

# Was bin ich – und wenn ja, wie viele?

## Identifikation und Analyse von Political Bots während des Bundestagswahlkampfs 2017 auf Twitter<sup>1</sup>

Fabian Pfaffenberger, Christoph Adrian und Philipp Heinrich

### 1 Einleitung

In einer Welt, in welcher der amtierende US-amerikanische Präsident Donald Trump seine politische Agenda hauptsächlich über Twitter kundtut, beantwortet sich die Frage nach der Relevanz dieses Social-Media-Kanals eigentlich von selbst. Generell bieten soziale Medien viele nützliche Funktionen, wie die Möglichkeit der Vernetzung und des interaktiven Austausches. Darüber hinaus nutzen immer mehr Deutsche soziale Netzwerke als Nachrichtenquelle, sodass unter den Internetnutzenden in Deutschland mittlerweile der Anteil der Personen, denen soziale Medien als Nachrichtenquelle dienen (31 %), sogar knapp den Anteil der Zeitungslesenden (29 %) übersteigt (Hölig und Hasebrink 2016, S. 535).<sup>2</sup> Die wahrgenommene Bedeutung von sozialen Medien steigt zusätzlich durch die zunehmende Einbindung von über soziale Medien geteilten Inhalten in der „regulären“ Berichterstattung, gepaart mit anhaltenden Aufforderungen der Medieninstitutionen, dem jeweiligen Angebot über einen der verschiedenen Social-Media-Kanäle zu folgen oder dort mitzudiskutieren. Kurz gesagt: Die sozialen Medien haben sich in den Informationsrepertoires etabliert.

Gleichzeitig bieten soziale Medien eine Möglichkeit, vorbei an den Selektionsmechanismen der traditionellen Medienakteure und -institutionen, eine breite Masse an

- 1 Diese Forschungsarbeit wurde in Teilen durch die Emerging Fields Initiative der Friedrich-Alexander-Universität Erlangen-Nürnberg unterstützt (Projekttitel: „Exploring the Fukushima Effect (EFE)“).
- 2 Basierend auf dem *Reuters Institute Digital News Survey 2016*. Es wurden 2035 Personen befragt. Die Umfrage wurde von dem Institut YouGov durchgeführt, „das auf der Basis von Online-Access-Panels Stichproben zog, die für die Internetnutzer der beteiligten Länder ab 18 Jahren repräsentativ sind“ (Hölig und Hasebrink 2016, S. 535).

Menschen direkt und nahezu ungefiltert anzusprechen. Dass damit im politischen Kontext, speziell bei Wahlen, gewisse Risiken einhergehen, verdeutlicht eine im Juli 2018 erschienene Meldung der *Washington Post*, der zufolge Twitter mehr als 70 Millionen Accounts im Mai und Juni desselben Jahres gesperrt habe (Timberg und Dwoskin 2018). Diese Welle an Sperrungen resultierte aus einer Anhörung im US-Kongress, bei der verschiedene amerikanische Social-Media-Unternehmen zum Einfluss von ausländischen Fake-Accounts auf die US-Präsidentschaftswahl 2016 befragt wurden. In der Anhörung relativierten die Social-Media-Unternehmen ihre vorherigen Aussagen über lediglich marginale Effekte auf die Wahl und offenbarten ein erheblich höheres Ausmaß an ausländischer Einflussnahme (Timberg und Dwoskin 2017). Konkret geht es um die gezielte Verbreitung von Desinformationen, teilweise mit Hilfe von (Social) Bots, deren Einfluss – besonders auf Twitter – bereits vor dem öffentlichen Eingeständnis von verschiedenen Studien (z. B. Bessi und Ferrara 2016; Kollanyi et al. 2016) angedeutet wurde. Hinter dem Begriff Social Bots verbergen sich „computer-generated programs that post, tweet, or message of their own accord“ (Howard und Kollanyi 2016, S. 1), welche teilweise zur Imitation menschlicher Akteure eingesetzt werden, um aktiv auf die öffentliche Gewichtung und Deutungsrichtung von Themen Einfluss auszuüben (Graber und Lindemann 2018, S. 52). Sofern dieser Einsatz im politischen Kontext geschieht, spricht man von *Political Bots*.

Die Sorge um die Beeinflussung des Bundeswahlkampfes durch Political Bots zeigte ein Appell von Bundeskanzlerin Angela Merkel an die anderen Parteien im Oktober 2016, im anstehenden Wahlkampf auf die Nutzung von Social Bots zu verzichten, dem letztendlich nur die *Alternative für Deutschland [AfD]* widersprach (Amann et al. 2016), wenngleich die Partei diese Aussage später wieder relativierte.

Dass Bots auch im Bundestagswahlkampf von Relevanz sein könnten bzw. dass die Sorge um den Einfluss von Bots nicht unberechtigt ist, zeigt eine Studie der *Amadeu Antonio Stiftung* (Baldauf et al. 2017). Für die Analyse wurden im Zeitraum zwischen April 2016 und Februar 2017 die zehn beliebtesten Beiträge der Facebook-Seiten verschiedener, dem rechten Spektrum zuzuordnender Social-Media-Akteure<sup>3</sup> mithilfe des Tools *FanPage Karma* gesammelt und qualitativ analysiert. Die Auswahl der Facebook-Seiten erfolgte bewusst und mit dem Ziel, „das breite Spektrum rechtsextremer und rechtspopulistischer Strömungen und Akteure exemplarisch“ abzubilden (Baldauf et al. 2017, S. 11). Die Untersuchung

---

3 Ausgewählte Facebook-Seiten: *Compact-Magazin*; *Ich bin stolz, deutsch zu sein*; *Lügenpresse*; *PI-News*; *AfD*; *Identitäre Bewegung*; *Ein Prozent für unser Land*; *PEGIDA*, *NPD*, *KenFM*

identifiziert sieben *toxische Narrative*<sup>4</sup>. Dies sind Erzählungen, die für Kommunikation bzw. sprachliches Verhalten, das seine Umgebung negativ beeinflusst, die entsprechenden Ereignisse und Interpretationen liefern (Baldauf et al. 2017, S. 6).

Der Effekt dieser toxischen Narrative kann durch den gezielten Einsatz von Bots zusätzlich verstärkt werden, indem durch eine große Anzahl von Reaktionen (z. B. Likes, Shares, Retweets, Kommentare etc.) bei anderen Nutzenden fälschlicherweise ein Eindruck der Zustimmung einer breiten Masse an Personen entsteht (Graber und Lindemann 2018, S. 57). Eine Erklärung für diesen Effekt liefert das Prinzip der sozialen Bewährtheit (Cialdini 2013), also der Tendenz eines Menschen, sich besonders in unüberschaubaren Situationen der Mehrheitsmeinung anzuschließen und die eigene Handlung aus den Handlungen Anderer abzuleiten. Erschwerend kommt hinzu, dass verschiedene Studien auf Probleme bei der Unterscheidung zwischen von Bots oder Menschen kreierten Inhalten hinweisen (Edwards et al. 2014; Everett et al. 2016; Freitas et al. 2016), woraus sich die Relevanz dieser Untersuchung ableitet.

Dementsprechend geht es in diesem Beitrag darum, die Verwendung potentieller Political Bots beim Bundestagswahlkampf auf Twitter zu ermitteln. Es stellt sich die Frage, ob und in welchem politischen beziehungsweise thematischen Umfeld Political Bots im Zuge des Bundestagswahlkampfes eingesetzt wurden. Die Auswahl des Social-Media-Kanals Twitter als Datenbasis erfolgt aus verschiedenen Gründen. Zwar nutzten 2016 lediglich 9,8 % der Internetnutzenden in Deutschland die Plattform Twitter allgemein und nur 4,1 % als Nachrichtenquelle<sup>5</sup> (Hölig und Hasebrink 2016, S. 539), doch speziell durch die Einbindung von bzw. Referenzen auf Tweets in den traditionellen Medien ergibt sich eine höhere Bedeutung. Zudem lässt sich im Vergleich zu Facebook eine Nachricht besser und schneller verbreiten oder ein bestimmtes Hashtag eher manipulieren, da Tweets grundsätzlich öffentlich und nicht nur für Nutzende, sondern auch für Nicht-Nutzende komplett sichtbar sind. Das wohl simpelste sowie entscheidendste Argument ist jedoch die Verfügbarkeit kostenloser Daten: Twitter ermöglicht unter anderem einen frei verfügbaren, aber restringierten Echtzeit-Zugang zum kompletten Twitter-Traffic<sup>6</sup>, während bei Face-

---

4 Hauptnarrative: *Der Untergang der Deutschen, Bedrohungen von außen, Bedrohungen von innen, Das Establishment manipuliert, Globale Verschwörung, Widerstand und Lösung, Repression und Widerstand*

5 Facebook wurde hingegen von 51,5 % der Personen generell bzw. 26,7 % als Nachrichtenquelle genutzt (Hölig und Hasebrink 2016, S. 539)

6 Das Datenvolumen der Search API-Abfrage (z. B. Suchterm-Filter) darf maximal ein Prozent des gesamten Twitter-Traffics zu einem Zeitpunkt betragen. Danach wird das Suchergebnis gedeckelt.

book nur Beiträge vorher spezifizierter (öffentlicher) Seiten oder Gruppen abrufbar sind. Einen API<sup>7</sup>-Zugriff auf alle (öffentlichen) Facebook-Beiträge gibt es nicht.

Die während der heißen Phase des Bundestagswahlkampfes gesammelten Tweets werden mit Hilfe eines zweistufigen Identifikations- und Validierungsverfahrens untersucht. Im ersten Schritt erfolgt anhand eines einfachen Nahduplikat-Erkennungsalgorithmus die Identifikation von potentiellen Bot-Twitter-Accounts, basierend auf dem Umfang der Verbreitung ähnlicher bis identischer Tweets. Diese auffälligen Accounts (typischerweise solche mit einem sehr hohen Nahduplikatan-teil) werden im zweiten Schritt einer manuellen qualitativen inhaltlichen Analyse unterzogen, um die Bot-Klassifizierung zu validieren. Zwar liefern die Ergebnisse wenig Indizien für umfangreiche Maßnahmen zur Beeinflussung des Wahlausgangs durch Political Bots, offenbaren aber zumindest den Versuch der Einflussnahme durch verschiedene Akteure und Gruppen.

