnis	3
Glossar	5
I. Einleitung (Summary)	7
Analyseebenen	8
Hintergrund ISD-Jahresrückblick: Hate Speech in Deutschland	8
II. Lautstarke Minderheit: #ichbinhier analysiert die Kommentare auf Facebook	9
1. Eine Facebook-Gruppe setzt sich für den Anstand ein	9
1.1 Tätigkeitsrahmen von #ichbinhier	9
1.2 Weitere qualitative Kriterien	10
2. Erste Erkenntnisse	10
III. Erhebung und Auswertung von Facebook-Daten	11
1. Hateful Speech messbar machen	11
1.1. Schaffen einer Datengrundlage	11
1.2 Hateful Speech	11
1.3 Identifizierung von „Hate-Speech-Accounts“	11
1.4 Eine kleiner Anteil User sorgt für Likes für Hateful Speech	12

Inhaltsverzeichnis

2. Wer steckt hinter den Accounts?	13
2.1 Vorüberlegungen	13
2.2 Die Rolle rechter Organisationen: AfD und Identitäre Bewegung	13
2.3 Koordination	14
2.4. Beispiele	15
IV. Hintergründe der Hate-Speech-Koordinierung	16
1.1 Hassverbreitung: Koordination durch organisierte Trollfabriken	16
1.2. Kampagnen	19
1.3. Die Rolle von Bots	22
1.4. Die Rolle der deutschen Medien	23
V. Limitationen	24
1.1. Klassifizierung von Hate-Speech-Accounts	24
1.2. E-Sympathisant	24
1.3. Unterschiede in der Vorgehensweise von #ichbinhier und Reconquista Germanica	24
Schlussfolgerungen	25
Empfehlungen: Koordinierte Zivilcourage gegen koordinierten Hass	26
Weiterführende Literatur	27
Endnotes	28

Glossar

#120db:

#120db ist eine Kampagne von *Aktivisten der Identitären Bewegung, welche auf die vermeintliche Gefährdung europäischer Frauen durch muslimische Einwanderer aufmerksam machen will.

„Der große Austausch“:

Eines der zentralen Konzepte der Identitären Bewegung behauptet, dass die einheimischen europäischen Bevölkerungen durch nicht-europäische Einwanderer ausgetauscht werden sollen.

Discord:

Discord ist eine Online Chat App, die ursprünglich für Fans von Videospiele entwickelt wurde, mittlerweile aber auch etliche rechtsextreme Kanäle umfasst.

E-Sympathisanten:

E-Sympathisanten sind in dieser Arbeit Accounts, nicht zwangsläufig reale Personen, die in den sozialen Medien Inhalte, die von Parteien bzw. Organisationen wie der Identitären Bewegung gepostet wurden, durch positive Interaktionen (Likes) unterstützen und so ihre Übereinstimmung mit den geteilten bzw. geposteten Inhalten signalisieren.

Hasskampagnen:

Hasskampagnen sind koordinierte Online-Aktionen mit einer hohen Konzentration von „Hate Speech“ mit dem Ziel, Gegenrede zu unterdrücken, Meinungen zu beeinflussen, Ressourcen der Redaktionen zu binden und Kommentatoren sowie Urheber eines Videos, Artikels oder Fotos direkt, beleidigend und herabwürdigend anzugehen. Im Rahmen dieser Veröffentlichung wird der Begriff „Shitstorm“ dem Begriff der „Hasskampagne“ gleichgesetzt.

Hasskommentare:

Hasskommentare sind Kommentare unter Posts in sozialen Medien, die unter unsere nachfolgende Definition von Hate Speech fallen.

Hate Speech (Hassrede):

Hate Speech als Begriff ist vergleichsweise neu und bislang nicht abschließend und verbindlich definiert. Das Ministerkomitee des Europarats bezeichnet

Hate Speech als „jegliche Ausdrucksformen, welche Rassenhass, Fremdenfeindlichkeit, Antisemitismus oder andere Formen von Hass, die auf Intoleranz gründen, propagieren, dazu anstiften, sie fördern oder rechtfertigen, unter anderem Intoleranz, die sich in Form eines aggressiven Nationalismus und Ethnozentrismus, einer Diskriminierung und Feindseligkeit gegenüber Minderheiten und Menschen mit Migrationshintergrund ausdrückt.“¹¹

Hateful Speech:

Als Hateful Speech werden in der Studie Posts und Kommentare bezeichnet, die in ihrer Gesamtheit deutlich mehr Hate Speech enthalten.

Identitäre Bewegung:

Die Identitäre Bewegung (IB) ist eine rechtsextreme Bewegung, die den „Erhalt der ethnokulturellen Identität“ und die „Verteidigung des Eigenen“ gegen Multikulturalismus, Globalisierung und Islamisierung als ihre wichtigsten Ziele ansieht. Sie wird wegen personeller Überschneidungen zum Rechtsextremismus bundesweit vom Verfassungsschutz beobachtet.

#Infokrieg:

#Infokrieg ist eine rechtsextreme deutschsprachige Discord-Gruppe, die unter anderem versucht hat, die Bundestagswahl 2017 mithilfe von koordinierten Online-Kampagnen zu beeinflussen. Der Name der Gruppe ist eine direkte Übersetzung von „Info Wars“, der Website der US-amerikanischen Verschwörungstheoretiker Alex Jones und Paul Joseph Watson.

Kandel:

Ende Dezember 2017 erstach in Kandel (Rheinland-Pfalz) ein afghanischer Flüchtling seine 15-jährige Ex-Freundin. Der Fall löste in den sozialen Medien Diskussionen aus und wurde vor allem von rechten Gruppen als Beispiel für eine gescheiterte Immigrationspolitik gedeutet.

KiKAgate:

In einer Folge der Serie „Schau in meine Welt“ portraitierte der Fernsehkanal Kinderkanal (KiKA) Ende November 2017 die Beziehung zwischen einem syrischen Flüchtling (dessen genaues Alter umstritten war) und einer 15-jährigen Deutschen. Die Sendung führte Anfang Januar 2018 zu einem Shitstorm

hauptsächlich durch rechte Gruppen in den sozialen Medien.

Massenmedien bzw. Mainstream-Medien:

Massenmedien erreichen indirekt über verschiedene Kanäle (visuell, auditiv, audio-visuell) eine nicht näher eingeschränkte oder definierte Empfängerschaft. Der sog. Medien-Mainstream kann als „ein mehr oder weniger weitgehender medialer Konsens in bestimmten Fragen, oder auch: eine Anzahl von Themen und Meinungen, die in einem bestimmten Zeitraum in der Medienlandschaft dominiert und damit eine ‚Hauptströmung‘ oder eine ‚Haupttrichtung‘ bildet“, charakterisiert werden.²

NetzDG:

Das Netzwerkdurchsetzungsgesetz (NetzDG) ist ein deutsches Gesetz, mit dem Hate Speech und Fake News in den sozialen Medien bekämpft werden sollen. Es verpflichtet die Plattformbetreiber, „offensichtlich rechtswidrige“ Inhalte innerhalb von 24 Stunden zu entfernen, nachdem sie ihnen gemeldet wurden.

Rechtsextremismus:

Laut dem Rechtsextremismus-Experten Cas Mudde (2000) zeichnen sich rechtsextreme Gruppen und Individuen durch mindestens drei der folgenden fünf Merkmale aus: Nationalismus, Rassismus, Xenophobie, Gegnerschaft zur Demokratie und die Sehnsucht nach einem starken Staat.

Rechtspopulismus:

Der Rechtsextremismus-Experten Cas Mudde (2004) bezeichnet Populismus "eine Ideologie, die davon ausgeht, dass die Gesellschaft in zwei homogene, antagonistische Gruppen getrennt ist, das 'reine Volk' und die 'korrupte Elite', und die geltend macht, dass Politik ein Ausdruck der *volonté générale* oder des allgemeinen Volkswillens sein soll."³

Reconquista:

Der Begriff „Reconquista“ (die Rückeroberung der iberischen Halbinsel von den Mauren) stellt aus Sicht der Identitären Bewegung eine historische Analogie zur aktuellen Situation dar, in der die Identitären sich ebenfalls mit einer muslimischen Invasion konfrontiert sehen.

Reconquista Germanica:

„Reconquista Germania“ ist eine rechtsextreme Discord-Gruppe, die das primäre Ziel verfolgt, den Online-Diskurs zu beeinflussen und somit Druck auf die Politik auszuüben. Den Mitgliedern der Gruppe werden auf Basis von strikten Hierarchien, klar verteilten thematischen und geographischen Verantwortungsbereichen und mit militärischem, an die Wehrmacht angelehntem Vokabular Anweisungen erteilt.

Remigration:

Die „Remigration“ (Zurückführung) nicht-europäischstämmiger Bevölkerungsgruppen ist eine der zentralen Forderungen der Identitären Bewegung, die darin die notwendige Antwort auf den „großen Austausch“ sieht.

Troll bzw. Trollaccount:

Unter einem Troll bzw. Trollaccount wird ein Nutzer verstanden, der durch Provokationen gezielt die Kommunikation stört und damit beim Gesprächspartner eine emotionale Reaktion hervorrufen möchte, mit dem Ziel, Konflikte zu schüren und Meinungen zu beeinflussen. Trolling bzw. Trollkampagnen oder Trollgruppen können damit als zielgerichtetes, gemeinsames Auftreten solcher Trollaccounts verstanden werden.

I. Einleitung (Summary)

Seit einigen Jahren ist eine immer gereiztere Stimmung in unserer Gesellschaft zu beobachten, die in besonderem Maße in den Kommentarspalten der sozialen Medien zum Ausdruck kommt. Beleidigungen und Bedrohungen, die Ausgrenzung und Anfeindung ganzer Bevölkerungsgruppen und sogar die unverhohlene Befürwortung von Gewalt sind auf Facebook, Twitter und Co. an der Tagesordnung. Eine sachliche, konstruktive Auseinandersetzung findet dort, wo abwertende, aggressive und pauschalisierende Kommentare dominieren, nicht mehr statt.

Ziel dieser gemeinsamen Studie vom ISD und ichbinhier e.V. ist es, die Ursprünge und Verbreitung von sogenannter Hate Speech im Netz besser zu verstehen, um darauf basierend effektivere Methoden zu deren Identifizierung bzw. Eindämmung zu entwickeln und damit die Diskussionskultur in den Kommentarspalten der sozialen Medien wieder so zu verbessern, dass sich auch die stillen Nutzer sozialer Medien wieder an Debatten zu gesellschaftlichen oder politischen Themen beteiligen möchten.

„[Gegenrede] habe ich früher mal versucht, jedoch wieder damit aufgehört, weil mir irgendwann meine Zeit und auch mein eigener Seelenfrieden zu schade waren [...]“

Facebook Kommentar bei SZ-online

Es wird daher einleitend in einem ersten Schritt (Kapitel II) der selbst festgelegte Tätigkeitsrahmen der Gruppe #ichbinhier näher erläutert, in dem zunächst auf Grundlage qualitativer Kriterien nach Hate Speech gesucht wurde. Die Identifizierung von Kommentaren als Hate Speech begegnet dabei gewissen Schwierigkeiten. Neben dem Ansatz, sich schlicht auf strafrechtliche Tatbestände wie den der Volksverhetzung zu stützen, nähert man sich dem Phänomen vor allem sprachwissenschaftlich (Kapitel II, Pkt. 1.1). Allerdings ließe sich auch dann, wenn eine verbindliche Definition zugrunde gelegt würde, im Einzelfall darüber streiten, ob ein bestimmter Kommentar unter diese Definition fällt oder nicht. Ein automatisiertes Auffinden von Hate Speech, welches sich an einer solchen linguistischen Definition orientiert, ist derzeit nicht möglich. Hinzu kommt, dass es Online-Aktivist*innen immer wieder gelingt, Einschüchterungs- und Hasskampagnen gegen Einzelakteure wie Organisationen so zu gestalten, dass sie die neuen Bestimmungen des NetzDG umgehen. Daher verfolgen die Autoren hier explizit nicht das Ziel, ausschließlich Hate Speech im zweifelsfreien, engeren

Sinne zu identifizieren. Im weiteren Verlauf wird daher der Begriff Hateful Speech eingeführt.

Wenig überraschend scheint die Erkenntnis, dass bestimmte Themen ein besonders hohes Kommentaraufkommen hervorrufen. Im Kapitel II Pkt. 1.2 und 1.3 werden die Grundlagen dafür beschrieben, Hate Speech bzw. Hateful Speech im Rahmen einer Datenanalyse ausfindig zu machen. Im Zuge dieser Datenauswertung hat sich gezeigt, dass viele Kommentarspalten von Accounts dominiert werden, die einen rechten bis rechtsextremen Hintergrund haben. (Kapitel II, Pkt. 3).

Es soll nicht davon abgelenkt werden, dass es neben rechtsextremen Gruppen wie Reconquista Germanica und #Infokrieg auch andere extremistische Akteure gibt, welche die sozialen Medien für ihre politischen Ziele missbrauchen. Allerdings haben sowohl #ichbinhier als auch das ISD und die OCCI in ihren Analysen der letzten Monate immer wieder festgestellt, dass es sich bei Hate Speech aus dem rechtsextremen Spektrum um ein besonders virulentes Problem handelt. Im deutschsprachigen Raum existiert nach den Beobachtungen dieser Initiativen gegenwärtig weder in linksextremistischen noch in islamistischen Kreisen ein vergleichbares Maß an koordinierten Hasskampagnen. Die Koordination der rechtsextremen Kampagnen ist Gegenstand des Kapitels IV.

Besonders islamistische Extremisten sind aufgrund der verstärkten Maßnahmen sozialer Medienunternehmen im letzten Jahr mittlerweile weniger präsent auf den großen Plattformen und in ihrer Kommunikation stark abhängig von verschlüsselten Nachrichten-Applikationen wie Telegram. Dies führt auch dazu, dass islamistische Gruppen, die bis Anfang 2017 häufig noch offen extremistische Inhalte gepostet haben, sich mittlerweile auf unpolitische religiöse Posts beschränken (Fielitz et al. 2018).

Die Namen und Accounts einzelner User wurden zum Schutz ihrer Privatsphäre ausgeblendet. Es wurde sichergestellt, dass die erfassten Daten in aggregierter Form, die keine Rückschlüsse auf individuelles Verhalten zulassen, veröffentlicht werden.

Für die vorliegende Untersuchung der Verbreitung von Hate Speech auf Facebook und der Intensität der Beteiligung rechtsextrem orientierter User an Hate Speech wurden die folgenden drei Analyseebenen untersucht, die in der Folge detaillierter beschrieben werden.

Analyseebenen

Analyse-Level 1: Jahrestrends-Analyse

Zunächst wurde mithilfe des Digital Analytics Tools Crimson Hexagon eine Jahresanalyse zu den allgemeinen Entwicklungen in Bezug auf Hass im Netz durchgeführt. Insbesondere Trends im Volumen von koordinierten Hass-Posts und geteilten Medienartikeln wurden festgehalten. Insgesamt wurden hierfür 1,6 Mio. rechtsextreme Posts im Zeitraum von Februar 2017 bis Februar 2018 ausgewertet.

Analyse-Level 2: Hasskommentaranalyse

Mithilfe der statistischen Programmiersprache R von (R-Studio) wurden in Verbindung mit dem frei verfügbaren R Facebook Plugin sowie Facebooks eigenen API-Funktionen (API Version 2.9) drei Datensätze ausgewertet. Datensatz I bezieht sich auf #ichbinhier Aktionen. Datensatz II enthält 700 Kommentarspalten, die vermutlich Hateful-Speech enthalten. Datensatz III enthält einen Querschnitt aus 1000 Kommentarspalten.

Analyse-Level 3: Hassverbreitungsmethoden

Die Datenanalyse wurde ergänzt mit einer 6-monatigen ethnografischen Recherche innerhalb der verschlüsselten Chats deutscher Rechtsextremisten auf der Gaming-Plattform Discord, wo die Hasskampagnen koordiniert wurden. Diese qualitative Langzeitbeobachtung ermöglichte uns, die Strategien und Techniken zur Hassverbreitung zu analysieren und die auf Level 2 aufgezeigten Effekte zu erklären.

Hintergrund ISD-Jahresrückblick: Hate Speech in Deutschland

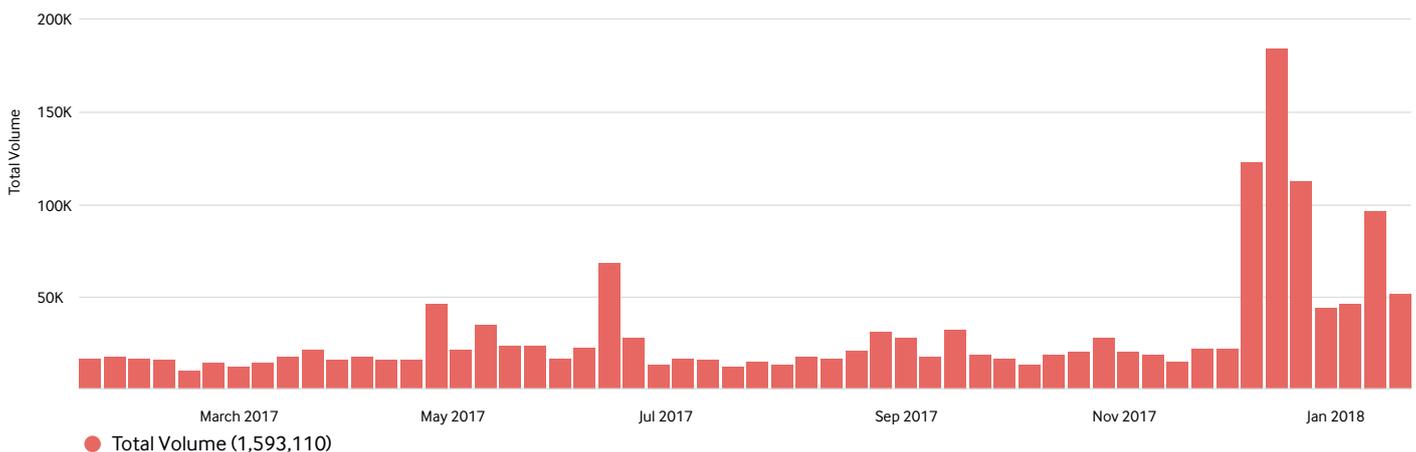
Mithilfe des Digitalanalyseinstruments Crimson Hexagon, das linguistische Analysemöglichkeiten mit Machine Learning kombiniert, konnten die Entwicklungen der Anzahl von Posts in Zusammenhang mit rechtsextremen Hashtags und Schlüsselwörtern untersucht werden. Insgesamt analysierten wir über 1,6 Mio. Posts auf Twitter und öffentlichen Facebook-Seiten im Zeitraum von Februar 2017 bis Februar 2018.

Auf Basis unserer 6-monatigen ethnografischen Recherchen in rechtsextremen Foren und Chatgruppen wurden Schlüsselwörter und Hashtags identifiziert, die im Zusammenhang mit koordinierten Hasskampagnen in rechtsextremen Discord-Gruppen festgelegt und verwendet wurden. Die folgenden wiederkehrenden Hashtags und Schlüsselwörter wurden dementsprechend für die Jahresanalyse verwendet:

#120dB, #Kandel, #kandelistueberall, Flüchtlingskriminalität, #NetzDG, #Zensur, #120DEZIBEL, Merkelstan, Remigration, Grenzdicht, #dergroßeaustausch, halbneger, Masseneinwanderung, Überfremdung, Islamisierung, Krimigranten, Einzelfall, Reconquista.

Wir trainierten den verwendeten Algorithmus darauf, zwischen positiven und negativen Erwähnungen der Schlüsselwörter und Hashtags zu unterscheiden sowie in welchen Medien diese erwähnt wurden.

Dabei stellten wir eine starke Zunahme von Posts im Zusammenhang mit koordinierten Hasskampagnen in sozialen Netzwerken (Twitter und öffentliche Facebook-Seiten) fest. Während es zwischen Februar 2017 und November 2017 im Durchschnitt etwa 90.000 Posts pro Monat im Zusammenhang mit rechtsextremistisch koordinierten Kampagnen gab, identifizierten wir zwischen Dezember 2017 und Februar 2018 im Durchschnitt über 300.000 solcher Posts pro Monat.



Um Hate Speech von zum Teil berechtigter Kritik oder auch schwarzem Humor abzugrenzen, orientierte sich das Moderatoren-Team zunächst einmal an der Definition, die der Duden für den Begriff der Hetze anbietet. Ganz ähnliche Definitionen lassen sich auch in anderen enzyklopädischen Werken finden. Danach handelt es sich bei Hetze um die „Gesamtheit unsachlicher, gehässiger, verleumderischer, verunglimpfender Äußerungen oder Handlungen (...), die Hassgefühle, feindselige Stimmungen und Emotionen gegen jemanden oder etwas erzeugen“.

Daneben zogen die Moderatoren die Veröffentlichungen des Ministerkomitees des Europarates von 2007 und von Prof. Anatol Stefanowitsch von der Freien Universität Berlin zu Rate. Auf dieser Grundlage verstehen die Moderatoren unter dem Begriff der Hate Speech die gezielte Herabsetzung und Demütigung von einzelnen Personen, Gruppen, Gemeinschaften oder Ethnien. Personen werden aufgrund bestimmter, ihnen zugeschriebener Eigenschaften - Abstammung, Herkunft, Hautfarbe, Geschlecht, sexuelle Orientierung, Religion, körperliche Einschränkung, aber auch Parteizugehörigkeit, Weltanschauung, Beruf o. a. - als "anders" markiert und ausgegrenzt. Anknüpfend an diese Eigenschaften werden bewusst Desinformationen oder falsche Aussagen verbreitet, es werden Stereotype und Vorurteile bedient und die Betroffenen werden aufgrund ihrer Gruppenzugehörigkeit beleidigt, bedroht, verächtlich gemacht und entmenschlicht: Beispiele sind Begriffe wie "Asylantenflut" oder "Viehzeug". Dies steigert sich bis hin zur Befürwortung von oder Appellen zu körperlicher Gewalt gegen die Betroffenen.

1.2 Weitere qualitative Kriterien

Die Moderatoren betrachteten allerdings nicht nur den Inhalt und die Tonalität der verfassten Kommentare nach Maßgabe der vorstehend dargestellten Kriterien. Daneben verfolgten sie die Entwicklung der Kommentarspalten unter Artikeln auf Facebookpräsenzen der großen Medienhäuser. Hierbei fiel auf, dass das Kommentaraufkommen stark mit dem Thema des betreffenden Artikels korrelierte:

Vor allem Berichte zu Ausländerkriminalität oder terroristischen Anschlägen sowie Artikel mit Bezug auf Flucht oder auf Muslime in Deutschland, aber auch (wenn auch deutlich seltener) Beiträge mit einem sozialpolitischen Hintergrund wurden von Kommentatoren signifikant stärker frequentiert. Unter solchen Beiträgen fanden sich innerhalb einer Stunde mehrere hundert Kommentare. Die ersten Kommentare

zu diesen Themen wiesen schon nach wenigen Minuten zwei- oder gar dreistellige Like-Zahlen auf.

Offenbar, und das konnte über den gesamten Zeitraum seit Gründung der Gruppe bis Januar 2018 festgestellt werden, gibt es einen direkten Zusammenhang zwischen den Themen im Spannungsfeld „Zuwanderung“ und der Anzahl der Kommentare bzw. deren Like-Zahlen. Inhaltlich positionierten sich die Verfasser der als problematisch identifizierten Kommentare besonders stark gegen Zuwanderung, Muslime und Politiker.

Es ist zu betonen, dass #ichbinhier bei Häufungen aller Arten von abwertenden Kommentaren aktiv wird. Beispielsweise finden sich in den Kommentarspalten bisweilen auch Angriffe auf die Würde von AfD-Politikern, denen #ichbinhier-Mitglieder dann ebenfalls widersprechen. Allerdings nehmen diese Kommentare selten in der oben beschriebenen Form quantitative und qualitative Ausmaße an wie in den anderen Bereichen, und ihnen wird auch ohne Zutun von #ichbinhier schnell widersprochen. Daher wird #ichbinhier in diesem Bereich sehr selten koordiniert aktiv.