---

## 2 Forschungsstand

Gerade bei der Betrachtung des Einsatzes von Bots und Fake News in den sozialen Medien in Zeiten des Wahlkampfes und der häufig damit einhergehenden negativen Konnotation tritt die eigentliche Multidimensionalität bzw. die Multifunktionalität eines *Bots* schnell in den Hintergrund. Während der vermehrte Einsatz von automatisierten Social-Media-Accounts durch Nachrichtenorganisationen zur Ausbreitung von Informationen neben Chancen auch Herausforderungen mit sich bringt (Lokot und Diakopoulos 2015), dürften zumindest die durch einen japanischen Twitter-Bot<sup>8</sup> erstellten Erdbebenwarnungen als gänzlich unproblematisch angesehen werden.

Die Diversität wird besonders durch das von Stieglitz et al. (2017) vorgestellte Bot-Kategorisierungsschema veranschaulicht, welches die Bot-Operationalisierungen verschiedener Studien, basierend auf den bestehenden, in der Literatur verbreiteten Unterscheidungsmerkmalen *Intention* (Ferrara et al. 2016) sowie *Imitation menschlichen Verhaltens* (Boshmaf et al. 2013), klassifiziert.

---

7 Application Programming Interface; Programmier-Schnittstelle zur Abfrage, Einspeisung und Manipulation von Daten.

8 Account: @earthquake\_jp; [https://twitter.com/earthquake\\_jp](https://twitter.com/earthquake_jp)

Nachahmung menschlichen Verhaltens	Absicht		
	bösartig	neutral	gutartig
	hoch: social bots		
	<ul style="list-style-type: none"><li>- Astroturfing Bots</li><li>- Social-Bot-Netzwerke in polit. Konflikten</li><li>- Infiltration einer Organisation</li><li>- Influence Bots</li><li>- Sybils</li><li>- Doppelgänger Bots</li></ul>	<ul style="list-style-type: none"><li>- Humoristische Bots</li></ul>	<ul style="list-style-type: none"><li>- Chatbots</li></ul>
	niedrig bis gar nicht		
	<ul style="list-style-type: none"><li>- Spambots</li><li>- Fake Accounts zur Kontrolle &amp; Steuerung von Bot-Netzwerken</li><li>- Paybots</li></ul>	<ul style="list-style-type: none"><li>- Nonsensebots</li></ul>	<ul style="list-style-type: none"><li>- Newsbots</li><li>- Recruitment Bots</li><li>- Public Dissemination Account</li><li>- Erdbebenwarnbots</li><li>- Editing Bots, Anti-Vandalism Bots bei Wikipedia</li></ul>

**Abb. 1** Schema zur Kategorisierung von Social Media Bot Accounts (nach Stieglitz et al. 2017, S. 7)

Auch wenn Abb. 1 lediglich ausgewählte Beispiele und nicht alle in der Studie untersuchten Artikel enthält, offenbart sich schnell eine Art Muster. Zum einen überwiegen in der Stichprobe die schädlichen Bots, 52 von 103 Artikeln nutzen das Wort „malicious“ bei der Beschreibung der jeweiligen Bot-Operationalisierung. Gleichzeitig zeigt sich, dass gutartige Bots größtenteils nützliche automatisierte Informationen liefern.

Mit der dargestellten Vielzahl unterschiedlicher Bot-Arten geht eine ebensolche Fülle an Ansätzen zur Detektion einher: von recht simplen Operationalisierungen, die auf singulären Indikatoren wie die Anzahl täglich versendeter Tweets (> 50) (Kollanyi et al. 2016) oder der Anzahl an täglichen Favorites (> 50) (Wilke 2017) beruhen, über die Tweet-URL<sup>9</sup> (Hegelich und Janetzko 2016, S. 579), Geo-Tagging (Guo und Chen 2014) oder die Nutzung vorhandener Bots (Lokot und Diakopoulos 2015) bis hin zu komplexen Machine-Learning-Algorithmen (Alarifi et al. 2016; Davis et al. 2016; Ratkiewicz et al. 2011; Varol et al. 2017) oder Deep-Learning-Ansätzen (Cai et al. 2017), die verschiedene User-, Freundes- und Netzwerkmerkmale sowie zeitliche, textliche und inhaltliche Merkmale der Tweets einbeziehen bzw. kombinieren.

9 Damit ist die URL gemeint, von welcher der Tweet gesendet wurde. Mit Hilfe der URL wurde im Fall von Hegelich und Janetzko (2016, S. 579) die Verbindung zu dem Programm *twifarm* hergestellt, welches die Verwaltung einer großen Anzahl von Bots ermöglicht.

Doch gerade bei der Berücksichtigung der textbasierten Inhalte eines Tweets gehen die Bot-Identifikationsalgorithmen selten über die Auswertung von Verteilungskennzahlen, wie Häufigkeit bzw. Anteil bestimmter POS-Tags innerhalb eines Tweets sowie die Anzahl und Entropie (Informationsgehalt) der Wörter eines Tweets (Varol et al. 2017, S. 3) oder die Anzahl der Zeichen pro Tweet (Alarifi et al. 2016, S. 336), hinaus. Speziell im Hinblick auf die im Wahlkampf vermutete Verwendung toxischer Narrative dürften die konkreten Inhalte der Tweets von stärkerer Bedeutung sein. Deshalb wird in der vorliegenden Untersuchung der korpuslinguistische Ansatz von Schäfer et al. (2017) zur Identifikation von Bot(nets) aufgegriffen. Im Fokus dieses Ansatzes steht die Analyse von Nahduplikaten (ND, *Near Duplicates*) – das sind Tweets mit identischem oder nahezu identischem Inhalt (siehe Abschnitt 3). Zwar hat diese Methode im Vergleich zu den erwähnten komplexeren Bot-Identifikationsmodellen offensichtliche Limitationen, jedoch ist das Kopieren bzw. Retweeten eines Tweets (mit geringfügigen Abweichungen) eine effiziente Strategie um ein bestimmtes Thema zu pushen, welche zudem in der vorherigen Forschung (vgl. Hegelich und Janetzko 2016; Howard und Kollanyi 2016; Kollanyi et al. 2016) bereits nachgewiesen wurde (Schäfer et al. 2017, S. 300).

## 2.1 Bots im politischen Kontext

Die Vielfältigkeit der Operationalisierung und verschiedener Arten von Bots zeigt sich zwar mit Blick auf die Anzahl an Publikationen zum Einsatz von Bots im politischen Kontext, jedoch nicht in Bezug auf die Anlässe. Nur wenige Publikationen widmen sich der Thematik außerhalb politischer Abstimmungen, wie z. B. die Studien zur Identifikation und Untersuchung von Botnets im syrischen Bürgerkrieg (Abokhodair et al. 2015) oder im Ukraine-Konflikt (Hegelich und Janetzko 2016). Der Großteil der Forschung fokussiert sich auf den unmittelbaren Zeitraum vor politischen Wahlen, wie beispielsweise in Venezuela (Forelle et al. 2015), Japan (Schäfer et al. 2017) oder den Vereinigten Staaten von Amerika (Kollanyi et al. 2016) – dem wohl prominentesten Beispiel im internationalen Kontext. Bei allen Unterschieden, die mit den variierenden Strukturen der politischen Systeme, Stichproben und zur Bot-Identifikation gewählten Operationalisierung einhergehen, offenbart sich mit Blick auf deren Ergebnisse die grundlegende Tendenz, dass Bots inhaltlich vorwiegend zur Verbreitung radikaler und ultra-konservativer Positionen eingesetzt werden. Ebenso verwenden radikale oder ultra-konservative Parteien Bots häufiger als gemäßigte Akteure. Der deutlichste Effekt zeigt sich dabei in den USA, wo während des Wahlkampfes nach Schätzungen knapp ein Fünftel der auf

Twitter generierten Inhalte mit Wahlbezug von Bots kreiert wurde (Bessi und Ferrara 2016), die zudem häufiger pro Trump als pro Clinton waren (Kollanyi et al. 2016).

Im europäischen Kontext traten Bots vor allem beim Brexit-Votum in Erscheinung. Howard und Kollanyi (2016, S. 1) bescheinigen den Bots eine kleine aber strategisch wichtige Rolle innerhalb der Debatte um das Referendum, da diese besonders den Austritt befürwortende Hashtags bewarben. Die Analyse eines 13.493 Accounts umfassenden Botnets von Bastos and Mercea (2017) kommt zu ähnlichen Ergebnissen, obgleich der Fokus ihrer Untersuchung auf der Verbreitung von „hyperpartisan news“ lag. Dennoch dienen die eingesetzten Bots ebenfalls hauptsächlich zur schnellen Verbreitung von Informationen und weniger zur gezielten anspruchsvollen Beeinflussung der komplexen Brexit-Debatte auf Twitter. Zwar gab es bei den französischen Präsidentschaftswahlen 2017 ebenfalls den Versuch der Manipulation in Form der *#macronleaks* Kampagne. Jedoch zeigt Ferrara (2017), dass hauptsächlich ausländische Personen mit bestehenden Vorkenntnissen von bzw. Interesse an rechten Nachrichteninhalten die Kampagne aufgreifen.

Die Forschung zum Einsatz und Einfluss von Bots bei politischen Abstimmungen in Deutschland vor der Bundestagswahl 2017 ist rar. Zwar gibt es Studien zu den ebenfalls 2017 durchgeführten Wahlen des Bundespräsidenten (Neudert et al. 2017a) und des nordrhein-westfälischen Landtags (Brachten et al. 2017), welche jedoch beide nur einen geringen bzw. nicht wahrnehmbaren Einsatz von Political Bots finden.