2. Erste Erkenntnisse

Bezogen auf den eingangs aufgezeigten Tätigkeitsrahmen war es über den gesamten Zeitraum bis Januar 2018 auffällig, dass letztlich fast ausschließlich Medien-Artikel, welche die o.g. Sachthemen behandelten, zu koordinierten #ichbinhier-Aktionen führten, obwohl die Moderatoren sich bei der Auswahl nicht in erster Linie an den Themen, sondern an dem Kommentaraufkommen und dem Anteil von Hate Speech in einer Kommentarspalte orientierten. Dies lässt die (zunächst qualitative) Vermutung zu, dass (auf-)hetzende Kommentare unter Medien-Artikeln des o.g. Tätigkeitsrahmens eher von Anhängern des rechtspopulistischen oder rechtsextremen Spektrums stammen. Auf den vom Moderatorenteam beobachteten, reichweitenstarken Medienseiten ließen sich zumindest weder ein hohes Kommentaraufkommen noch Hate Speech ausmachen, die dem linken Spektrum zuzuordnen gewesen wären.

Dies war der Anlass, den Versuch zu unternehmen, relevante Medien-Artikel anhand von Aktivitätsmustern automatisiert zu identifizieren. Die gewählte Vorgehensweise und die gewonnenen Erkenntnisse sind Gegenstand des nachfolgenden Abschnitts.

III. Erhebung und Auswertung von Facebook-Daten

1. Hateful Speech messbar machen

1.1. Schaffen einer Datengrundlage

Ziel war es, verschiedene Daten aus den Kommentarspalten der Medien auf Facebook zu sammeln und im Kontext von Hate Speech zu untersuchen. Dazu wurden die folgenden Fragen beantwortet:

Wie groß ist die Gruppe von Personen bzw. in diesem Falle Accounts, die in den Kommentarspalten der Medien-Artikel Hate Speech verbreitet? Lassen sich die Accounts politisch einordnen? Inwiefern gehen diese Accounts gemeinsam und koordiniert vor?

Da die Gesamtheit aller Hasspostings nicht klar beschreibbar und zugänglich ist, wurde eine Datengrundlage aus drei Datensätzen erstellt, die, unter den technischen Gegebenheiten und den Bedingungen dieser Studie, so repräsentativ wie möglich ist. Die hier beschriebenen Untersuchungen wurden mit R Studio durchgeführt (technische Details: siehe Supplement).

Zuerst wurden alle Top-Level-Kommentare aus den koordinierten #ichbinhier-Aktionen im Monat Dezember 2017 heruntergeladen und analysiert (Datensatz I).

#ichbinhier-Aktionen alleine bieten keinen generellen Überblick über die Kommentarspalten von Medien auf Facebook insgesamt. Dieser Überblick ist aber notwendig, um das Verhalten der Accounts längerfristig und umfassender beobachten zu können. Daher wurden ausgehend von #ichbinhier-Aktionen im Dezember weitere 700 ähnliche Kommentarspalten bei 16 Medien im Januar analysiert (Datensatz II). Es wird gezeigt, dass es sich hierbei um Kommentarspalten mit einem erhöhten Anteil an Hate Speech handelt.

Um einen Gesamtüberblick zu gewinnen, die Methodik zu validieren und Koordination nachzuweisen wurde ein Querschnitts-Datensatz heruntergeladen (Datensatz III). Dabei wurden die letzten 100 Posts von 10 Medienseiten geladen.

1.2 Hateful Speech

Wie bereits erläutert, ist der Begriff der Hate Speech schwierig zu fassen. Da eine algorithmisierte und automatisierte Analyse bisher nicht möglich ist, wurde daher an dieser Stelle eine vereinfachte Definition von Hate Speech, der Begriff „Hateful Speech“, benutzt:

Als „Hateful Speech“ wird eine Gruppe von Kommentaren bezeichnet, die signifikant höhere Anteile an Hate Speech beinhaltet als die Gesamtheit aller Kommentare auf Facebook. Das Ziel ist es demnach nicht, Hate Speech zu identifizieren (was infolge der Auswirkungen des NetzDG und der Moderationstätigkeit auf den Medienseiten vermutlich nicht ausreichend Daten liefern würde), sondern die Wahrscheinlichkeit für das Antreffen von Hate Speech in den untersuchten Daten so zu erhöhen, dass die Analyse des ursächlichen Verhaltens im Datenmaterial möglich wird. Es handelt sich bei „Hateful Speech“ um ein methodisches Konstrukt. Die Schlussfolgerungen dieser Arbeit beruhen nicht auf der Bewertung von einzelnen Kommentaren als Hate Speech und sind somit so objektiv wie möglich.

Durch dieses Verfahren kann erstens sichergestellt werden, dass eine große Bandbreite an Themen und Formen von Hate Speech Eingang in die Stichprobe findet und das Risiko, dass es Verzerrungen in Richtung bestimmter Themen gibt, gering gehalten wird. Zweitens ist die Wahrscheinlichkeit hoch, dass alle aktuellen Debatten und Themen entdeckt werden, die zu einem bestimmten Zeitpunkt in der allgemeinen und speziellen (Facebook-)Öffentlichkeit dominant und durch besonders starke Hate Speech geprägt sind.

Bei #ichbinhier-Aktionen (Datensatz I), die auf ein deutlich erhöhtes Aufkommen von Hate Speech schließen lassen, gelten aus methodischer Sicht alle Kommentare als Hateful Speech, die keine Likes von #ichbinhier-Mitgliedern erhalten. Zur Erinnerung: „Hateful Speech“ ist hier kein normatives Label für den einzelnen Kommentar. Grundlage ist hier das große und überparteiliche Netzwerk von #ichbinhier. Es kann hier unabhängig von der Bewertung einzelner Aktionen davon ausgegangen werden, dass diese Daten in ihrer Gesamtheit mit höherer Wahrscheinlichkeit Hate Speech enthalten als normale Kommentare.

1.3 Identifizierung von „Hate-Speech-Accounts“

Auf der Basis und unter Zuhilfenahme von Datensatz I („Aktionen von #ichbinhier im Dezember 2017“) wurde in einem mehrstufigen Prozess der größere Datensatz II erzeugt:

Schritt 1: Da es bei dieser Analyse darum geht, das Likeverhalten der Nutzer zu analysieren, wurden zunächst angelehnt an die vorstehende Definition Accounts identifiziert, die mit Hateful-Speech-Kommentaren positiv, z. B. in Form von „Likes“, interagieren. Aus Datensatz I („Aktionen von

#ichbinhier im Dezember 2017“) wurden die 1.000 aktivsten Nicht-ichbinhier-Accounts ausgewählt, die bei mindestens vier #ichbinhier-Aktionen im Betrachtungszeitraum Dezember 2017 Hateful-Speech-Kommentare unterstützt haben („wahrscheinliche Hate-Speech-Accounts“).

Schritt 2: Um auch Beiträge in die Analyse miteinzubeziehen, die nicht als besonders prägnante oder prominente Fälle Gegenstand von #ichbinhier-Aktionen wurden, wurden diejenigen Medien-Artikel bei 16 Medienseiten ausgewählt, die eine intensive Beteiligung dieser im Dezember 2017 ermittelten 1.000 Accounts aufwiesen, woraus geschlossen werden kann, dass diese Medien-Artikel mit einer hohen Wahrscheinlichkeit Hateful-Speech-Kommentare enthalten. Selbst wenn einzelne Accounts falsch zugeordnet worden wären, so erlaubt die große Anzahl gleichzeitig auftretender Accounts diese Schlussfolgerung.

Schritt 3: Kommentare mit Likes von #ichbinhier-Mitgliedern wurden von der Analyse ausgeschlossen, denn es kann davon ausgegangen werden, dass diese in ihrer Gesamtheit mit geringerer Wahrscheinlichkeit Hateful Speech enthalten. Eine „Verunreinigung“ der Daten durch koordinierte #ichbinhier-Aktionen wird so ebenfalls ausgeschlossen. Auch Kommentare mit weniger als 10 Likes wurden nicht berücksichtigt, um die Datenmenge gering zu halten und da hier aufgrund der geringen Like-Zahl die Wahrscheinlichkeit größer ist, Kommentare anhand ihrer Likes falsch zu klassifizieren.

Es ergab sich ein Datensatz von insgesamt ca. 14 000 Kommentaren, die jeweils mehr als 10 Likes erhielten, verfasst in den Kommentarspalten unter 700 Medien-Artikeln, mit wiederum insgesamt ca. 600 000 Likes („Datensatz II“).

Das Vorgehen bei der Auswahl des Untersuchungsmaterials sollte sicherstellen, dass die Analysegrundlage so repräsentativ wie möglich ist. Dabei wurde erstens auf die Vielfalt und große Zahl der Mitglieder des #ichbinhier-Netzwerks sowie zweitens auf das Nachverfolgen von besonders aktiven Autoren von Hate Speech geprägten Kommentaren/Likes gesetzt. Dadurch kann die Wahrscheinlichkeit, dass relevante und wichtige von Hate Speech geprägte Debatten übersehen wurden, als gering eingeschätzt werden. Da es sich bei der Datengrundlage jedoch nicht um eine systematische Zufallsstichprobe handelt und es sich wegen der technischen Rahmenbedingungen von Facebook auch nicht darum handeln kann, bleibt hinsichtlich der konkreten Befunde immer eine gewisse Unsicherheit, weshalb geringere Unterschiede zwischen Gruppen oder Typen nur mit Vorsicht interpretiert werden sollten. Die tatsächlichen Zustände können darüber hinaus auch

für spezielle Zeiträume oder isolierte Themen oder Forenbereiche von den hier vorgestellten Befunden abweichen.

Um zu überprüfen, ob der extrahierte Datensatz tatsächlich mehr Hate Speech enthielt, wurden manuell von zwei verblindeten und unabhängigen Untersuchern Stichproben mit der Definition von Hate Speech (s. o.) abgeglichen. Es konnte eindeutig bestätigt werden, dass Datensatz II signifikant ($p < 0.0001$) mehr Hate Speech als ein zufälliges Set an Facebook-Kommentaren enthielt. (für weitere Details siehe Supplement)

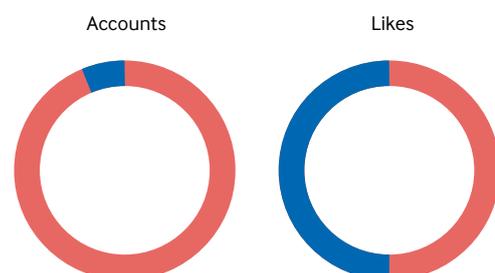
1.4 Eine kleiner Anteil User sorgt für Likes für Hateful Speech

Mithilfe von Datensatz II lässt sich die Anzahl der Hate-Speech-Accounts in den Kommentarspalten auf Facebook ungefähr quantifizieren.

Zusätzlich zu den 1.000 Accounts („wahrscheinliche Hate-Speech-Accounts“), die der Identifikation derjenigen Medien-Artikel dienten, unter denen ein hoher Anteil an Hateful-Speech-Kommentaren gemäß obiger Definition verfasst wurde, wurden in diesem Datensatz knapp 5.500 weitere hochaktive Accounts identifiziert. Diese entsprechen etwa 5% der an den mit Likes an Hateful Speech beteiligten Accounts in Datensatz II. Diese Accounts waren insgesamt für ca. 50% aller Likes für Hateful-Speech-Kommentare in den untersuchten 700 Medien-Artikeln verantwortlich. Das aktivste Prozent der Nutzer generierte gar 25% der Likes für Hateful-Speech-Kommentare. Aus dieser Abschätzung wurden Kommentare, die Likes von Mitgliedern der Gruppe #ichbinhier erhielten, ausgenommen, da die Aktionsgruppe koordiniert vorgeht und diese Daten daher das Ergebnis verfälschen könnten. Der Fokus liegt hier also ganz auf den möglichen Hateful-Speech-Kommentaren.

Die identifizierten 5% repräsentieren gerade einmal 0,02% aller monatlichen Facebook-Nutzer (für den Monat September 2017: ca. 31 Millionen Deutsche).

Eine kleine Gruppe hoch aktiver Accounts dominiert demnach eine große Anzahl der Kommentarspalten und interagiert positiv mit Hateful Speech.



2. Wer steckt hinter den Accounts?

2.1 Vorüberlegungen

Nachdem im vorangegangenen Abschnitt gezeigt werden konnte, dass besonders aktive Accounts einen Großteil der Hateful-Speech-Kommentarspalten dominieren, stellen sich nun die folgenden Fragen:

- Stellen Hateful-Speech-Accounts einen repräsentativen Querschnitt der Bevölkerung dar, oder
- sind diese Accounts einem bestimmten politischen Lager zuzuordnen?

Dazu muss zunächst eine allgemeine Definition für einen „Sympathisanten von Parteien und Organisationen“ gefunden werden. In diesem Zusammenhang kann die Überlegung zugrunde gelegt werden, dass Accounts Inhalte, die von Parteien bzw. Organisation gepostet wurden, durch positive Interaktionen (Likes) unterstützten und so ihre Übereinstimmung mit den geteilten bzw. geposteten Inhalten signalisiert haben. Für diese Accounts wird im Weiteren der Begriff „E-Sympathisant“ verwendet. Wichtig ist hier der Begriff „Account“. Eine reale Person kann mehrere Accounts besitzen, in rechtsextremen Foren wird sogar explizit die Erstellung mehrerer solcher so genannten „Sockenpuppen“ angeregt.

Um ein möglichst breit gefächertes Spektrum abzubilden, wurden die Likes für die jeweils neuesten 200 Posts von den primären Facebook-Seiten der jeweiligen Organisationen über die Facebook-Schnittstelle geladen. Der „E-Sympathisant“ einer Organisation ist hier so definiert, dass er mindestens ein Like bei diesen 200 Posts hinterlassen haben muss. Der Analyseschwerpunkt lag hierbei auf allen im Bundestag vertretenen Parteien. Dabei wurde jeweils nur die Hauptseite einer Partei ausgewertet. Seiten der Landesverbände (Ausnahme CSU) oder des Personals wurden nicht berücksichtigt, um maximale Vergleichbarkeit sicherzustellen. Zusätzlich wurde die Aktivität von E-Sympathisanten der Identitären Bewegung (IB Deutschland) und Björn Höckes untersucht, da beide in den verschlüsselten Discord-Foren „Reconquista Germanica“ und „Infokrieg“ eine große Rolle spielen.

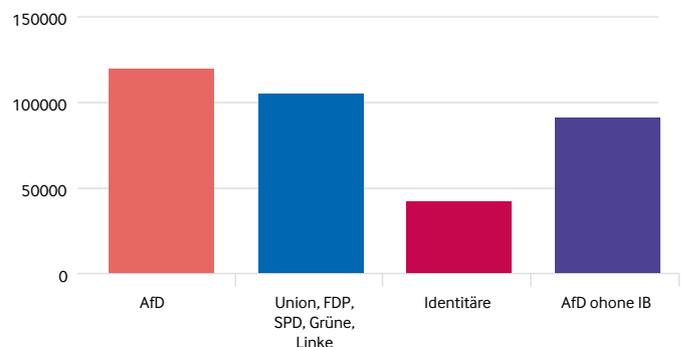
Die so identifizierten E-Sympathisanten von Parteien oder Organisationen wurden mit den bereits analysierten 700 Kommentarspalten der Medien-Artikel abgeglichen. Dadurch konnte ein Drittel der Likes für Hateful Speech in diesen 700 Kommentarspalten

E-Sympathisanten einer im Bundestag vertretenen Partei zugeordnet werden.

2.2 Die Rolle rechter Organisationen: AfD und Identitäre Bewegung

Besonders auffällig: In diesen untersuchten 700 Kommentarspalten unter den Medien-Artikeln, die Hateful-Speech-Kommentare enthalten, finden sich etwa genauso viele Likes von E-Sympathisanten der AfD wie von E-Sympathisanten aller anderen Parteien zusammen.

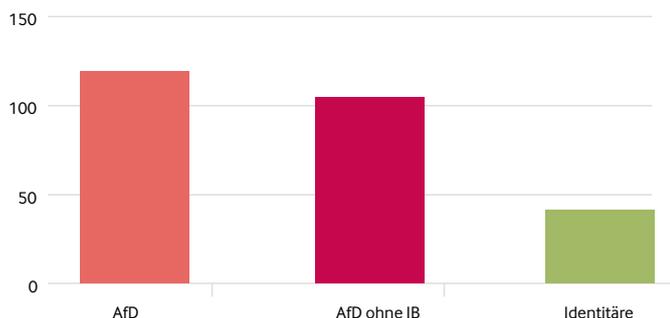
Likes von E-Sympathisanten der Parteien



Darüber hinaus sind Accounts, die als E-Sympathisanten der vom Verfassungsschutz beobachteten und rechtsextrem eingestuften Identitären Bewegung (IB) eingeordnet werden können, messbar aktiver als die Accounts aller anderen untersuchten Gruppen, wobei diese Gruppe insgesamt deutlich kleiner ist als z. B. die der E-Sympathisanten der AfD. So verteilten E-Sympathisanten der IB Likes in einer Größenordnung, die einem Drittel der Likes von AfD-E-Sympathisanten entspricht, sodass Accounts der E-Sympathisanten der Identitären Bewegung im Durchschnitt deutlich aktiver waren als AfD-E-Sympathisanten und E-Sympathisanten anderer im Bundestag verteilter Parteien.

Im arithmetischen Mittel war die IB mit jeweils 40 durch positive Interaktion als E-Sympathisanten identifizierten Accounts unter den 700 Artikeln aus Datensatz II aktiv, während der AfD 107 Accounts von E-Sympathisanten zuzuordnen waren. Zieht man AfD-E-Sympathisanten ab, die gleichzeitig E-Sympathisanten der IB sind (Schnittmenge AfD-IB), so ergeben sich im Schnitt 83 „reine“ AfD-E-Sympathisanten unter einem der 700 Artikel. Hier zeigt sich wieder die im Vergleich zur AfD extrem hohe Aktivität der IB-Accounts.

In den Medienberichten, die zu #ichbinhier-Aktionen führten, war die Beteiligung der IB noch intensiver:

arithm. Mittel d. Accounts bei #ichbinhier
Aktionen im Jan. 18

Unter solchen Medien-Artikeln waren im Beobachtungszeitraum E-Sympathisanten der Identitären Bewegung durchschnittlich mit 62 Accounts beteiligt. Bei dem sog. Shitstorm beim Kinderkanal („KiKAgate“) Anfang Januar 2018 waren es sogar mindestens 400 aktive, mit der Identitären Bewegung sympathisierende Accounts.

Die hier genannten Zahlen sind eher als untere Schranke für das Engagement der Identitären Bewegung zu sehen. Vermutlich konnten nicht alle E-Sympathisanten als solche identifiziert werden, beispielsweise, wenn sie mit Zweitaccounts aktiv waren oder eben einfach im beobachteten Zeitraum nicht mit der Hauptseite der Organisation interagiert hatten. Auffällig bleibt, dass nicht nur vereinzelt der Identitären Bewegung nahestehende Accounts auftreten. Vielmehr verfestigte sich die Erkenntnis eines konstant intensiven und koordinierten Auftretens von Accounts, die positiv mit der Facebook-Seite der Identitären Bewegung interagiert hatten.

Zusätzlich fiel ein starker Zusammenhang zwischen dem Auftreten von E-Sympathisanten der Identitären Bewegung und solchen von Björn Höcke auf, die sich unter den hochaktiven Profilen fanden. Kein anderer Politiker wird bei Reconquista Germanica und anderen Discord-Servern derart (z. B. in militaristischer Aufmachung oder in Form von Memes) glorifiziert.

Insgesamt sind diese Zahlen konsistent mit der Zusammensetzung des rechtsextremen „Reconquista Germanica“ Discord Servers.



Ser Arthur Dayne heute um 22:01 Uhr

Eine grob umrissene Bewerbung wäre hilfreich und irgendwas, womit man dich verifizieren kann: Social Media Aktivitäten, Youtube, etwaige Mitgliedschaft bei IB oder AfD (auch Fördermitglied), Bilder von "gesinnungsfreundlichem Material"...

Figure 1: Rekrutierer von Reconquista Germanica gibt AfD- und IB-Mitgliedern bewusst einen einfachen Zugang

2.3 Koordination

Die Beobachtung von hoch aktiven Accounts von E-Sympathisanten der Identitären Bewegung ist konsistent mit dem „Handbuch für Medienguerillas“,

das Martin Sellner, Co-Leiter der Identitären Bewegung Österreich, als Handreichung und Anleitung unter anderem auf seiner Webseite verlinkte. Darin ist zu lesen:

„Folge und Like die Accounts von allen Parteien, insbesondere den Grünen, bekannten Feministinnen, Regierungslakaien wie Till Schweiger oder Böhmermann und sämtlicher Propaganda-Regierungspresse, wie ARD, ZDF, Spiegel und dem Rest der Fake-News-Mischpoke.[...] Alles oben genannte ist effektiver, wenn es nicht nur von einem, sondern von einer Gruppe von Leuten gemacht wird. Sprich Dich mit Deinen Freunden ab, und wenn es zum Showdown kommt, trommle sie zusammen, um gemeinsam gegen die Lügen und das Gift zu kämpfen. Wenn Du keine Freunde hast, lege Dir mehrere Accounts an und betreibe sie parallel. Bau Dir eine Armee von Sockenpuppen auf. Der Mensch ist ein Herdentier. Er ist eher gewillt einer Gruppe von Menschen zu folgen, als einem einzelnen.“

Eine Frage, die sich daraus ergibt, ist die nach dem Grad der Koordinierung von Accounts, die in Kommentarspalten mit Hateful-Speech-Kommentaren in Erscheinung treten.

Vor diesem Hintergrund wurden die zum Ende des Betrachtungszeitraumes jeweils letzten 100 Facebook-Medien-Artikel von SPIEGEL ONLINE, FOCUS Online, WELT, tagesschau, ZDF heute, Kronen Zeitung, BILD, HuffPost Deutschland, ZEIT ONLINE und der Frankfurter Allgemeinen Zeitung geladen. Anschließend wurden alle Kommentare mit mehr als 10 Likes analysiert. Es ergab sich ein Datensatz (Datensatz III) von 6.000 Kommentaren mit insgesamt 300.000 Likes, die aus 1.000 Kommentarspalten stammen.

Für jeden Account wurde dessen Aktivität berechnet. Als Maß für Aktivität wurde die Anzahl der Kommentarspalten verwendet, in der dieser Account mit Likes aktiv war. Nun wurde für jeden Kommentar bestimmt, wie aktiv die Accounts in diesen Kommentarspalten sind, die diesen Kommentar mit „Gefällt mir“ markiert haben. Dabei zeigt sich, dass einige Kommentare von deutlich aktiveren Accounts gelikt wurden als andere.

In der Realität liken hochaktive Accounts deutlich häufiger gemeinsam als man dies probabilistisch erwarten würde. „Zufall“ kann also mit hoher Wahrscheinlichkeit ausgeschlossen werden (weitere Details siehe Supplement).

Auch die weitergehende Analyse der E-Sympathisanten in diesem Datensatz zeigt, dass viele Likes hochaktiver Accounts vor allem von Sympathisanten der AfD und insbesondere der rechtsextremen Identitären Bewegung stammten. Sympathisanten der IB und der AfD sind wie in Datensatz II signifikant aktiver als Sympathisanten anderer Parteien und treten meistens gemeinsam auf.

Die Kommentare mit den aktivsten Accounts sind zumeist Pro-Trump und pauschal gegen die Mainstream-Medien, gegen die Grünen, gegen Merkel, die EU und gegen Geflüchtete gerichtet (siehe Beispiel Sample 2).

Diese hochaktiven und in bedeutenden Teilen als E-Sympathisanten rechtsextremer Organisationen einzuordnende Accounts scheinen also gezielt bestimmte Narrative und Inhalte in den Kommentarspalten so zu verbreiten, dass sie durch die Masse ihrer Erwähnung und ihrer positiven Interaktionen als „Mehrheitsmeinung“ erscheinen.

Konsistent mit dem „Handbuch für Medienguerillas“ kann sich nur der Schluss ergeben, dass genau diese Wirkung erwünscht ist. Rechte Akteure versuchen durch hochaktive und koordinierte Accounts den Eindruck einer Mehrheitsmeinung zu suggerieren.