Im Bezug zur Bundestagswahl 2017 zeigen sich erste Ergebnisse schon mit Blick auf das kurz vor der Wahl stattfindende Kanzlerduell am 3. September 2017. Dort stellen Grimme et al. (2017) durch das Streaming ausgewählter Hashtags während der Debatte und anschließender Analyse der gesammelten Daten fest, dass überproportional viele Accounts, die jünger als einen Monat sind, gleichzeitig die Popularität von *#kanzlerduell* und *#verräterduell* durch hochfrequente Verwendung erhöhen. Ebenso zeigt sich eine Überschneidung zwischen Accounts, die sowohl das Hashtag *#afd* als auch *#verräterduell* nutzen. Kurz nach dem Ende des Duells wurde jedoch bekannt, dass es sich dabei um eine gezielte Aktion des rechten Netzes *Reconquista Germanica* handelte, das auf den Einsatz von Bots verzichtete (Schmehl 2017). Neudert et al. (2017b, S. 3–4) schreiben Bots aufgrund des geringen Umfangs automatisiert verbreiteter Nachrichten eine geringe Bedeutung im Wahlkampf zu. Zugleich zeigen weitere Ergebnisse der Studie eine im Verhältnis übergroße Menge an Tweets mit Bezug auf die AfD. Grimme et al. (2018) fanden zudem in der Zeit unmittelbar vor den Wahlen (10. bis 25. September) Aktivitäten von Bots der Freien Wähler.

Diese und die folgenden Ergebnisse sollten jedoch mit Blick auf den Kontext der Wahl mit besonderer Vorsicht betrachtet werden. Zum einen vollzog sich der gesamte Bundestagswahlkampf während der noch immer anhaltenden Debatte um ausländi-

sche Einflussnahme – vor allem auch durch die gezielte Nutzung von Bots – auf die US-amerikanische Präsidentschaftswahl. Es ist daher nicht auszuschließen, dass dieser Umstand bei allen Akteuren, aber besonders bei den nutzenden Personen und den Betreibenden von Social-Media-Plattformen, zu einer Sensibilisierung im Hinblick auf die Identifikation, Meldung und Sperrung auffälliger Accounts geführt haben könnte. Ebenfalls lässt sich aus dem zahlreichen Verschwinden vieler Bot-Accounts nach dem Brexit-Referendum (Bastos und Mercea 2017) die Vermutung ableiten, dass die Anzahl der Bots während des Wahlkampfes grundsätzlich höher ist.

---

### 3 Methodik

Der Basisdatensatz besteht aus 2.939.811 originären Tweets (keine Retweets über den Marker *RT*) von insgesamt 318.451 Benutzern, die zwischen dem 1. August und 27. September 2017 über die Streaming-API (*statuses/filter*) gesammelt wurden. Die Sammlung basiert auf einer Liste von 122 Keywords (siehe Anhang), anhand derer der Datenstrom der API gefiltert wurde. Da der Datensatz aus einem anderen Forschungsprojekt übernommen wurde und zum Analysezeitpunkt bereits bestand, konnte kein Einfluss auf die Auswahl der Keywords genommen werden.

Eine Datensammlung über die Streaming API mit Keyword-Filtern hat mehrere Vorteile gegenüber anderen Ansätzen. So ist im Vergleich zur Nutzung des allgemeinen 1 % Samples der API, das ein Abbild der gesamten Twitter-Aktivität generiert, bei spezifischen Keyword-Filtern die Wahrscheinlichkeit deutlich höher, einen möglichst vollständigen Datensatz für die Untersuchung *relevanter* Tweets zu erhalten, auch wenn hier die Gefahr von Rate Limits<sup>10</sup> besteht (Pfaffenberger 2018). Zudem bietet die Stichprobe keine zuverlässige Repräsentativität (vgl. Morstatter et al. 2013; Tromble 2018). Eine Ex-post-Nutzung der REST APIs (Search API) wäre in diesem Forschungskontext sogar hinderlich, da bereits gelöschte, widerrufen oder gesperrte Tweets nachträglich nicht mehr abrufbar sind. Sofern Twitter in der Zwischenzeit Bots identifiziert und gesperrt oder gelöscht hat, würden deren Tweets daher im gesammelten Datensatz fehlen.

Somit sind Keyword-Filter die einzige praktikable Methode zur Sammlung potentiell relevanter Tweets zur Bundestagswahl, wenngleich dieser Ansatz auch mehrere Nachteile mit sich bringt. Qualität und Quantität der Daten hängen stark von der

---

10 Unter Rate Limits versteht man die Deckelung des Datenstroms, wenn das Volumen der Keyword-Filter 1 % des momentanen Twitter-Volumens übersteigt (Pfaffenberger 2018, S. 58).



Wahl der Keywords ab – ein ungeeignetes Keyword kann den Datensatz schnell „verwässern“, wohingegen fehlende Begriffe unter Umständen die Aussagekraft der Ergebnisse mindern. Zudem erschweren sprachliche Überschneidungen eine korrekte Auswahl deutschsprachiger Tweets. Ein Beispiel ist das Kürzel *BTW*, das einerseits im deutschen Sprachraum für Bundestagswahl, international aber auch für das Englische „by the way“ (dt. „übrigens“) verwendet wird. *EU* steht weltweit nicht nur für die Europäische Union, sondern bedeutet auf Portugiesisch „ich“. Dementsprechend ist einerseits eine vorsichtige Wahl der Keywords entscheidend, andererseits mussten die erfassten Tweets nachträglich noch über einen Sprachfilter selektiert werden (um beispielsweise Tweets auf Portugiesisch und Englisch zu entfernen). Die verwendeten 122 Keywords können den Anspruch auf Vollständigkeit gewiss nicht wahren, sollen aber als Abbild der gesellschaftlichen Diskurse zu dieser Zeit dienen.

Eine erste Sichtung der Tweet-Frequenz pro Tag zeigt eine Zunahme der durch die Keyword-Liste erfassten (relevanten) Twitter-Aktivität zur Bundestagswahl, weshalb der Analysezeitraum nachträglich auf die heiße Phase vor der Bundestagswahl bis zwei Tage nach dem Wahltermin begrenzt wird. Konkret bedeutet das eine Beschränkung der Analyse auf den einmonatigen Zeitraum vom 27. August bis 26. September. Aufgrund technischer Probleme steht für den 17. September nur eine sehr geringe Anzahl an Tweets zur Verfügung, sodass dieser Tag bei der späteren qualitativen Betrachtung nicht berücksichtigt wird. Der Analysedatensatz besteht aus 2.034.801 *originären Tweets* und enthält keine reinen Retweets.

Da eine manuelle Analyse bei dieser Anzahl an Tweets unmöglich ist, bedarf es algorithmisch gestützter Methoden zur Identifikation von Social Bots. Wie bereits in Kapitel 2 beschrieben, existiert mittlerweile eine Vielzahl unterschiedlicher Ansätze zur Erkennung (teil-)automatisierter Twitter-Accounts. Die folgende Analyse basiert auf einem Nahduplikat-Erkennungsalgorithmus, der Tweets anhand ihrer Ähnlichkeit zu Clustern gruppiert. Dieser computerlinguistische Ansatz betrachtet folglich keine quantitative Metriken, wie die Tweet-Aktivität, Frequenz oder die Nutzer-Vernetzung, sondern analysiert die verbreiteten textbasierten Inhalte.

Die gesammelten Tweets werden zunächst mit einem auf internetbasierte Kommunikation spezialisierten Tool tokenisiert<sup>11</sup>, d. h. in einzelne Wörter und Satzzeichen zerlegt. Darauf folgt die Normalisierung dieser Tweets, welche alle im Tweet enthaltenen URLs, Mentions, RT-Marker sowie Interpunktion, Leer- und Sonderzeichen (inklusive @ und #) entfernt. Der letzte Schritt umfasst die Erkennung exakter Duplikate der normalisierten Tweets. Dazu wird jedem normalisierten Tweet ein Hashwert zugewiesen und alle Tweets mit identischem Hashwert einem

---

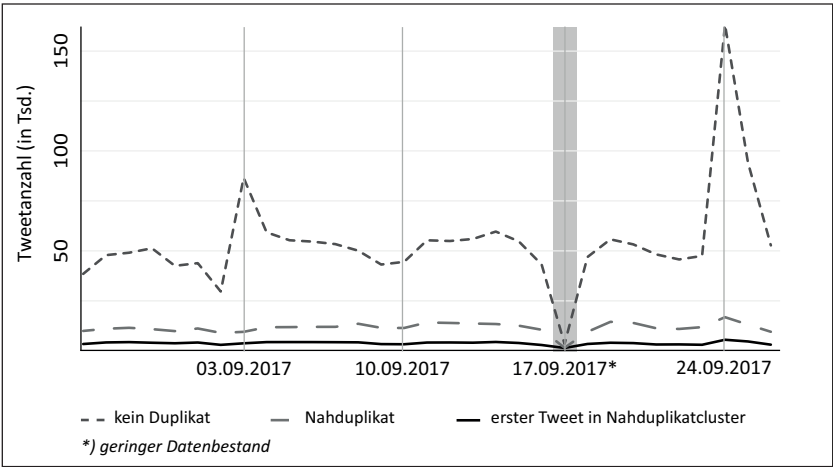
11 SoMaJo (Proisl und Uhrig 2016).

Cluster zugeordnet. Anhand ihres makroskopischen Duplikatstatus erfolgt die Einordnung der Tweets in drei Kategorien: „kein Duplikat“ (0), „echtes Duplikat“ (1) und „erster Tweet in einer Reihe von Duplikaten“ (9). Als Nahduplikate gelten somit im Folgenden ähnliche Tweets, die nach Tokenisierung und Normalisierung identische Hashwerte aufweisen. Alle ähnlichen Tweets innerhalb der Cluster erhalten Status 1, wobei der jeweils erste Tweet (nach Zeitstempel) eines Clusters den Status 9 erhält. Tweets sind demzufolge Nahduplikate mit Duplikatstatus 1, wenn sie eine starke Ähnlichkeit zum ersten Tweet des Nahduplikatclusters (mit Status 9) aufweisen. Alle anderen einzigartigen Tweets erhalten den Status 0. Tab. 1 zeigt die Häufigkeitsverteilung im Analysezeitraum nach Duplikatstatus und Account-Anzahl, Abb. 2 stellt die Werte im Zeitverlauf dar.

**Tab. 1** Häufigkeitsverteilung der Tweets nach Duplikatstatus und Accounts im Analysezeitraum

Status	Tweet-Anzahl (in %)	Urheber-Accounts
0 – Kein Duplikat	1.643.214 (80,8)	205.331
1 – Nahduplikat	315.995 (15,5)	67.330
9 – Erster Tweet eines Nahduplikatclusters	75.592 (3,7)	21.097
Gesamt	2.034.801 (100)	245.492

Analysezeitraum: 27. August bis 26. September 2017



**Abb. 2** Duplikatstatus der gesammelten Tweets im Zeitverlauf

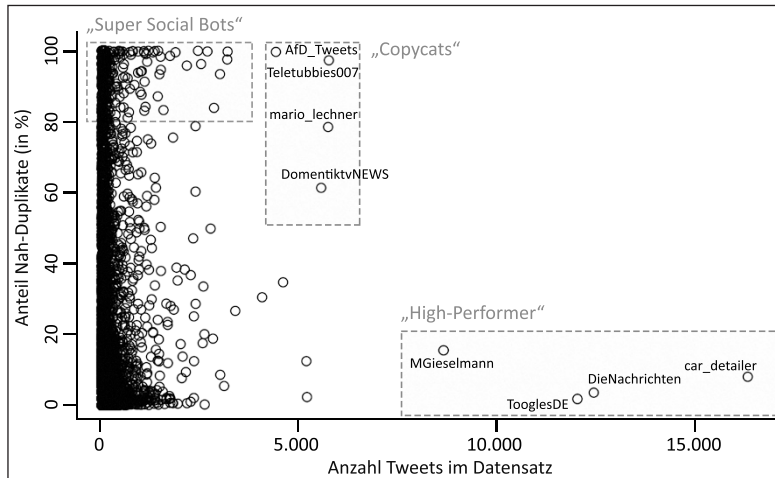
Wie erwartet, sind die meisten Tweets im Analysezeitraum (1.643.214, 80,8 %) keine Duplikate. Dem gegenüber stehen 315.995 Tweets, die sich auf 75.592 Cluster von 21.097 unterschiedlichen Nutzern verteilen. Es fällt auf, dass der tägliche Anteil an Nahduplikaten ( $M = 0.16$ ,  $SD = 0.03$ ) und Nahduplikatclustern ( $M = 0.04$ ,  $SD = 0.01$ ) im Zeitverlauf stabil bleibt. Entgegen allen Vermutungen gibt es am Tag der Bundestagswahl zwar einen absoluten Anstieg an Nahduplikaten, aber in Relation zum sehr hohen Tweet-Volumen (insgesamt 181.526 erfasste Tweets) einen anteilmäßigen Tiefpunkt mit nur neun Prozent Nahduplikaten. Eine allgemeine Erklärung hierfür könnte eine insgesamt stärkere Aktivität auf Twitter am Wahltag sein, die über die relativ konstante, potentiell automatisierte Verbreitung von Nahduplikaten hinausgeht. Allerdings kann an dieser Stelle auch eine rein technische Erklärung nicht ausgeschlossen werden: Je früher ein Tweet im Untersuchungszeitraum getätigt wird, desto höher ist die generelle Wahrscheinlichkeit, dass ein weiterer Tweet zu späterem Zeitpunkt ein Nahduplikat ist. Somit gibt es für Tweets am Ende des Untersuchungszeitraums eine geringere Wahrscheinlichkeit, dass es weitere ähnliche Tweets gibt.

Da eine Aussage über Bot-Aktivitäten auf einem makroskopischen Niveau nicht sinnvoll ist und eine vollständige Analyse aller Accounts, Nahduplikate oder Nahduplikatcluster den Rahmen dieses Beitrags sprengen würde, werden im folgenden Kapitel einzelne interessante Accounts innerhalb der Nahduplikatcluster identifiziert und deren Tweets betrachtet.

---

## 4 Betrachtung auffälliger Accounts

Zur Identifikation potentiell interessanter Accounts – hier: Political Bots – gibt es mehrere Möglichkeiten. So könnte man die Tweet-Aktivität jedes Nutzers, die Account-Eigenschaften oder die geteilten Inhalte beziehungsweise deren Ähnlichkeit betrachten. Für eine einfachere Auswahl werden zunächst der Zusammenhang zwischen der Anzahl erfasster Tweets je Nutzer im Datensatz und dem Anteil an Nahduplikaten grafisch dargestellt (siehe Abb. 3). Da eine erste Betrachtung der Häufigkeitsverteilung nach Nutzern bereits verdeutlichte, dass der überwiegende Teil der im Datensatz erfassten Accounts wenige Tweets während des Analysezeitraums teilte, wurden aus Gründen der Sparsamkeit und Relevanz nur Accounts mit mehr als 50 Tweets im Datensatz berücksichtigt.



**Abb. 3** Nutzerverteilung anhand Tweet-Anzahl und Anteil Nahduplikate  
(Basis: Nutzer mit mehr als 50 Tweets im Datensatz)

Es zeigt sich, wie erwartet, trotz Ausschluss wenig aktiver Accounts noch immer eine Häufung von Nutzern am linken Rand. Es gibt also viele Accounts mit einer geringen Tweet-Anzahl, wobei der Anteil an Nahduplikaten zu beiden Extremen stark variiert. In dieser Gruppe sind besonders die Accounts von Interesse, die eine hohe Anzahl an Nahduplikaten, gleichzeitig aber eine insgesamt moderate Tweet-Häufigkeit aufweisen. Ein hoher Anteil ähnlicher Tweets bei relativ geringer Tweet-Anzahl wird in dieser Analyse als ein Indiz für Social Bots gewertet. Schließlich würde eine sehr hohe Anzahl sehr ähnlicher Tweets zum einen bei anderen Nutzern leicht Misstrauen gegenüber dem Account beziehungsweise den Verdacht von Bots wecken. Ebenso bestünde die Gefahr, dass Twitter den Account sperrt. Zum anderen ist ein Netzwerk mehrerer moderat aktiver Bots wirkungsvoller als einzelne sehr aktive Bots. Wird eine Botschaft/Narrative/Nachricht/Meinung/Information durch mehrere Accounts verbreitet, steigt die Wahrscheinlichkeit, dass sie wahrgenommen und ihr geglaubt wird, als wenn einzelne Nutzer isoliert ständig einen minimal abgewandelten Tweet posten. Deshalb agieren wirkungsvolle Social bzw. Political Bots eher in Gruppen „unter dem Radar“ und verbreiten ähnliche Inhalte in einem nur moderaten Umfang (*Astroturfing*<sup>12</sup>).

12 Unter *Astroturfing* versteht man die zentrale Steuerung scheinbar unabhängiger Akteure im politischen Kontext, um eine spontane Graswurzelbewegung vorzutäuschen.

Zur detaillierteren Analyse werden im Folgenden Accounts mit weniger als 4.000 Tweets bei einem Anteil an Nahduplikaten von mehr als 80 % betrachtet und als *Super Social Bots [SSB]* bezeichnet (in der Grafik oben links). Daneben fallen zwei weitere in sich relativ homogene Cluster ins Auge: eine Gruppe von vier Accounts mit sehr hoher Tweet-Frequenz aber geringer Nahduplikatrate (*High-Performer*) und vier Accounts mit etwa 5.000 Tweets und einem relativ hohen Anteil an Nahduplikaten (*Copycats*). Nach der algorithmus-basierten Identifikation potentieller Bots erfolgt nun in den nächsten Unterkapiteln eine quantitativ-qualitative Betrachtung dieser Account-Gruppen.

4.1 Super Social Bots

Das Cluster der Super Social Bots besteht aus 221 Accounts mit ähnlicher Tweet-Häufigkeit und Nahduplikatrate im Datensatz: Weniger als 4.000 Tweets im Erhebungszeitraum, die sich inhaltlich relativ ähnlich sind (80-100 % Nahduplikate). Die quantitative Betrachtung der Account-Aktivität hinsichtlich Datum und Uhrzeit liefert erste Anhaltspunkte für automatisierte Twitter-Nutzung. So gibt es Accounts, die konstant über den ganzen Tag verteilt eine hohe Anzahl an Tweets verbreiten, während andere ein menschlicheres Twitter-Verhalten mit Ruhezeiten aufweisen (s. Abb. 4). Dennoch besitzt eine rein auf zeitliche Frequenzmuster bezogene Betrachtung eine geringe Aussagekraft. Wie bereits erwähnt, besitzen intelligente Bots die Fähigkeit, menschliches Verhalten zu imitieren, oder sind bewusst nur dann aktiv, wenn ein Großteil der Zielgruppe Twitter aktiv nutzt.

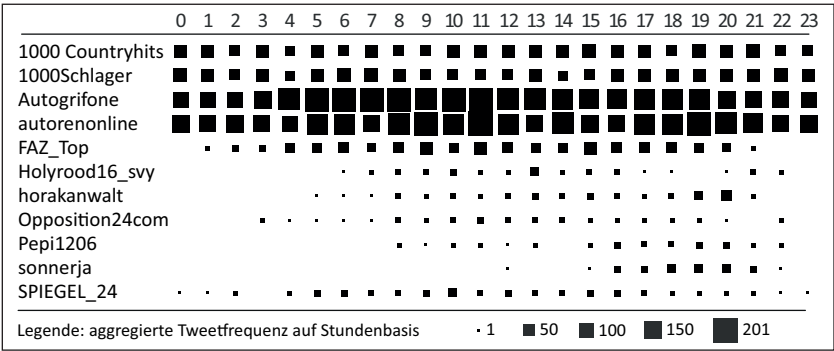
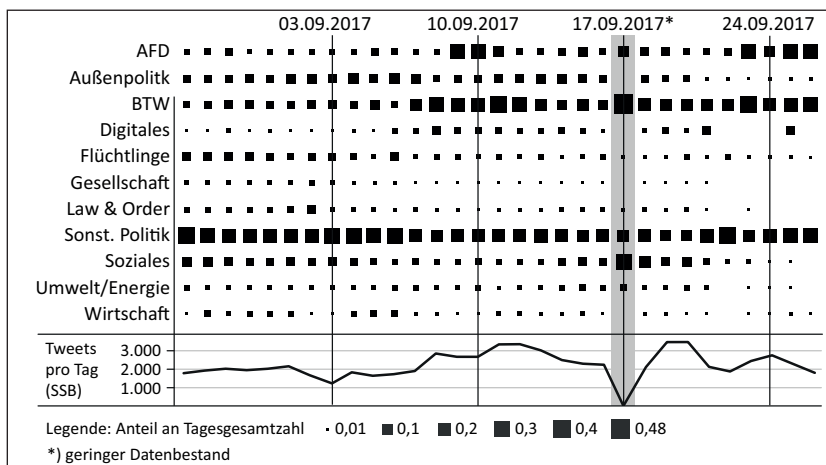


Abb. 4 Aktivitätsmuster ausgewählter Nutzer nach Uhrzeit



Social Bots nach Signalwörtern des Wörterbuchs durchsucht und binär kodiert (0 nicht vorhanden / 1 vorhanden).