2.4. Beispiele

Sample 1 von Kommentaren aus **Datensatz II** (Hateful-Speech) welche von den Prüfern einstimmig als Hetze identifiziert wurden:

- „Na dann mal viel spass in algerien. Beim Ziege hüten , nebebei 10 kinder bekommen, und dein morchid holt, wie es zur tradition gehört eine algerische frau.“
- „Wie lange soll das so noch weiter gehen? Diese Kultur passt hier einfach nicht her. Und das ganze ist wahrscheinlich passiert weil er in seiner Ehre gekränkt wurde. Mein Beileid an die Familie. Man kann nur erahnen wie stark der Hass und die Wut und zugleich die Trauer ist.“
- „Diese Art der Kriminologie eines senilen SPDlers ist kriminell. Auch ihre Verbreitung ist kriminell.“
- „Und wo ist das Urteil für die Politiker die mit illegaler Grenzöffnung Menschen auf dem Gewissen haben und noch haben werden. Sollen unsere Kinder wenn die das überleben 50 Jahre oder mehr warten, worauf bis der Kadaver dieser Schweine verurteilt wird. Natürlich wenn er schuldig ist korrekt aber wie schon einer meiner Vorredner sagte warum lässt man sich so lange Zeit.“
- „Wir fühlen uns in Deutschland dermaßen verarscht. Diese ewigen Relativierungsversuche die Islamische Zuwanderung von dem dramatischen Anstieg der Kriminalität zu trennen ist eine Zumutung an die Intelligenz der Bürger. Der Islam breitet sich aus und Deutschland hat sich in nur 2 Jahren verändert wie in 60 Jahren nicht.“
- „die meisten sind illegale und keine flüchtlinge daher ist der demkansatz vom start her falsch und man kann unterentwickelten nicht höhere ziele vermitteln ,die leute kommen aus einer mittelalter gesellschaft ,als wie soll das gehen ? es ist zwingend notwendig sie in ihre heimat zurückzuführen und vor allem die entwicklungshilfe die für vielstreitigkeiten in den betroffenen ländern führt zu streichen“

Wie bereits erwähnt, finden sich in Datensatz II signifikant mehr Kommentare, die Hate Speech oder Hetze beinhalten.

Sample 2 von Kommentaren aus **Datensatz III** (Querschnitt) mit mindestens fünfmal so aktiven Accounts wie der Durchschnitt hinter den Likes (#ichbinhier bewusst ausgeblendet):

- *"Es sollten einfach Mal alle kriminellen Flüchtlinge und Migranten abgeschoben werden...der Schutz sollte sich auf wirklich bedrohte Flüchtlinge konzentrieren."*
- *„Einfach lächerlich diese Anzahl! Deutschland ist und bleibt eine Bananenrepublik! aus Liebe zu Deutschland“*
- *„Und Rechten" ist klar... Lass dich einfach zu klump schlagen dann bist kein Rechter“*
- *GEZ abschaffen. Ich schaue und höre mir das nicht an. Sehr oft sehr einseitig recherchiert, Stimmungsmache, völlig überzogene Gehälter für Redakteure (Kleber etc.) usw."*
- *"ich traue den öffentlichen Sendern auch nicht mehr haben in den Flüchtlings fragen uns zu oft belogen und für dumm verkauft“*
- *bei den ÖR gibt es keine neutrale Berichterstattung und das wäre der eigentliche Auftrag, wofür man auch gerne bezahlen würde“*
- *"Neue Gesichter [Politiker], altes leid!!!"*
- *"Wäre doch super! Dann arbeiten wir alle wieder für 6€/std und unsere Politiker und Banker und Wirtschaftsführer sind wieder glücklich!"*
- *"Er [Macron] verliert in Frankreich immer mehr an Zustimmung , in moment wohl ca. 10%. So viel planen sollte er wohl nicht“*

Dieses Sample wurde aus Datensatz III nach keinem anderen Kriterium als der hohen Aktivität (Aktivität hier: Zahl der Kommentarspalten, in denen ein Account mit Likes aktiv war) der Accounts hinter den Likes ausgewählt. Hier handelt es sich nach Ansicht der Prüfer nicht um Hate Speech, aber offensichtlich in der Tendenz um Kommentare von Nutzern, die dem rechtspopulistischen Spektrum zuzuordnen sind. Damit wird bestätigt, was sich Datensatz II schon zeigte: Rechte Accounts dominieren durch ihre extrem hohe Aktivität die Kommentarspalten und pushen ihre Narrative.

Sample 3: Zum Vergleich einige zufällig ausgewählte Kommentare mit Usern sehr geringer Aktivität aus **Datensatz III:**

- *„Gemüse und Obst kann man waschen/ schälen! Notfalls würd ich mir das auch, wie an Wurst-/ Käsetheke, von Angestellten des Marktes geben lassen (1 Job geschaffen + Plastik vermieden). Wenn man will, geht alles!“*
- *"nicht beim Aufstieg, sondern beim Abstieg...und nicht aus 7400 Metern Höheso viele Fehler in einer Meldung. Peinlich..... Warum man Urubko und Bielecki mit keinem Wort erwähnt, bleibt mir schleierhaft....."*
- *"Ich teste gar nichts. Mein Baby darf so zur Welt kommen, wie es ist und es wird einfach nur geliebt. Fertig."*
- *"Es heißt ja nicht gleich, dass sich jeder Fotograf nennen möchte, nur weil er bearbeitete Bilder (oder auch unbearbeitete) bei Facebook oder Instagram hochlädt...n\n\nIch sehe jedenfalls oft tolle Bilder von Hobbyfotografen und sehr langweilige oder auch schlechte von „professionellen“."*
- *„Wo die Liebe hinfällt! Gott sei Dank gibt es heute nicht die Hürden, die es zu meiner Zeit gab. Da war es schon schwierig, einen Partner in aus einer anderen Konfession zu heiraten, geschweige denn, eine andere Nationalität! Ich bin froh, dass sich hier viel geändert hat!“*
- *"Der Schrift nach zu urteilen, eine ältere Person. Evtl. war sie mit einem temperamentvollen Welpen heillos überfordert. Ausserdem anzunehmen, dass sie finanziell nicht auf Rosen gebettet ist."*

IV. Hintergründe der Hate-Speech-Koordinierung

1.1 Hassverbreitung: Koordination durch organisierte Trollfabriken

Auf Basis unserer 6-monatigen ethnografischen Recherchen in rechtsextremen Foren und Chatgruppen in Deutschland analysieren wir die Koordination von Hass-, Einschüchterungs- und Manipulationskampagnen im Netz.

Die US-Wahlen zeigten erstmals das Ausmaß des Einflusses, den koordinierte Kampagnen von Trollfabriken auf unsere politischen und gesellschaftlichen Strukturen haben können. Der Wahlsieg Donald Trumps, den die amerikanische Alt-Right im Netz als Resultat ihrer Online-Trollkampagnen feierte, motivierte Rechtsextremisten auf der ganzen Welt zu versuchen, den politischen Online-Diskurs in ihrem Land zugunsten rechtspopulistischer Parteien zu beeinflussen, indem sie in geschlossenen Chatgruppen gezielt Kampagnen zur Verbreitung von Hass gegen Minderheiten, das politische Establishment und die Mainstream-Medien planten. Allein im Jahr 2017 konnten zahlreiche koordinierte Versuche von rechtsextremen Gruppen im Netz beobachtet werden, die Wahlen in Frankreich, den Niederlanden, Deutschland und Österreich zu beeinflussen.⁵

Im Sommer 2017 schlossen sich auch erstmals deutsche rechtsextreme Gruppierungen mit dem Ziel zusammen, den Ausgang der Bundestagswahl zu Gunsten der AfD zu beeinflussen. Die Diskussionen dazu fanden zunächst auf 4Chan und 8Chan statt, wurden jedoch schnell in verschlüsselte Nachrichtenapplikationen wie Telegram und Discord verlagert. Eine zentrale Rolle spielten dabei die Kanäle #Infokrieg und Reconquista Germanica in der verschlüsselten Gaming-Applikation Discord. Die Gruppe Reconquista Germanica hatte am ersten Tag nach ihrer Gründung bereits knapp 500 Mitglieder, am Tag vor der Bundestagswahl etwa 5.000 und im Februar 2018 vor ihrer ersten Schließung durch den Betreiber mehr als 7.000. Seit Ende April hat die Gruppe, die sich jetzt „Revivica Germanica“ nennt, deutlich über 10.000 Mitglieder.

Die Autoren haben mit großer Überraschung zur Kenntnis genommen, dass die Bundesregierung in ihrer Antwort auf eine Kleine Anfrage der Fraktion DIE LINKE erklärt hat, dass sie „Reconquista Germanica“ bislang nicht als rechtsextrem einstuft. Wie aus den Screenshots in dieser Studie hervorgeht, sind nationalistische, rassistische, xenophobe und zum Teil sogar nationalsozialistische Inhalte und Symbole derart bestimmend in der Gruppe, dass uns unklar ist, wie die Bundesregierung zu dieser Einschätzung kommt.

Bereits der Gruppenname enthält die muslimfeindliche Botschaft, dass Deutschland, wie einst die iberische Halbinsel, von islamischen Invasoren gewaltsam „zurückerober“ werden müsse.

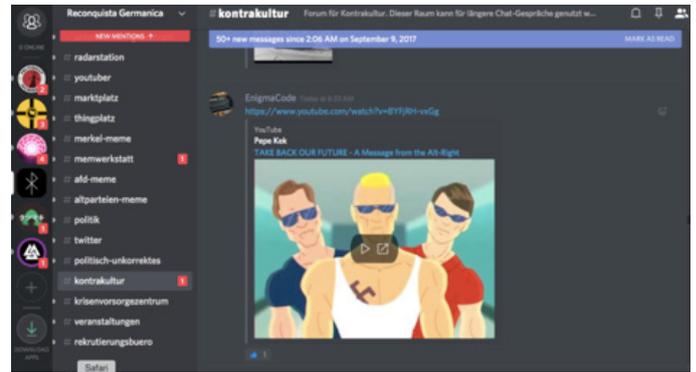


Figure 2: Nationalsozialistische Symbolik im Discord-Kanal Reconquista Germanica

Die Organisationsstrukturen sind streng hierarchisch mit unterschiedlichen strategischen und regionalen Ebenen. Der selbsternannte Oberbefehlshaber Nikolai Alexander erteilt Befehle unter Verwendung militärischer Sprache mit Bezug auf die Wehrmacht (s. unten). Neben den Discord-Kanälen werden auch die hauptsächlich unter Rechtsextremen populäre Twitter-Alternative Gab.ai und der russische Facebook-Ersatz VK.com zur Koordination von und Mobilisierung für Trolling-Kampagnen verwendet.



Die Autoren haben mit großer Überraschung zur Kenntnis genommen, dass die Bundesregierung „Reconquista Germanica“ bislang nicht als rechtsextrem einstuft.



Begriff	Verwendung bei RG	Bedeutung
Fall Blau	Wiedererrichtung des Servers nachdem dieser gelöscht wurde	Erster Teil der Sommeroffensive der Wehrmacht im Jahr 1942 während des Deutsch-Sowjetischen Krieges
Paladin	Zweithöchster Rang in RG	Spitzname von Hermann Göring, „zweiter Mann hinter Hitler“
Heeresgruppe Süd	Regionalgruppe von RG	Heeresgruppe der Wehrmacht für den Polenfeldzug 1939
Bewährungsbattalion	Niedrigster Rang von RG	Aus verurteilten Soldaten bestehende Wehrmachtseinheiten im Zweiten Weltkrieg, die zur „Bewährung“ an die Front geschickt wurden



durch Trolling zu einer verbalen Überreaktionen zu provozieren, um sie öffentlich bloßzustellen. In „Clear and Hold“-Missionen wird hingegen darauf abgezielt, gegnerische Hashtags (#refugeeswelcome) mit den eigenen Inhalten zu infiltrieren, um User, die solche Hashtags affirmierend verwenden, aus der Timeline zu drängen. Und unter „Search and destroy“-Operationen wird schlicht ein bestimmter, auf Discord vorher vereinbarter Post massiv mit Memes und Kommentaren „zugesпамmt“.

In gezielten „Raids“ auf Facebook, Twitter und YouTube werden Uhrzeiten und Hashtags verabredet, zu denen Mitglieder der rechten Trollgruppen „Altparteien“-Politiker, linke Aktivisten oder Mainstream-Medien angreifen, den Kommentarbereich von Videos mit provokativen Nachrichten „zusammen“ oder besonders viele „Dislikes“ verteilen. Hauptthemen sind dabei Masseneinwanderung, eine vermeintliche Islamisierung, Identität, Freiheit und Tradition.