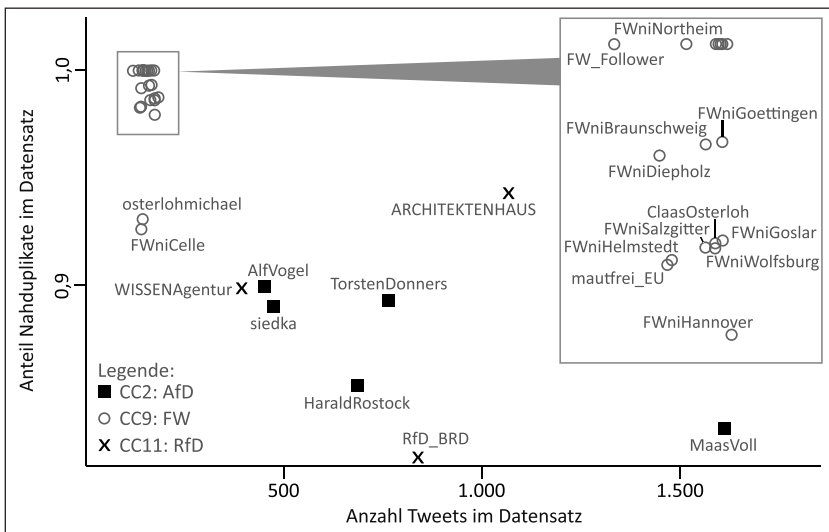
Die Betrachtung der Häufigkeiten zeigt, dass die Themenkomplexe Bundestagswahl, Soziales und Sonstige Politik bei den erfassten Tweets der Super Social Bots einen großen Anteil haben. Wie vermutet, beinhalten viele Tweets Begriffe, die der AFD zugeordnet werden können, wie *#nichtmeinekanzlerin* oder *#merkelmussweg*. Abb. 6 stellt die Häufigkeit ausgewählter Themen im Zeitverlauf dar. Die Flüchtlingsthematik nimmt im Laufe der Zeit kontinuierlich ab und spielt eine deutlich geringere Rolle, als ursprünglich vermutet. Besonders ins Auge fallen die Anteils-häufungen von AfD-Themen am 9. und 10. sowie am 23. September. Die Spitzen am 9. und 10. September resultieren aus einer koordinierten Aktion rechtspopulistischer Gruppen: Hier stechen vor allem die Hashtags *#ToxischeNarrative* und *#LügenSpiegel* beziehungsweise *#nichtmeinspiegel* hervor, die während des Bundestagswahlkampfes auch die Aufmerksamkeit von Medienbeobachtern erreichte. Die massenhafte Verwendung von *#ToxischeNarrative* war eine Reaktion auf die in der Einleitung erwähnte Studie der Amadeu-Antonio-Stiftung, die mehrere Hauptnarrative rechtspopulistischer Social-Media-Akteure identifizierte. Der Spiegel widmete sich in dieser Zeit mit dem Beitrag „Aufmarsch der Trolle“ der rechtsextremen Onlineaktivistengruppe *Reconquista Germanica* und der Beeinflussung von politischen Debatten in sozialen Medien (von Hammerstein et al. 2017). Sowohl die Studie als auch der Spiegel-Beitrag riefen in rechtspopulistischen



**Abb. 6** Themenverteilung der SSB im Zeitverlauf

Social-Media-Kreisen ein erhebliches Maß an negativer Resonanz hervor – die oben genannten Hashtags hatten auf Twitter binnen kurzer Zeit eine hohe Popularität. Das legt die Vermutung nahe, dass es sich bei der Verbreitung der Hashtags um gezielte und gesteuerte Maßnahmen rechtspopulistischer Aktivisten handelt (Hensel 2017).

Aufgrund der Erkenntnisse der quantitativen Betrachtung empfiehlt sich eine detailliertere Auseinandersetzung mit ausgewählten Accounts. Für diesen Zweck wurden alle 221 Super Social Bots auf Basis einer Stichprobe von jeweils bis zu 200 Tweets (sofern vorhanden) näher charakterisiert. Die manuelle inhaltliche Analyse, die gezielt nach thematischen Ähnlichkeiten und strukturellen Mustern sucht, ergab elf in sich homogene Nutzer-Gruppen, sogenannte *Content Cluster* (CC). Während die meisten dieser Cluster einen werblichen Charakter (Spam) aufweisen, bei denen die Tweets lediglich Links zu Online-Shops mit Trending Hashtags verknüpfen, oder simple News-Bots sind, die Nachrichtenmeldungen verbreiten, deckt die Analyse auch drei Cluster auf, die jeweils nahezu identische politische Botschaften verbreiten. Zur Beantwortung der Forschungsfrage, ob und in welchem Ausmaß politische Social Bots im Bundestagswahlkampf aktiv sind, folgt nun eine ausführliche Betrachtung dieser drei Cluster. Abb. 7 zeigt die Aktivitätsmerkmale der betrachteten Accounts.



**Abb. 7** Aktivitätsmerkmale ausgewählter Content Cluster der Super Social Bots



### 4.1.1 Content Cluster 2 – AfD

Content Cluster 2 besteht aus fünf Accounts, die AfD-nahe Inhalte teilen. Das Cluster weist weniger eine phrasenbasierte, als eine thematische Homogenität der Accounts auf. Alle User posten aber jeweils in sich ähnliche Tweets. User @MaasVoll verzeichnet die höchste Aktivität innerhalb des Clusters. Seine Tweets weisen eine sehr ähnliche Struktur auf. Sie beginnen meistens mit „BUNDESTAGSWAHL 2017 !!!“, gefolgt von nahezu identischen Wahlraufrufen für die AfD oder rechtspopulistischen Äußerungen. Gelegentlich werden Meldungen aus (neo-konservativen) Medienkreisen geteilt. User @TorstenDonners kombiniert AfD-Wahlwerbung mit typischen AfD-nahen sowie wahlkampfspezifische Hashtags (*#nichtmeinekanzlerin*, *#afd*, *#wahlbeobachter*). Massenhaft identisch verbreitete, inhaltlich sinnlose und meist fehlerbehaftete Phrasen (s. Tab. 2) legen den Verdacht nahe, dass es sich hier um einen selbstständig agierenden Bot handelt. Da es jedoch unter dem Pseudonym Torsten Donnerstein auch einen gleichnamigen YouTube-Kanal<sup>15</sup> gibt, der rechtspopulistische, verschwörungstheoretische und weltuntergangsbeschwörende Videobeiträge erstellt, liegt wahrscheinlich ein zumindest teilautomatisierter Account vor.

Ein ähnliches Twitter-Verhalten besitzt User @siedka. Auch dieser verwendet nur wenige grundlegende Formulierungen, die aber zumindest einen Link zu anderen Tweets regierungskritischer, rechtspopulistischer Accounts beinhalten. Neben dieser vermeintlich automatisierten Verbreitung von Tweets findet gelegentlich auch eine höchstwahrscheinlich menschlich betriebene Interaktion mit Nutzern in Form von Kommentaren und Gegenfragen statt, weshalb es sich vermutlich auch hier um einen größtenteils automatisierten Account handelt. Nutzer @AlfVogel verbreitet noch simpler strukturierte Tweets, die ein nahezu identisches Repertoire an gleichen Hashtags in Verbindung mit anderen (vermutlich zu diesem Zeitpunkt populären) Hashtags beinhalten.

Während die anderen Accounts des Clusters eher wahl- bzw. ziellos politische Parolen verbreiten, übernimmt @HaraldRostock eine interaktivere Rolle und verknüpft AfD-Wahlwerbung und regierungskritische Botschaften mit Mentions von Nutzern. Die meisten der Tweets enthalten Links zu einer Webseite<sup>16</sup>, die rechtspopulistische und verschwörungstheoretische Inhalte verbreitet. Dass @HaraldRostock in manchen Tweets auch den Account der AfD-Politikerin Beatrix von Storch mit Wahlwerbung anspricht, verdeutlicht zugleich, dass die Auswahl

---

15 <https://www.youtube.com/user/Donnerstein88>

16 <https://tagebuch-ht.weebly.com>



zeigten ganz offen ihren Parteibezug. Die Profile teilten alle sehr ähnliche und kaum variierende Inhalte, weshalb eine detaillierte Betrachtung jedes einzelnen Accounts überflüssig ist. Es weichen jedoch drei Accounts von der Nomenklatur ab: Mit @ClaasOsterloh und @osterlohmichael nahmen überraschenderweise zwei offizielle Accounts der niedersächsischen Landespolitiker Claas und Michael Osterloh bei der massenhaften Verbreitung identischer Tweets teil. @mautfrei\_EU diente als themenspezifischer Account gegen die geplante PKW-Maut, verbreitete aber (trotzdem) die gleichen Inhalte wie die anderen Nutzer des Clusters.

**Tab. 2** Ausgewählte typische Tweets der betrachteten SSBs

Cluster	Nutzer	Typischer Tweet
CC2	AlfVogel	#merkelmussweg #darumAfD wählen #noch15Tage bis zur Wahl #NichtmeineKanzlerin Merkel ist (Yoda ! ) :D #GehWählen #AfD bitte #btw17
CC2	AlfVogel	#afd #gehwählen #nichtmeinekanzlerin #NichtMeinSpiegel #LügenSpiegel – #S04VfB – #BSCSVW – #Lindenstrasse – #EsIstZeit #noch14tage Auf Auf
CC2	HaraldRostock	@janine_pieper @AfD Neues aus der Anstalt <a href="https://t.co/ol8Nb95hyv">https://t.co/ol8Nb95hyv</a> ----- <a href="https://t.co/50mD-WJnyBy">https://t.co/50mD-WJnyBy</a>
CC2	HaraldRostock	@Beatrix_vStorch Warum ich die AfD wähle ? Die Antwort finden Sie hier – HP : – <a href="https://t.co/ol8Nb95hyv">https://t.co/ol8Nb95hyv</a>
CC2	MaasVoll	BUNDESTAGSWAHL 2017 !!! ES GIBT NUR EINE „ALTER-NATIVE“ ZUR DERZEITIGEN REGIERUNG IN DEUTSCHLAND !!!
CC2	MaasVoll	Bundestagswahl 2017 !!! Deutschland braucht dringender denn je eine „Alternative“ zur derzeitigen Regierung !!!!!
CC2	siedka	Wählt AFD, es ist kein Platz für Rechtsbeuger wie die CDU und Spiegel Online Propaganda Verein #gehwahlen #Nicht-MeinSpiegel #AfD #annewill
CC2	siedka	Nur die AFD kann uns retten. Es ist unfassbar was diese Gaukler uns vormachen.
CC2	TorstenDonners	hab den #LügenSpiegel satt den er ist #NichtMeinSpiegel, #NichtMeinSpiegel ja weil #LügenSpiegel – deshalb #Nicht-MeinSpiegel, #GehWählen #afd
CC2	TorstenDonners	@AfD @blaubeereis07 #gehwahlen bei der #btw17, die #AfD, – #noch6Tage dann bin ich #Wahlbeobachter denn es ist nötig #darumAfD

Cluster	Nutzer	Typischer Tweet
CC9	FWniCelle	Familien ohne Existenzangst & berufliche Nachteile. #FREI-EWÄHLER eben Die #anstaendigeAlternative <a href="https://t.co/TveLd7CqUb">https://t.co/TveLd7CqUb</a> #btw17 #ltwns
CC9	FWniGoettingen	Die Bundestagswahl wird eine Richtungsentscheidung für die Deutschland. #FREIEWÄHLER #anstaendigeAlternative <a href="https://t.co/mqsraAeABT">https://t.co/mqsraAeABT</a>
CC9	FWniWolfsburg	Bildung von Anfang an. #FREIEWÄHLER eben Die #anständigeAlternative <a href="https://t.co/hK2pma6gRF">https://t.co/hK2pma6gRF</a> #btw17 #ltwns
CC11	ARCHITEKTEN-HAUS	#ARCHITEKTENHAUS #Bundestagswahl 24.09.2017: Warum ich als #Christ die #AfD mit gutem Gewissen #wähle! <a href="https://t.co/YpxD9PZ1zT">https://t.co/YpxD9PZ1zT</a>
CC11	RfD_BRD	Das #Geheimnis der Grundlage #Flüchtlingskrise von #Bundeskanzlerin Frau Dr. #Merkel ist gelüftet: <a href="https://t.co/1MM-GUAh7uq">https://t.co/1MM-GUAh7uq</a>
CC11	WISSEN Agentur	#Massenmigration nach #Europa muss sofort #gestoppt werden! Oder #Buergerkrieg & #Religionskrieg: <a href="https://t.co/X9QuVHEwii">https://t.co/X9QuVHEwii</a>

#### 4.1.3 Content Cluster 11 – Rettung für Deutschland (RfD)

Content Cluster 11 besteht aus drei Accounts, die ähnliche rechtspopulistische, verschwörungstheoretische Inhalte verbreiten und einen Bezug zur politischen Gruppierung *Rettung für Deutschland (RfD)* aufweisen. Hinter dieser Gruppierung verbirgt sich Uwe Melzer, der unter anderem auch die Unternehmen *WISSEN Agentur* und *Öko-Architektenhaus* betreibt, die jeweils auf Twitter unter eigenem Account aktiv sind. Alle drei Accounts machen Werbung für die AfD und beinhalten meistens Links auf die eigene Webseite.

## 4.2 High-Performer

Die vier Accounts des Highperformer-Clusters sind typische News- beziehungsweise Retweet-Bots. Hier sticht vor allem die sehr hohe Tweet-Anzahl hervor, wohingegen der Nahduplikatanteil vergleichsweise gering ist, auch wenn sich der Aufbau der Tweets innerhalb der Accounts jeweils ähnelt. User @car\_detailer mit 16.357 Tweets teilt aktuelle Angebote mit Verweis auf die jeweilige Webshop-Seite. @DieNachrichten teilt Nachrichtenmeldungen, ähnlich wie @MGieselmann, der sich jedoch nicht nur auf Nachrichten beschränkt und zudem immer retweetet. @TooglesDE ist ein automatisierter Account des Unternehmens Toogles.AI und

verknüpft in seinen Tweets mehrere ähnlich thematische Hashtags mit Links zur eigenen Webseite, wie zum Beispiel „Deine Meinung zu #Wirtschaft + #Präsident + #Grüne + #Frankfurt + #Kanzlerin bildest du hier – – <https://t.co/OUavg5J3SD>“. Die Wordcloud in Abb. 9 stellt die häufigsten Begriffe innerhalb des Clusters dar. Insgesamt zeigt sich kein Bezug zur Bundestagswahl, weshalb dieses Cluster nicht näher betrachtet wird.



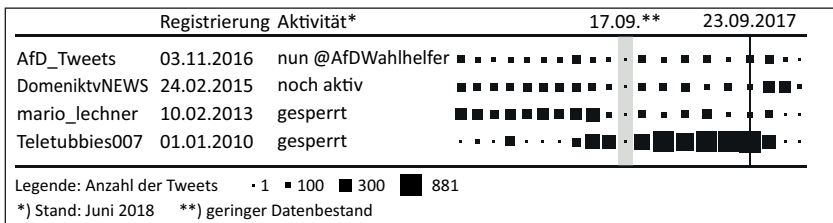
**Abb. 9** Wordcloud der häufigsten Begriffe der High-Performer Accounts

### 4.3 Copycats

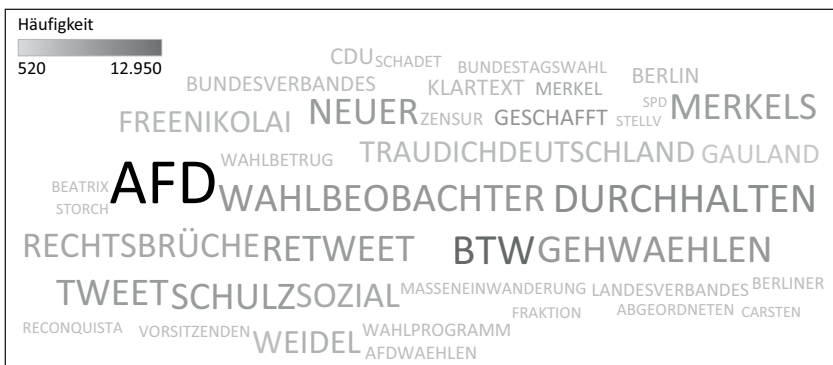
Das Cluster der „Copycats“ besteht aus vier Accounts, die ein ähnliches Aktivitätsmuster in Bezug auf Tweet-Häufigkeit und Nahduplikatanteil haben und überwiegend dem rechtspopulistischen, AfD-nahen Spektrum zugeordnet werden können. Nutzer @AfD\_Tweets ist ein klassischer Retweet-Bot, der Meldungen von AfD-Accounts (wie dem von Frauke Petry, Beatrix von Storch oder dem AfD-Bundesverband) unkommentiert als Verlinkung weiterleitet. Dabei sind 4.436 von 4.450 Tweets im Beobachtungszeitraum Nahduplikate.<sup>18</sup> Der Account @Teletubbies007 (5.630/5.786 ND) postet in nahezu allen Fällen den identischen Tweet „Durchhalten #btw17 #gewhaehlen #AfD #Wahlbeobachter #traudichDeutschland für #Weidel und #Gauland #FreeNikolai gegen #Zensur #Reconquista“. Die Berücksichtigung

<sup>18</sup> @AfD\_Tweets variiert nur den Accountnamen und Link. Da Hyperlinks von der Nahduplikatanalyse ausgeschlossen wurden, unterscheiden sich die Tweets nur im erwähnten Account-Namen des Retweets.

der Häufigkeitsverteilung in Abb. 10 zeigt zudem, dass der Account vor allem kurz vor der Bundestagswahl aktiv war. Mit einer durchschnittlichen Frequenz von zwei Tweets pro Minute zwischen 4 und 23 Uhr sowie zeitweise sieben Tweets pro Minute am Tag vor der Bundestagswahl, liegt es nahe, dass @Teletubbies007 ein Bot für den AfD-Wahlkampf ist. @mario\_lechner (4.527/5.767 ND) dient primär als Multiplikator des Autors Thorsten Schulte, der in den zwei verlinkten YouTube-Videos (in 45 % aller Tweets des Accounts) für sich und sein rechtspopulistisch, verschwörungstheoretisches und regierungskritisches Buch „Kontrollverlust“ wirbt. Ein typischer Tweet ist „@Ralf\_Stegner Merkels Rechtsbrüche <https://t.co/eRzqiDIRwb> Das nennt Schulz sozial! Lohnt sich!“. Einzig @DomeniktvNEWS (3.430/5.592 ND) weist keinen direkten inhaltlichen Bezug zur AfD auf und teilt dagegen breit gefächert Nachrichtenmeldungen verschiedener Medien. Die Wordcloud in Abb. 11 stellt die häufigsten Begriffe der vier Accounts dar und verdeutlicht den überwiegend rechtspopulistischen Bezug.



**Abb. 10** Aktivitätsmuster der Copycats



**Abb. 11** Wordcloud der häufigsten Begriffe der Copycat-Accounts

## 5 Zusammenfassung

Die vorliegenden Daten können keine Belege für automatisierte Twitter-Aktivitäten liefern, sondern nur Indizien. Ob beziehungsweise welche dieser Accounts nun in welchem Ausmaß einer algorithmisch basierten Steuerung unterliegen, kann nur vermutet werden.