Diese Operationen sind häufig äußerst erfolgreich. So schafften es die rechtsextremen Aktivisten beispielsweise, 15 Tage vor der Bundestagswahl sieben von ihren Hashtags (u.a. #TraudichDeutschland, #nichtmeinekanzlerin, #merkelmussweg, #reconquista) in den Top-20 Hashtags zu platzieren.



Figure 3: Rechtsextreme Trollgruppen: #Infokrieg (oben) und Reconquista Germanica (unten)

Im sogenannten „Krisenvorsorgezentrum“ auf „Reconquista Germanica“ wurden gewalt- und waffenverherrlichende Inhalte sowie Hinweise dazu geteilt, wie sich die User selber Elektroschocker bauen können.

Dabei setzen die rechtsextremen Trolle auf unterschiedliche Techniken und Taktiken zur Manipulierung der Algorithmen der sozialen Medien, der koordinierten Verbreitung von Hass im Netz und der gezielten Beeinflussung des Online-Diskurses.

In sogenannten „Sniper“-Missionen versuchen die rechtsextremen Aktivisten, bekannte Persönlichkeiten

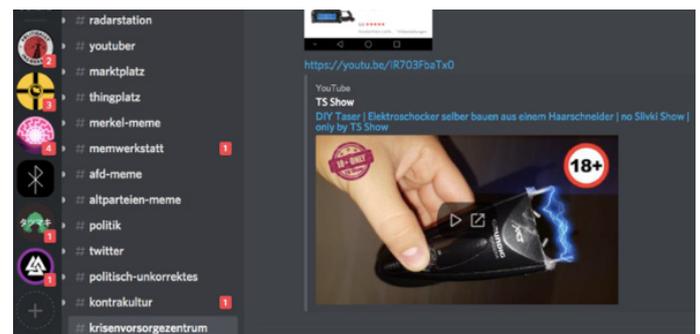
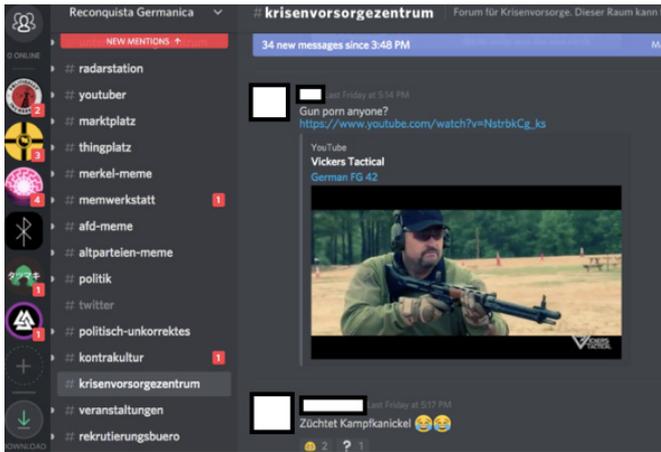


Figure 4: Anleitung zum Bauen eines Elektroschockers



und waffenverherrlichende Inhalte im "Krisenvorsorgezentrum" von "Reconquista Germanica"



Figure 5: Der Gebrauch des von „Reconquista Germanica“ gepushten Hashtag #traudichdeutschland nimmt in den zwei Wochen vor der Wahl deutlich zu.

Die Administratoren von #Infokrieg weisen ihre Follower an, sich mehrere Social Media-Profile (idealerweise außerhalb ihrer eigenen Filterblase) zuzulegen und mit dem Programm „Ghostbuster“ gleichzeitig über mehrere Accounts zu posten, um unverdächtig zu bleiben. Zudem sollten sie über nicht-politische Themen (Autos, Fußball) tweeten, um die Zahl der

eigenen Follower zu erhöhen. Daneben legen sich einige der rechtsextremen Trolls auch Parodie-Accounts zu, um die politischen Gegner zu trollen und Verwirrung zu stiften. Dies bedeutet auch, dass die Untersuchung in Kapitel III 2.2. vermutlich die Zahl dieser Accounts unterschätzt hat.



Figure 6: Der Gebrauch des von „Reconquista Germanica“ gepushten Hashtag #traudichdeutschland nimmt in den zwei Wochen vor der Wahl deutlich zu.



Figure 7: Anweisungen zur Verwendung mehrerer Accounts über das „Ghostbrowser“-Programm und die Wichtigkeit von Accounts außerhalb der rechtsextremen Filterblase (in der dann beispielsweise über Fußball oder Autos getweetet wird, um Follower zu gewinnen).

In den Discord-Gruppen „Reconquista Germanica“ und „#Infokrieg“ werden zudem provokante politische Memes entwickelt und in Meme-Datenbanken gesammelt, um diese dann auf sozialen Medien wie Facebook und Twitter zu verbreiten. Um an politisch sensibles Material zu gelangen, werden innerhalb von „Reconquista Germanica“ auch Anleitungen verteilt, wie IP-Adressen versteckt werden können, wie man automatisierte Social-Media-Accounts erstellt und wie sich verhindern lässt, dass diese von den Plattformbetreibern heruntergenommen werden.



1.2. Kampagnen

Trotz des formellen parteiinternen Unvereinbarkeitsbeschlusses, der eine Kooperation zwischen AfD-Mitgliedern und vom Verfassungsschutz beobachteten extremistischen Gruppen wie der Identitären Bewegung verhindern soll, gibt es immer wieder Berichte, die auf eine Zusammenarbeit hinweisen. Auch in unserer Analyse der von den rechtsextremen Trollfabriken organisierten Kampagnen stoßen wir immer wieder auf AfD-Accounts, welche die Hashtags dieser Kampagnen in ihren Tweets verwenden. Durch die hohe Anzahl an Followern, welche diese AfD-Accounts auf sozialen Medien haben, fungiert die AfD damit für die rechtsextreme Identitäre Bewegung als ein wichtiges Vehikel, um ihre Inhalte einem weiteren Publikum zugänglich zu machen (Mainstreaming).

Dabei handelt es sich zum Teil um durchaus prominente Mitglieder der Partei, darunter mehrere Bundestagsabgeordnete, welche die Inhalte der rechtsextremen Identitären aufgreifen und in den sozialen Medien verbreiten. So benutzte neben den Mitgliedern des Bundestags Udo Hemmelgarn und Martin Reichardt, welche dem rechten und völkischen Lager der Partei zugeordnet werden, auch die Bundesvorsitzende Alice Weidel den Hashtag #KiKAGate und stellte die Frage, ob KiKA „unsere Mädchen auf Ehen mit islamischen Einwanderern“ vorbereiten würde.

Figure 8: Schaubild der „militärischen“ Rangordnung innerhalb des Discord-Kanals „Reconquista Germanica“.



Figure 9: Online Meme-Datenbank, in der selbst entwickelte Memes in der Gruppe geteilt und gespeichert werden, um bei geeigneten Kampagnen eingesetzt zu werden

Udo Hemmelgarn, MdB  **@UdoHemmelgarn** Following

"#AufbruchInsUngewisse" war schäbig. 😞
Den berechtigten anti-ARD-Protest entsetzter Beitragszahler dann auch noch zu diffamieren, hat Folgen. 🗣️ Nach #KIKAGate + #heuteshow-Skandal ein neuer Eklat. 📺 #Programmbeschwerde #Rundfunkrat #AfD #Maischberger #Lanz

Beatrix von Storch  **@Beatrix_vStorch**
Ich dachte dümmer geht nimmer. #AufbruchInsUngewisse ist SOO unterirdisch. Und jetzt müsst ihr auch noch die Protestwelle als „gezielt beeinflusst“ verunglimpfen? Wacht ihr irgendwann nochmal auf? #Absurdistan #NoGEZ #AfD
twitter.com/daserste/statu...

Translate from German
12:56 AM - 15 Feb 2018

87 Retweets 196 Likes 

8 87 196

Alternative NRW  **@AlternativeNRW** Folge ich

Unerträglich, dass eine mutige Mutter, die Mädchen schützen will, als "Nazi" diffamiert wird! 😞 Damit laden die, die eine solch couragierte Frau verunglimpfen, schwere Schuld auf sich! R.I.P., Mia Valentin! #Kandel #kandelistueberall #120dB #db120 #Schulz

Junge Freiheit  **@Junge_Freiheit**
„Die nennen mich jetzt im Internet Nazi. Das ist doch verrückt. Nichts liegt mir ferner.“ Eine Mutter, die an der Frauendemo in #Kandel das Mikrofon ergriffen hat, im Gespräch mit der JF.
jungefreiheit.de/kultur/gesells...

17:19 - 9. Feb. 2018

246 Retweets 502 „Gefällt mir“-Angaben 

Junge Alternative Hessen  **@ja_hessen** Folgen

Die Junge Alternative Hessen steht hinter allen mutigen Bürgern, die ihren Protest heute lautstark auf Hamburgs Straßen bringen. Seid laut und seid stolz.

#AfD #JungeAlternative #MerkelMussWeg #120db #Hamburg #hh1202

19:49 - 12. Feb. 2018

42 Retweets 136 „Gefällt mir“-Angaben 

3 42 136

Während es in manchen Fällen möglich ist, dass aufgrund inhaltlicher Überschneidungen Kampagnen der Identitären Bewegung unbeabsichtigt amplifiziert werden, scheint Absicht in anderen Fällen deutlich wahrscheinlicher. So verwenden die „Junge Alternative Hessen“, der immer wieder eine Nähe zur Identitären Bewegung vorgeworfen wird, der AfD-Rechtsaußen Martin Reichardt und der Account der „Alternative NRW“ sogar den eindeutig der Identitären Bewegung zuzuordnenden Hashtag #120db.⁷

Dr. Alice Weidel  **@Alice_Weidel** Following

"Minderjähriger"  Asylbewerber (schnell alternd & bärtig) über seine 16-jährige  Freundin:
»Ich mag nicht, dass andere Typen mit ihr reden. Sie gehört mir.«
Bereitet #KiKa unsere Mädchen auf Ehen mit islamischen Einwanderern vor?
#Malvina #KIKAGate

Translate from German



Beziehung zu deutschem Mädchen: Wirbel um Flüchtlings-Doku im Kinderka...
Aufregung um den Kinderkanal. In einer Sendung erzählt eine 16-jährige Deutsche, wie ihr syrischer Freund ihr Leben veränderte. Nun rechtfertigt sich KiKa und beda...

„ Durch die hohe Anzahl an Followern, welche diese AfD-Accounts auf sozialen Medien haben, fungiert die AfD damit für die rechtsextreme Identitäre Bewegung als ein wichtiges Vehikel, um ihre Inhalte einem weiteren Publikum zugänglich zu machen. „

Um gezielt eine einzelne Kampagne zu analysieren und einen noch weitergehenden und direkten Vergleich der verschiedenen Account-Aktivitäten und -Ausrichtungen zu ermöglichen, wurden Kommentarspalten aus Datensatz III zum Holocaust-Gedenktag analysiert.

Hier zeigte sich, dass die hochaktiven Accounts das Augenmerk vor allem auf Antisemitismus unter Geflüchteten lenken wollen, anstatt des Holocausts zu gedenken. Dieselben Accounts unterstützen auch ein „Ende des Schuldkults“. Außerdem werden gezielt und koordiniert Äußerungen der Holocaust-Überlebenden Hanni Levy kritisiert, die zu einem respektvollen Umgang mit Flüchtlingen gemahnt hatte.

Ähnliches ließ sich im verschlüsselten Discord Forum „Reconquista Germanica“ beobachten.



Figure 10: Die Likes für diesen Kommentar kamen von Accounts, die 60% aktiver waren als der Durchschnitt. Mindestens 15% stammen von E-Sympathisanten der IB. Eine Aussage über die Integrität des Kommentators ist nicht möglich, es geht nur um die Likes.



Figure 11: Dieser Kommentar steht bei der Facebook Seite der WELT unter einem Sharepic mit der Aufschrift „Wir dürfen niemals vergessen“, welches die Züge vor Auschwitz zeigt. Er wurde von mehreren E-Sympathisanten der IB gelickt.