Der Ansatz, über eine Nahduplikatanalyse Bots zu identifizieren, hat – wie andere Ansätze auch – Schwächen. Zum einen werden über diese Methode nur User oder Usergruppen mit sehr ähnlichen Tweets ermittelt, wodurch die Sensitivität (*Recall*) des Klassifikationsverfahrens eingeschränkt wird. Zum anderen kann bei den entsprechend als potentielle Bots identifizierten Accounts keine sichere Aussage darüber getroffen werden, ob es sich dabei tatsächlich um Bots oder aber nur um bot-ähnlich agierende Menschen handelt, worunter demnach der positive Vorhersagewert (*Precision*) des Verfahrens leidet. So könnten Accounts mit einer hohen Nahduplikatraten auch von Menschen betrieben werden, die bestimmte Tweets manuell kopieren und – wenn überhaupt – leicht abgewandelt teilen. Denkbar wären auch menschlich gesteuerte Accounts, die durch Programme teilautomatisiert wurden (*bot-assisted human*) beziehungsweise Bots, deren Aktivität durch Menschen mehr oder weniger überwacht und gelenkt werden (*human-assisted bot*). Man spricht in diesem Zusammenhang auch von *Cyborgs* (Chu et al. 2012). Das hier angewendete zweistufige Verfahren mit einer nachgelagerten manuellen qualitativen Analyse zufällig ausgewählter Tweets (und Berücksichtigung quantitativer Kennzahlen<sup>19</sup>) der vorher erkannten potentiellen Political Bots mindert das Risiko von falschen Bot-Klassifizierungen, jedoch besteht insgesamt eine gewisse Unsicherheit, ob und in welchem Umfang es sich bei den identifizierten Accounts um tatsächliche Bots handelt.

Demgegenüber ermöglicht dieses Verfahren eine textbasierte Erkennung inhaltlich interessanter Accounts im Hinblick auf politisch motivierte Beeinflussung bei moderatem Ressourcenverbrauch. Während rein quantitative Verfahren meist mit einem geringen Aufwand verbunden sind, dafür aber nur eine abstrakte Liste von Nutzern und deren Kennzahlen ausgeben (wie der *Bot Score* von *Botometer*<sup>20</sup>), haben rein qualitative Ansätze einen immensen Zeitbedarf für inhaltliche Analysen. Eine manuelle Betrachtung aller 21.097 Nutzer im hier vorliegenden Datensatz hätte Wochen gedauert. Das angewendete zweistufige Verfahren reduziert zunächst die Zahl „interessanter“ Accounts anhand quantitativer Parameter (Tweet-Häufigkeit

---

19 Wie Tweetaktivität pro Tag und Stunde oder Zeitpunkt der Accounterstellung.

20 <https://botometer.iuni.iu.edu/>

und Nahduplikatrate), wodurch die darauffolgende qualitative Betrachtung dieser überschaubaren Menge an Accounts weniger Zeit in Anspruch nimmt.

Die Ergebnisse zeigen, dass es im Vorfeld der Bundestagswahl trotz vorherigen Dementi durch die Partei AfD-bezogene, zumindest bot-ähnliche politische Aktivitäten auf Twitter gab. Inwieweit dieses Vorgehen durch die AfD selbst initiiert oder gesteuert wurde, bleibt offen. Möglich ist beispielsweise auch eine (un-) koordinierte Aktion politisch rechter Aktivisten und Sympathisanten oder Parteianhänger. Überraschend ist die offensichtliche Bot-Aktivität der Freien Wähler, die im Zuge der Bundestagswahl und der kurz darauffolgenden Landtagswahl in Niedersachsen mit einem klassischen Bot-Netzwerk aktiv war, das mit einer sehr simplen Systematik meist identische Tweets über mehrere Accounts verbreitete.

Trotz der gewiss vorhandenen methodischen Einschränkungen deckt die Nahduplikatanalyse deutliche Ähnlichkeiten zwischen den Inhalts- und Aktivitätsmustern mehrerer Accounts auf. So verwenden die Nutzer des AfD-Clusters sehr ähnliche und teilweise identische Formulierungen und Hashtags. Auch innerhalb der Super Social Bots finden sich mehrere Nutzgruppen, die im Sinne der AfD aktiv sind. Gleichzeitig ermittelt der Nahduplikat-Algorithmus auch Accounts, die offensichtlich keine Political Bots sind, wie simple Multiplikator-Accounts von Medienakteuren (z. B. der Frankfurter Allgemeinen Zeitung).

Zusammenfassend betrachtet ist die Methode der Nahduplikatanalyse im Hinblick auf die Bot-Identifikation weniger effektiv als beispielsweise ungleich komplexere Machine-Learning-Ansätze. Schließlich ignoriert eine rein textbasierte Erkennung von Bots deren andere Einflussmöglichkeiten wie die Einwirkung durch Interaktion oder die Erzeugung einer vermeintlichen Mehrheitsmeinung durch die massenweise Generierung von Retweets. Gleichzeitig liefert der Ansatz aber eine verbesserte Ausgangsposition zur Untersuchung von massenhaft geteilten Inhalten. Prinzipiell erscheint eine Kombination der Nahduplikatanalyse mit anderen Verfahren sinnvoll, um beispielsweise unterscheiden zu können, ob Bots lediglich als Multiplikatoren vorhandener Nachrichten oder ebenfalls als Initiatoren originärer Tweets fungieren.

Eine viel grundlegendere Frage ist jedoch, wie stark der Einfluss von Political Bots, auf Twitter oder anderen Social-Media-Kanälen, im Wahlkampf tatsächlich ist. Beispielsweise ist Twitter in Deutschland zweifelsohne ein Nischenmedium. Über das Teilen von Tweets und die nachgelagerte Verbreitung durch Massenmedien ist zwar eine nicht quantifizierbare Beeinflussung der Debatte in den Medien und der Öffentlichkeit möglich. Wie stark diese Debatte aber durch reine Bot-Aktivitäten beeinflusst werden kann, bleibt fraglich. Viele der identifizierten Political Bots haben nur eine geringe Anzahl an Followern. Auch wenn die tatsächliche Reichweite von Tweets, die über die Zahl der Favorites und Retweets hinausgeht, unbekannt ist,



bedarf es großer Bot-Netzwerke, um die Bedeutung und Relevanz von Themen und Meinungen auf Twitter zu manipulieren. Ein Spill-Over-Effekt auf die Medienberichterstattung und die gesamte Öffentlichkeit ist noch schwieriger und folglich unwahrscheinlicher. Diese Studie versucht die Auseinandersetzung mit Bots, weg von der rein mechanischen Identifikation von Accounts hin zur gleichzeitigen Herausarbeitung besonders prominenter und potentiell toxischer Narrative. Denn auch wenn die komplexen Verfahren der Bot-Erkennung sicherlich zur Eindämmung von Fake News beitragen können, ist die Kenntnis sowie die Auseinandersetzung mit den geteilten Inhalten nicht weniger bedeutsam.

## Literatur

- Abokhodair, N., Yoo, D., & McDonald, D. W. (2015). Dissecting a social botnet: Growth, content and influence in Twitter. *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing – CSCW '15*, 839–851. doi:10.1145/2675133.2675208.
- Alarifi, A., Alsaleh, M., & Al-Salman, A. (2016). Twitter turing test: Identifying social machines. *Information Sciences* 372, 332–346. <https://doi.org/10.1016/j.ins.2016.08.036>
- Amann, M., Knaup, H., Müller, A.-K., Rosenbach, M., & Wiedmann-Schmidt, W. (2016, 22. Oktober). Digitale Dreckschleudern. Wie „Social Bots“ den politischen Diskurs manipulieren. *Der Spiegel*, S. 44.
- Baldauf, J., Dittrich, M., Hermann, M., Kollberg, B., Lüdecke, R., & Rahthje, J. (2017). *Toxische Narrative. Monitoring rechts-alternativer Akteure*. Berlin: Amadeu Antonio Stiftung.
- Bastos, M. T., & Mercea, D. (2017). The Brexit botnet and user-generated hyperpartisan news. *Social Science Computer Review*. doi:10.1177/0894439317734157
- Bessi, A., & Ferrara, E. (2016). Social bots distort the 2016 U.S. Presidential election online discussion. *First Monday* 21(11). doi:10.5210/fm.v21i11.7090
- Boshmaf, Y., Muslukhov, I., Beznosov, K., & Ripeanu, M. (2013). Design and analysis of a social botnet. *Computer Networks* 57(2), 556–578.
- Brachten, F., Stieglitz, S., Hofeditz, L., Kloppenborg, K., & Reimann, A. (2017). Strategies and influence of social bots in a 2017 German state election – A case study on Twitter. *arXiv:1710.07562*.
- Cai, C., Li, L., & Zengi, D. (2017). Behavior enhanced deep bot detection in social media. 2017 *IEEE International Conference on Intelligence and Security Informatics (ISI)*: 128–130. Peking, China. doi:10.1109/ISI.2017.8004887.
- Chu, Z., Gianvecchio, S., Wang, H., & Jajodia, S. (2012). Detecting automation of Twitter accounts: Are You a Human, Bot, or Cyborg? *IEEE Transactions on Dependable and Secure Computing*, 9(6): 811–824. doi:10.1109/TDSC.2012.75.
- Cialdini, R. B. (2013). *Die Psychologie des Überzeugens*. Bern: Huber.
- Davis, C. A., Varol, O., Ferrara, E., Flammini, A., & Menczer, F. (2016). BotOrNot: A system to evaluate social bots. *Proceedings of the 25th International Conference Companion on World Wide Web*: 273–274. arXiv:1602.00975

- Edwards, C., Edwards, A., Spence, P.R., & Shelton, A.K. (2014). Is that a bot running the social media feed? Testing the differences in perceptions of communication quality for a human agent and a bot agent on Twitter. *Computers in Human Behavior* 33, 372–376.
- Everett, R.M., Nurse, J.R. C., & Erola, A. (2016). The anatomy of online deception: What makes automated text convincing? *Proceedings of the 31st Annual ACM Symposium on Applied Computing*, Pisa/Italien, 1115–1120. doi: 10.1145/2851613.2851813.
- Ferrara, E. (2017). Disinformation and social bot operations in the run up to the 2017 French presidential election. *SSRN Electronic Journal*. doi:10.2139/ssrn.2995809.
- Ferrara, E., Varol, O., Davis, C., Menczer, F., & Flammini, A. (2016). The rise of social bots. *Communications of the ACM* 59(7), 96–104.
- Forelle, M. C., Howard, P. N., Monroy-Hernandez, A., & Savage, S. (2015). Political bots and the manipulation of public opinion in Venezuela. *SSRN Electronic Journal*. doi:10.2139/ssrn.2635800.
- Freitas, C., Benevenuto, F., Veloso, A., & Ghosh, S. (2016). An empirical study of socialbot infiltration strategies in the Twitter social network. *Social Network Analysis and Mining* 6, 23.
- Graber, R., & Lindemann, T. (2018). Neue Propaganda im Internet. Social Bots und das Prinzip sozialer Bewährtheit als Instrumente der Propaganda. In K. Sachs-Hombach & B. Zywiets (Hrsg.), *Fake News, Hashtags & Social Bots: Neue Methoden populistischer Propaganda* (S. 51–68). Wiesbaden: Springer Fachmedien.
- Grimme, C., Assenmacher, D., & Adam, L. (2018). Changing perspectives: Is it sufficient to detect social bots? In: G. Meiselwitz (Hrsg.), *Social computing and social media. User experience and behavior. SCSM 2018. Lecture Notes in Computer Science* (S. 445–461). Cham: Springer.
- Grimme, C., Assenmacher, D., Preuss, M., Adam, L., & Lütke Stockdiek, J. (2017). Bundestagswahl 2017: Social-Media-Angriff auf das #kanzlerduell? PropStop Technischer Bericht. <http://www.propstop.de/wp-content/uploads/2017/09/bundestagswahl-2017-social-media.pdf>. Zugegriffen: 14. Juli 2018.
- Guo, D., & Chen, C. (2014). Detecting non-personal and spam users on geo-tagged Twitter network. *Transactions in GIS* 18(3), 370–384.
- Hegelich, S., & Janetzko, D. (2016). *Are social bots on Twitter political actors? Empirical evidence from a Ukrainian social botnet*. Vortrag bei der Tenth International AAAI Conference on Web and Social Media, ICWSM 2016. Köln, 17.-20. Mai 2016.
- Hensel, G. (2017, 11. September). Rechte Bots im Bundestagswahlkampf: Vor 7 Tagen ging's los. *Fearless Democracy*. <https://fearlessdemocracy.org/btw-2017-bots/>. Zugegriffen: 07. Juli 2018.
- Hölig, S., & Hasebrink, U. (2016). Nachrichtennutzung über soziale Medien im internationalen Vergleich. *Media Perspektiven* (11), 534–548.
- Howard, P.N., & Kollanyi, B. (2016). Bots, #Strongerin, and #Brexit: Computational propaganda during the UK-EU referendum. *SSRN Electronic Journal*. doi:10.2139/ssrn.2798311
- Kollanyi, B., Howard, P., & Woolley, S. (2016). Bots and automation over Twitter during the first U.S. Presidential debate: COMPROP Data Memo 2016.1. – <http://blogs.oii.ox.ac.uk/politicalbots/wp-content/uploads/sites/89/2016/10/Data-Memo-First-Presidential-Debate.pdf>. Zugegriffen: 28. Juni 2018.
- Lokot, T., & Diakopoulos, N. (2015). News bots. *Digital Journalism* 4(6), 682–699.
- Morstatter, F., Pfeffer, J., Liu, H., & Carley, K.M. (2013). Is the sample good enough? Comparing data from Twitter's streaming API with Twitter's Firehose. *arXiv:1306.5204*

- Neudert, L.-M., Kollanyi, B., & Howard, P.N. (2017a). Junk news and bots during the German Federal Presidency election: What were German voters sharing over Twitter? *COMPROP DATA MEMO 2017.2*. <http://blogs.oii.ox.ac.uk/politicalbots/wp-content/uploads/sites/89/2017/03/What-Were-German-Voters-Sharing-Over-Twitter-v6-1.pdf>. Zugegriffen: 26. Juni 2018. –
- Neudert, L.-M., Kollanyi, B., & Howard, P.N. (2017b). Junk news and bots during the German parliamentary election: What are German voters sharing over Twitter. *COMPROP DATA MEMO 2017.7*. [http://comprop.oii.ox.ac.uk/wp-content/uploads/sites/93/2017/09/ComProp\\_GermanElections\\_Sep2017v5.pdf](http://comprop.oii.ox.ac.uk/wp-content/uploads/sites/93/2017/09/ComProp_GermanElections_Sep2017v5.pdf). Zugegriffen: 25. Juni 2018.
- Proisl, T. & Uhrig, P. (2016). SoMaJo: State-of-the-art tokenization for German web and social media texts. In: *Proceedings of the 10th Web as Corpus Workshop (WAC-X) and the EmpirIST Shared Task*, 57–62. Berlin: Association for Computational Linguistics.
- Pfaffenberger, F. (2018). What you tweet is what we get? *Publizistik* 63(1), 53–72.
- Ratkiewicz, J., Conover, M.D., Meiss, M., Goncalves, B., Flammini, A., & Menczer, F.M. (2011). Detecting and tracking political abuse in social media. *Proceedings of the ICWSM 11*, 297–304.
- Schäfer, F., Evert, S., & Heinrich, P. (2017). Japan's 2014 General Election: Political bots, right-wing internet activism, and Prime Minister Shinzō Abe's hidden nationalist agenda. *Big Data* 5(4), 294–309.
- Schmehl, K. (2017, 3. September). Diese geheimen Chats zeigen, wer hinter dem Mem-Angriff #Verräterduell aufs TV-Duell steckt. *BuzzFeed*. <https://www.buzzfeed.com/de/karstensmehl/willkommen-in-der-welt-von-discord-teil1>. Zugegriffen: 07. Juli 2018.
- Stieglitz, S., Brachten, F., Ross, B., & Jung, A.-., Katharina. (2017). Do social bots dream of electric sheep? A categorisation of social media bot accounts. arXiv:1710.04044.
- Timberg, C., & Dwoskin, E. (2017, 30. Oktober). Russian content on Facebook, Google and Twitter reached far more users than companies first disclosed, congressional testimony says. *The Washington Post*. [https://www.washingtonpost.com/business/technology/2017/10/30/4509587e-bd84-11e7-97d9-bdab5a0ab381\\_story.html?utm\\_term=.43b78783d13e](https://www.washingtonpost.com/business/technology/2017/10/30/4509587e-bd84-11e7-97d9-bdab5a0ab381_story.html?utm_term=.43b78783d13e). Zugegriffen: 10. Juli 2018.
- Timberg, C., & Dwoskin, E. (2018, 6. Juli). Twitter is sweeping out fake accounts like never before, putting user growth at risk. *The Washington Post*. <https://www.washingtonpost.com/technology/2018/07/06/twitter-is-sweeping-out-fake-accounts-like-never-before-putting-user-growth-risk/>. Zugegriffen: 18. Juli 2018.
- Varol, O., Ferrara, E., Davis, C. A., Menczer, F., & Flammini, A. (2017). *Online human-bot interactions: Detection, estimation, and characterization*. arXiv:1703.03107
- von Hammerstein, K., Höfner, R., & Rosenbach, M. (2017, 12. September). Wahlkampf mit Fake-News. Wie die Rechtsextremen mit Hetze mobil machen. *Spiegel Online*. <http://www.spiegel.de/spiegel/wahlkampf-mit-trollen-wie-die-rechten-im-netz-mobil-machen-a-1167063.html>. Zugegriffen: 18. Juli 2018.
- Wilke, T. (2017). Social Bots – die unberechenbare Armee im Wahlkampf. In J. Böttger, R. Güldenpopp, & M. Voigt (Hrsg.), *Wahlanalyse 2017: Strategie, Kampagne, Bedeutung* (S. 88–92). Berlin: epubli.

## Anhang

### Keyword-Liste

Abrüstung	Energiewende	KITA	Röttgen
Adoption	Erneuerbare	Klima	Russland
AFD	EU	Klimaschutz	Schäuble
Altersarmut	Euro	Kraft	Schuldenbremse
Altmaier	Europa	Krankenversicherung	Schulz
Arbeitslosengeld	Euroraum	Kretschmann	Seehofer
Arbeitslosigkeit	Familien	Kriminalität	Sicherheit
Asyl	Familienarbeitszeit	Landwirtschaft	Sicherheitspolitik
Atomausstieg	FDP	Leitkultur	Sozialdemokraten
Außengrenze	Flucht	Lindner	SPD
Außenpolitik	Flüchtling	Linke	Spitzensteuer
BAföG	Flüchtlingspolitik	Lohnausgleich	Staatsangehörigkeit
Bankenunion	Fluchtursachen	Lohngleichstellung	Staatsbürgerschaft
Befristete Arbeitsverträge	Frauenquote	Massentierhaltung	Studiengebühren
Betreuungsgeld	Freihandelsabkommen	Menschenrechte	Terror
Bildung	Gabriel	Merkel	Türkei
Bundestagswahl	Ganztage	Migration	Union
Bündnis 90	Gauland	Mindestlohn	Vermögenssteuer
Bürgerversicherung	Gerechtigkeit	Mindestquote	Volksabstimmung
Bürokratie	Gleichstellung	Mindestrente	Vorratsdatenspeicherung
Cannabis	Grundeinkommen	Mütterrente	Waffenexporte
CDU	Grundsicherung	Naturschutz	Wagenknecht
CSU	Grünen	Netzpoltik	Wehrpflicht
Datenschutz	Handelsabkommen	Obergrenze	Wirtschaft
Digitalisierung	Hartz	Özdemir	Wohlstand
Doppelpass	Höcke	Peter	Wohnraum
Drogenpolitik	Homoehe	Petry	Wohnungsbau
Ehe-für-Alle	Integration	Quoten	Zuwanderung
Einwanderung	Investitionen	Rente	
E-Mobilität	Kinderarmut	Rentenniveau	
Energie	Kindergeld	Riexinger	