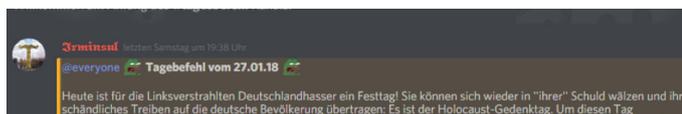


Figure 12: Screenshot aus dem verschlüsselten Discord Forum „Reconquista Germanica“



Figure 13: Dieser Post des ZDF über die Holocaust-Überlebende Hanni Levi weist in seiner Kommentarspalte Like-Aktivität von 104 E-Sympathisanten der rechtsextremen Identitären Bewegung auf. Es waren außerdem 66 der aktivsten Accounts zugegen, die sich am Shitstorm #KiKagate beteiligt hatten. Alle Accounts interagierten positiv mit Kommentaren, die Hanni Levi kritisierten.



Konzertierte rechtsextreme Aktionen müssen frühzeitig erkannt werden, so dass die Zivilgesellschaft Gegenmaßnahmen ergreifen kann.



Im Gegensatz zu den hochaktiven Accounts, die eben beschriebene und von „Reconquista Germanica“ verbreitete gemeinsame Narrative in den Mainstream übertragen wollen, wird mit Kommentaren, die den Holocaust-Gedenktag als solchen würdigen, deutlich weniger interagiert.

Insbesondere in der Analyse zum Holocaust-Gedenktag fällt auf, dass einzelne Accounts nicht nur deutlich aktiver sind als andere, sondern dass auch sehr viele sehr aktive Accounts gleichzeitig auftreten, die auch noch in vielen Fällen der rechtsextremen Identitären Bewegung angehören. Das lässt den Schluss zu, dass es sich dabei in vielen Fällen um ein koordiniertes Auftreten dieser sehr aktiven Accounts handelte.

Um zu bestätigen, dass diese Beobachtung kein Einzelfall, sondern ein gemeinsames strategisches Vorgehen der hochaktiven Accounts beschreibt, wurde der Shitstorm gegen den Kinder-Kanal von ARD und ZDF („KiKAgate“) ebenfalls ausführlich untersucht. Bei diesem Shitstorm waren Accounts der Identitären Bewegung besonders aktiv und auch im rechtsextremen Forum „Reconquista Germanica“ wurde zur Teilnahme daran aufgerufen. Die 1.000 aktivsten Accounts, die am Shitstorm gegen den Kinderkanal aktiv an Likes für Hateful Speech beteiligt waren, waren bei Kommentaren zum Holocaust-Gedenktag mit tausenden Interaktionen zu finden.

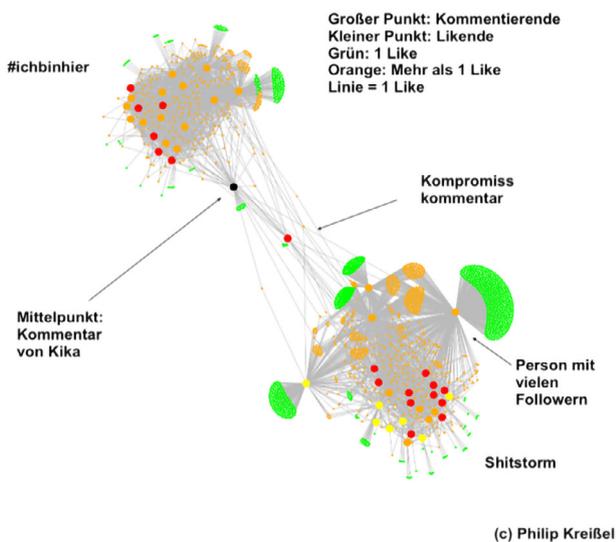


Figure 14: Visualisierung des KiKA-Shitstorms – Die am Shitstorm beteiligten Accounts „ liken“ sich gegenseitig „hoch“.

1.3. Die Rolle von Bots

Es kommt auf Twitter zudem zu einer Verstärkung der geplanten Hasskampagnen durch russischsprachige Bots, wenngleich es so scheint, als ob dies nicht mit den rechtsextremen Online Trolls abgestimmt ist. In der verschlüsselten Gruppe #Infokrieg beispielsweise wunderten sich mehrere User darüber, dass bestimmte Tweets viral gegangen waren, und stellten überrascht fest, dass viele der Retweets von russischsprachigen Bots ausgingen.



Auf Facebook gestaltet sich eine Analyse von Bots im Allgemeinen schwieriger: Es wurden von verschiedenen Autoren Kriterien zur Identifizierung von automatisiertem Verhalten vorgeschlagen. Dazu zählen: sehr hohe Aktivität, ganztägige Aktivität etc. Allerdings treffen diese Kriterien genauso auf koordiniertes Verhalten von rein menschlich gesteuerten koordinierten Aktionen zu. Daher ist eine eindeutige Identifizierung und Abgrenzung von automatisiertem Verhalten mithilfe der vorliegenden Daten technisch nicht möglich. Prinzipiell stellen die hier vorgestellten Erkenntnisse aber ein „Proof of Concept“ dar, weil gezeigt werden konnte, dass nur wenige Accounts nötig sind, um das Bild in den Kommentarspalten zu

ändern. Es ist leicht möglich, dass nicht-staatliche und staatliche Akteure Fake Accounts erstellen, um damit die öffentlichen Debatten in den Kommentarspalten bei bestimmten Themen vollständig zu kontrollieren. Ein Beweis des Vorliegens staatlicher Kampagnen ist mit den vorliegenden Daten nicht möglich.

1.4. Die Rolle der deutschen Medien

In den 1,6 Mio. analysierten rechtsextremen Posts im vergangenen Jahr wurden Links zu den folgenden Medien-Webpages am häufigsten geteilt:

Meistgeteilte Links zu Medien-Webpages	Meistgeteilte Medien-Artikel
Philosophia-Perennis.com Refcrime.info Welt.de Tichyseinblick.de JungeFreiheit.de Bild.de Focus.de	1. Bild: „Hier posiert Mias Killer in Smoking“, 28. Dezember 2018 2. Junge Freiheit „Die Zensur macht Twitter kaputt“, 31. Mai 2017 3. Welt, „Vorbestrafter Sexualstraftäter soll 13-jährige vergewaltigt haben“, 2. Februar 2018

In Zusammenhang mit den Hasskampagnen-Hashtags #KiKagate, #Kandelistueberall und #120dB wurden Links zu folgenden Medien-Webseiten am häufigsten geteilt:

Meistgeteilte Links zu Medien-Webpages	Meistgeteilte Medien-Artikel
Philosophia-Perennis Welt.de Bild.de Compact-online.de Focus.de pi-news.net Krone.at	1. Welt, Wirbel um Flüchtlings-Doku im Kinderkanal KiKA“, 9. Januar 2018 2. Bild, „Wegen eines Mädchens stach der Jugendliche zu“, 18. Januar 2018 3. Bild: „Hier posiert Mias Killer in Smoking“, 28. Dezember 2018

Im Gegensatz zu den AfD-Accounts und den russischen Medien wie RT und Sputnik griffen deutsche Mainstream-Medien wie die BILD zwar die Themen von #Kandel und #KiKA auf, verwiesen allerdings nicht auf die Kampagnen der Identitären Bewegung.

1.5 Die Rolle russischer Medien

Abgesehen von AfD-Accounts spielen auch russische Medien wie RT und Sputnik eine wichtige Rolle als Multiplikatoren für die Hasskampagnen der Identitären Bewegung. So griffen sowohl RT („120 Dezibel – Aktivistinnen starten Kampagne gegen ‚importierte‘ Gewalt an Frauen in Europa“) als auch Sputnik („Frauen gegen ‚Merkels Gäste‘ – Rechtsruck oder Selbstverteidigung?“) Anfang Februar die Kampagne der Identitären Bewegung #120db in ihren Berichten auf.



V. Limitationen

1.1. Klassifizierung von Hate-Speech-Accounts

Die Einordnung der 1.000 Hate-Speech-Accounts (vgl. Kapitel 2.3) erfolgte auf der Grundlage ihrer besonders häufigen positiven Interaktion mit Hateful Speech. Selbst wenn einzelne Interaktionen fälschlicherweise als Interaktionen mit Hate Speech klassifiziert wurden, so ist die Wahrscheinlichkeit einer falschen Einordnung als Hate-Speech-Account relativ gering, weil jeder dieser Accounts mindestens in vier Kommentarspalten positiv mit vermuteter Hateful Speech interagiert hat

Es ist weiter zu erwähnen, dass Artikel mit Hateful Speech ohne die Beteiligung der 1.000 Hate-Speech-Accounts möglicherweise nicht erkannt worden wären. Ein Selektions-Bias wäre die Folge. Allerdings fand sich, gleichsam in der Gegenprobe, auch im sekundären Auswahlzeitraum (Januar 2018) in allen Medien-Artikeln, in denen #ichbinhier aktiv wurde, eine hohe Beteiligung der identifizierten Hateful-Speech-Accounts, währenddessen außerhalb der Aktivitätsschwerpunkte dieser 1.000 Hate-Speech-Accounts kein nennenswertes Vorkommen von Hateful Speech zu verzeichnen war.

Umgekehrt wurden aber hunderte Kommentarspalten mit Hateful Speech und Beteiligung der Hateful-Speech-Accounts erkannt, in denen #ichbinhier nicht aktiv wurde, weil zum Beispiel bereits Gegenrede vorhanden war oder insgesamt zu wenig Interaktion vorlag. Für die Identifikation von Hateful Speech war diese Einschränkung also kein Hindernis und führte allenfalls dazu, dass die Ergebnisse das wahre Ausmaß der Like-Zahlen bei wenigen Accounts und die Beteiligung rechtsextremer Accounts unterschätzen.

1.2. E-Sympathisant

Die Definition des E-Sympathisanten mag eine gewisse Unschärfe enthalten, weil ein Like selbstverständlich keine zweifelsfreie Zuordnung eines Accounts zu einer Partei zulässt. Dennoch ist der immense Unterschied zwischen der Repräsentanz von E-Sympathisanten der AfD gegenüber allen anderen Parteien so signifikant, dass dies keinesfalls allein der Methodik geschuldet sein kann. Bei der Identitären Bewegung ist eher davon auszugehen, dass deren Anzahl unterschätzt wurde, da sie oft mit Fake Accounts arbeitet, die sich nicht eindeutig der IB zuordnen lassen.

1.3. Unterschiede in der Vorgehensweise von #ichbinhier und Reconquista Germanica

Die Facebook-Gruppe #ichbinhier wird koordiniert in den Kommentarspalten im Sinne einer Gegenrede aktiv, wenn Hateful-Speech-Kommentare identifiziert wurden. Es handelt sich also um eine Reaktion, #ichbinhier wird nicht aktiv, um a priori das Meinungsbild in den Kommentarspalten zu ändern, wie z.B. Reconquista Germanica es tut.

Außerdem macht #ichbinhier **seine Koordination durch einen Hashtag kenntlich**, um Transparenz zu schaffen. Dies könnte #ichbinhier als Nachteil ausgelegt werden, #ichbinhier hält es aber für moralisch geboten, seine koordinierte Vorgehensweise offenzulegen. Aktionen aus rechtsextremen Gruppen versuchen bewusst, ihren Charakter, ihre Herkunft und ihr Vorgehen zu verschleiern.

Interessanterweise erhalten aber manche Kommentare mit Hateful Speech zum Teil Likes von im Durchschnitt deutlich aktiveren Accounts als die auch koordinierten #ichbinhier-Kommentare mit vielen positiven Interaktionen. Dafür gibt es keine andere plausible Erklärung als einen extrem hohen Grad an Koordination.

Eine Schlussfolgerung aus diesen Daten kann nur sein, dass sich die Aktivitäten von #ichbinhier nicht gegen sachliche konservative Meinungen, sondern gegen koordinierte rechtsextreme Aktionen auf den Facebook-Medienseiten richten. Durch koordinierte rechtsextreme Aktionen soll ein Meinungsbild suggeriert werden, das so in der Realität nicht existiert.



Durch koordinierte rechtsextreme Aktionen soll ein Meinungsbild suggeriert werden, das so in der Realität nicht existiert.



Schlussfolgerungen

Hate Speech unter Medienartikeln der großen deutschsprachigen Newsseiten auf Facebook wird von einer – gemessen an der Anzahl aller Nutzer – kleinen Gruppe von Accounts produziert, „gepusht“ und verbreitet. Die Verbreitung geschieht oft inhaltlich und zeitlich koordiniert. Ziel dieser Koordination ist es, politische Inhalte in der Mitte der Gesellschaft zu platzieren und den Eindruck zu erwecken, die kommunizierten Einstellungen zu diesen Inhalten entsprächen der Einstellung des Durchschnittsnutzers. Auf diese Weise soll eine verzerrte Darstellung des Meinungsbildes erreicht werden, die wiederum auf die tatsächlichen Meinungen der User zurückwirkt.

Die Koordination gemeinsamer Aktionen erfolgt beispielsweise auf „Discord“ und in geschlossenen Gruppen auf Facebook. Die geschlossenen Gruppen haben eine inhaltliche und organisatorische Nähe zu Gruppierungen und Parteien, die am äußersten rechten Rand des politischen Spektrums zu verorten sind. Die einzelnen User sympathisieren mit den Inhalten diesen Organisationen. In vielen Fällen sind die Aktionen so gut organisiert, dass sie die Durchführung von Hasskampagnen und „Shitstorms“ begünstigen.

Die Analysen belegen eindeutig ein strategisches Kalkül rechtsextremer Organisationen zur gezielten Meinungs- und Stimmungsmache in den sozialen Medien. Insbesondere Aktivisten aus dem rechtsextremen Spektrum zeigen sich auffällig bestrebt, extremistische Inhalte über die sozialen Netzwerke wie Facebook und Twitter in die Mitte der Gesellschaft zu überführen und über die schiere Masse an Kommentaren und Likes eine Meinungshoheit suggerieren. Für diese Organisationen stellen die sozialen Medien ein ideales Umfeld da, um mittels verdeckter Aktionen ihre Randideologie in die Mitte der Gesellschaft zu tragen. Vertreter des übrigen Meinungsspektrums, verfassungstreue Bürger und Minderheiten werden beleidigt, zermürbt und aus den Kommentarspalten vertrieben. Politiker und Medien greifen teilweise Inhalte aus Kampagnen extremistischer Organisationen auf und machen diese damit im gesellschaftlichen Diskurs „salonfähig“.

Aus der vermeintlichen quantitativen Überlegenheit der extremistischen Nutzer resultieren zwei Effekte:

Fehlinformationen, extremistische Meinungen, aber auch herabwürdigende oder beleidigende Sprache können in Verbindung mit sehr hohen und schnell wachsenden Like-Zahlen durchaus als gesellschaftlich vorherrschende Meinung wahrgenommen werden.

Die vorliegende Analyse zeigt in diesem Zusammenhang, dass 5% der bei Hateful-Speech aktiven Accounts 50% der Likes für herabwürdigende Kommentare generieren. Somit werden in der Online-Welt nicht nur Mehrheitsmeinungen suggeriert, sondern auch ein gesellschaftlicher Konsens dahingehend vorgetäuscht, dass Menschen anderer Herkunft, anderen Glaubens oder anderer Hautfarbe beleidigt und abgewertet werden können, ohne dass mit Sanktionen zu rechnen ist. Hierdurch wird die Meinungsbildung in der Bevölkerung manipuliert. Welchen Schaden die oben erläuterten Kampagnen beispielsweise mit Blick auf die gesellschaftlichen Bindungskräfte oder die politische Kultur langfristig anrichten wird, lässt sich bislang nicht prognostizieren.

Die Beeinträchtigung oder gar das Verdrängen des pluralistischen Diskurses ist eine weitere Folge der koordinierten Aktionen. Mehrfach-Accounts bekämpfen gezielt Gegenrede und tragen in vermeintlich numerischer Überlegenheit dazu bei, Meinungsäußerungen und das Engagement für sachlichen Diskurs zu unterdrücken. Damit geraten – bislang weitgehend unbeachtet – elementare Prinzipien des Grundgesetzes, wie z.B. der Schutz von Minderheiten sowie die Religions- und Meinungsfreiheit, zunehmend unter Druck.



Die vorliegende Analyse zeigt in diesem Zusammenhang, dass 5% der bei Hateful-Speech aktiven Accounts 50% der Likes für herabwürdigende Kommentare generieren.



Empfehlungen: Koordinierte Zivilcourage gegen koordinierten Hass

Die vorliegende Veröffentlichung richtet sich an die Verantwortlichen in Medien, Institutionen und politischen Parteien sowie Bürger, denen Werte wie Demokratie und Pluralismus am Herzen liegen. Sie soll für die Funktions- und Wirkungsweise koordinierten Hasses sowie die sich daraus ergebenden Folgen sensibilisieren.

Um der hieraus resultierenden Destabilisierung entgegenzuwirken und unsere Verfassung und die demokratischen Grundwerte zu schützen, sind gesamtgesellschaftliche Gegenmaßnahmen erforderlich. Aus Sicht der Autoren leiten sich aus den vorstehenden Erkenntnissen die nachfolgenden Empfehlungen ab:

Das verstärkte Auftreten der rechtsextremen Identitären Bewegung, bei der der Verfassungsschutz „Anhaltspunkte für Bestrebungen gegen die freiheitliche Grundordnung“ sieht, zeigt, dass eine verstärkte Identifizierung, Beobachtung und Analyse solcher digital organisierter illiberaler Gruppierungen im Netz unbedingt notwendig ist.

Sowohl die Medien als auch die zivilgesellschaftlichen Institutionen sind gehalten, über das konzertierte Setzen von Meinungs- und Deutungshoheiten gründlich und umfassend aufzuklären und die Bevölkerung weitreichend zu informieren und zu sensibilisieren. Auch in anderen Ländern der EU sind ähnliche Analysen und Initiativen erforderlich, um den gesellschaftlichen Frieden in Europa dauerhaft zu sichern. So existieren in anderen europäischen Ländern ebenfalls an #ichbinhier angelehnte Counter-Speech-Initiativen, die Unterstützung benötigen. Die Problematik beschränkt sich nicht auf Deutschland, sondern ist ein generelles Problem digitaler sozialer Netzwerke.

Insgesamt erfordert die Faktenlage eine bessere Zusammenarbeit zwischen den Betreibern sozialer Medien und verschlüsselter Plattformen und Anti-Hass-Initiativen. Konzertierte rechtsextreme Aktionen müssen frühzeitig erkannt werden, so dass die Zivilgesellschaft Gegenmaßnahmen ergreifen kann. Das Zusammentragen von Informationen und die Aufklärung über die Gefahr, die von rechtsextremen Organisationen ausgeht, sind hier der Schlüssel. Dies ist eine gesamtgesellschaftliche Aufgabe. Hier sind die Plattformbetreiber in der Verantwortung, die technischen Voraussetzungen für die Entwicklung von Frühwarnsystemen, die Hass-Akkumulation in sozialen Medien schnell identifizieren könnten, zu schaffen, um NGOs, Aktivisten, Journalisten und andere Internet-

User vor Online-Angriffen zu schützen. Der Zugang zu öffentlich auf Facebook verfügbaren Informationen sollte wieder über eine Schnittstelle möglich sein. Die großen Ausmaße rechtsextremer Aktivität in den sozialen Medien müssen akribisch dokumentiert und aufbereitet werden.

In erster Linie sind aber Nutzerinnen und Nutzer der sozialen Medien und die digitale Zivilgesellschaft gefragt, im Netz Zivilcourage zu zeigen, die Opfer von Shitstorms unabhängig von eigenen politischen Ansichten zu unterstützen und die Spaltung der Gesellschaft durch einen respektvollen Umgang auch mit den Verursachern von Hassrede zu verhindern. Hasskampagnen verschleiern absichtlich ihre Urheber, um den Eindruck einer breit getragenen gesellschaftlichen Wut zu suggerieren. Organisatoren von politischen Grassroots-Kampagnen in sozialen Medien und Kommentarspalten sollten daher ihre Sympathisanten aufrufen, sich mit Hilfe von Hashtags klar zur eigenen Kampagne zu bekennen und sich ausnahmslos an den Grundsätzen respektvoller Kommunikation zu orientieren.

Auch die Verantwortlichen in den Medien selbst, insbesondere in Social-Media-Redaktionen, müssen für die Problematik sensibilisiert werden. Aktive Moderation kann zu einer deutlich verbesserten Diskussionskultur führen und öffnet die Kommentarspalten einer breiteren Zielgruppe. Moderatoren können dabei auf konstruktive und differenzierte Kommentare gezielt antworten und damit zu einer größeren Sichtbarkeit dieser Kommentare beitragen. Leicht zu manipulierende Like-Zahlen verlieren dadurch an Bedeutung, und Medien können verhindern, dass ihre Kommentarspalten und ihre Reichweite als Gratis-Plattform für verfassungsfeindliche und extremistische Gruppen missbraucht werden.

Nicht zuletzt müssen Verfassungsschutz, Strafverfolgungsbehörden und Justiz bei der Bekämpfung extremistischer Strukturen die sozialen Medien deutlich stärker in den Fokus nehmen. Der Staat muss in die Lage versetzt werden, seine Institutionen und die demokratische Zivilgesellschaft vor gezielten Angriffen im Internet mit rechtsstaatlichen Mitteln zu schützen. Das NetzDG hat hierbei bisher sein Ziel verfehlt. Wird eine Verfassungsfeindlichkeit der für den koordinierten Hass verantwortlichen Organisationen festgestellt, sind diese auf der Grundlage rechtsstaatlicher Verfahren konsequent in ihren Handlungsmöglichkeiten einzuschränken.

Weiterführende Literatur

Baldauf, Johannes, Miro Dittrich et al. (2017) – Toxische Narrative: Monitoring Rechts-Alternativer Akteure

Bednarz, Liane und Giesa, Christoph (2015): Gefährliche Bürger: Die neue Rechte greift nach der Mitte

Bruns, Julian, Kathrin Glösel und Natascha Strobl (2014) - Die Identitären: Handbuch zur Jugendbewegung der Neuen Rechten in Europa

Davey, Jacob und Ebner, Julia (2017) – The Fringe Insurgency: Connectivity, Convergence and Mainstreaming of the Extreme Right

Ebner, Julia (2018) - Wut: Was Islamisten und Rechtsextreme mit uns machen

Ley, Hannes und Görig, Carsten (2018) - #ichbinhier: Zusammen gegen Fake News und Hass im Netz

Nagle, Angela (2017) – Kill All Normies: Online Culture Wars From 4Chan And Tumblr To Trump And The Alt-Right.

Phillips, Whitney. (2015). This is why we can't have nice things: Mapping the relationship between online trolling and mainstream culture.

Rost, Katja, Lea Stahel, und Bruno S. Frey (2016) – Digital social norm enforcement: Online firestorms in social media.

Wagner, Thomas (2017) - Die Angstmacher: 1968 und die Neuen Rechten

Weiß, Volker (2017) – Die Autoritäre Revolte: Die Neue Rechte und der Untergang des Abendlandes

Wendling, Mike (2018) – Alt-Right: From 4Chan to the White House

Endnotes

1. Europarat, Ministerkomitee, Empfehlung Nr. (97) 20
2. Vgl. weiterführend Bundeszentrale für politische Bildung: <http://www.bpb.de/apuz/231307/medien-im-mainstream?p=all>
3. Mudde, Cas: "The populist zeitgeist." *Government and opposition* 39.4 (2004): 541-563
4. Was ist überhaupt Hate Speech?" in: „Geh sterben‘ – Hate Speech und Kommentarkultur im Internet“, Hrsg. Amadeu-Antonio-Stiftung, 2015
5. Für weitergehende Analysen zu diesem Thema: Jacob Davey und Julia Ebner, "The Fringe Insurgency: Connectivity, Convergence and Mainstreaming of the Extreme Right", ISD (Oktober 2017). Verfügbar hier: http://www.isdglobal.org/wp-content/uploads/2017/10/The-Fringe-Insurgency-221017_2.pdf
<https://kleineanfragen.de/bundestag/19/1994-aktivitaeten-des-rechten-internetnetzwerkes-reconquista-germanica>
6. <https://kleineanfragen.de/bundestag/19/1994-aktivitaeten-des-rechten-internetnetzwerkes-reconquista-germanica>
7. <http://www.fr.de/rhein-main/junge-alternative-aufstrammem-kurs-nach-rechtsaussen-a-1451025>
8. <https://www.youtube.com/watch?v=-IM4wtEr4No>
9. <https://www.verfassungsschutz.de/de/aktuelles/zur-sache/zs-2016-001-maassen-dpa-2016-08>

ISD

Powering new
generations
against extremism

PO Box 7814, London
United Kingdom, W1C 1YZ
www.isdglobal.org