Die überschätzte Gefahr? Twitter-Bots im Europawahlkampf 2019

Fabian Pfaffenberger und Philipp Heinrich

1 Einleitung

Zwischen der „Realität digitaler Öffentlichkeit“ (Klinger 2019) und „The army that never existed“ (Kreil 2019) – in der medial geführten Debatte diskutieren nicht nur Journalistinnen und Journalisten, sondern auch Wissenschaft und Politik über „das Problem mit den Social Bots“ (Gensing 2020). Social Bots sind durch Computerprogramme gesteuerte Accounts, die eine menschliche Identität vortäuschen, wie Menschen im Internet kommunizieren und so auch für manipulative Zwecke eingesetzt werden können (Kind et al. 2017). Die öffentliche Debatte behandelt dabei einerseits die von Social Bots ausgehende Gefahr durch deren Verbreitung von Fake News, Beeinflussung von Meinungen oder Diffamierung politischer Gegner – und stellt andererseits auch elementare Fragen. Besonders die jüngste Diskussion zwischen dem Politikwissenschaftler Simon Hegelich und dem Medieninformatiker Florian Gallwitz sowie Datenanalyst Michael Kreil im Kontext einer Expertenbefragung für die *Enquete-Kommission KI des Deutschen Bundestages* verdeutlicht die Uneinigkeit der

F. Pfaffenberger (✉)
Lehrstuhl für Kommunikationswissenschaft, Friedrich-Alexander-Universität Erlangen-Nürnberg, Nürnberg, Deutschland
E-Mail: fabian.pfaffenberger@fau.de

P. Heinrich
Lehrstuhl für Korpus- und Computerlinguistik, Friedrich-Alexander-Universität Erlangen-Nürnberg, Erlangen, Deutschland
E-Mail: philipp.heinrich@fau.de

C. Holtz-Bacha (Hrsg.), *Europawahlkampf 2019,*
https://doi.org/10.1007/978-3-658-31472-9_4

Wissenschaft über die Relevanz oder gar Existenz von Social Bots im politischen Kontext (Gallwitz 2020a, b; Hegelich 2020a, b).

Vor allem in Bezug auf die intensiv geführte Debatte um den Einsatz von Social Bots zur Beeinflussung von Debatten, zum Agenda Setting oder zur Verbreitung von Spam und Fake News in sozialen Medien überwiegt häufig die negative Konnotation des Begriffs *Bot.* Wenngleich Social Bots populäre Anwendungsfälle sind und einer entsprechend intensiven wissenschaftlichen Betrachtung unterliegen, bleibt die eigentliche Multifunktionalität von Bots im Allgemeinen oft unbeachtet. Denn es gibt vielseitige Einsatzmöglichkeiten für Bots, die nicht moralisch oder ethisch fragwürdigen Absichten unterliegen: Gerade auf Twitter nutzen viele Nachrichtenunternehmen Bots, die alle oder parametrisch selektierte Meldungen (zum Beispiel Eilmeldungen oder Schlagzeilen aus dem Wirtschaftsressort) automatisiert teilen und auf die eigentliche Nachrichtenmeldung verlinken (BBC News Labs 2019). Ein weiteres Einsatzgebiet von Bots im Journalismus ist der sogenannte *Robojournalism* (Lokot und Diakopoulos 2016). Hier erfolgt eine (Teil-)Automatisierung der Nachrichten-Berichterstattung durch *News Bots,* die beispielsweise im Lokalsport einfache, vorstrukturierte Informationen zu kurzen Nachrichten-Meldungen verarbeiten. Auch die automatische Verbreitung von Informationen durch Bots unterliegt, wenn sie wie @*sentinel_bot* beispielsweise Satellitenbilder eines ESA-Satelliten teilen, keinen schlechten Absichten, sondern dient der Informationsverbreitung.

Zweifelsohne gibt es aber eine Vielzahl an ethisch, rechtlich oder zumindest moralisch fragwürdigen Anwendungsmöglichkeiten für Bots. Der Einsatz von Social Bots zur Beeinflussung von Meinungen oder zur Verbreitung von Fake News ist dabei zwar nur einer von vielen Anwendungsfällen, steht jedoch bei wissenschaftlichen und medialen Betrachtungen im Mittelpunkt. Im Vorfeld der Europawahl 2019 gab es wie vor der Bundestagswahl 2017 Diskussionen um das Risiko einer Einflussnahme auf das Wahlverhalten (Graff 2017; Tagesschau 2019) beziehungsweise dessen Relativierung (Lypp 2017; Reuter 2019). Insgesamt stellt sich jedoch zweifellos die Frage, in welchem Ausmaß automatisierte Accounts auf Twitter aktiv sind und ob es überhaupt nennenswerte Aktivitäten im vielfach diskutierten politischen Kontext gibt.

Diese Untersuchung möchte zu der bereits ausgedehnten Debatte beitragen und analysieren, ob es eine nennenswerte Aktivität automatisierter Twitter-Accounts im Zusammenhang mit der Europawahl 2019 gab. Hierfür wurden während der heißen Phase des Europawahlkampfs relevante Tweets gesammelt und anhand eines mehrstufigen Identifikations- und Validierungsverfahrens auf Bot-ähnliche Aktivitäten analysiert. Die Untersuchung stützt sich dabei nicht,

wie viele andere Studien zur Bot-Klassifikation, auf rein parameterbasierte Verfahren, sondern nutzt einen inhaltsbasierten Nahduplikat-Erkennungsalgorithmus zur Identifikation auffälliger Twitter-Accounts (mit einem hohen Anteil an Nahduplikaten). Diese vorselektierten Accounts werden im zweiten Schritt einer manuellen qualitativen Analyse unterzogen. Hiermit soll überprüft werden, ob es auffällige Accounts mit einer hohen Nahduplikatrate im Datensatz gibt, die politische Inhalte verbreiteten, und wie groß deren Anteil am Gesamtdatensatz ist. Die Betrachtung über Nahduplikate und deren Verteilung erlaubt jedoch nicht nur die Ermittlung auffälliger Accounts, sondern auch Aussagen über die Verteilung ähnlicher Botschaften. So lassen sich auch Account-Gruppen identifizieren, die jeweils exakte Duplikate oder leichte Abwandlungen eines Tweets verbreiten. Deshalb soll auch untersucht werden, ob es Account-Cluster gibt, die ähnliche Botschaften teilten.

2 Social Bots – Definitionen und Klassifikationen

Verschiedene Arbeiten befassten sich bereits mit einer Übersicht und Kategorisierung von Arten, Eigenschaften und Einsatzmöglichkeiten von Bots (vgl. u. a. Gorwa und Guilbeault 2018; Pieterson et al. 2017; Stieglitz et al. 2017b). Während anfangs vor allem klassische *Spambots* untersucht wurden, erfahren mittlerweile vor allem *Social Bots* sowie sogenannte *Sockpuppets* und *Trolle* eine große Aufmerksamkeit in der Wissenschaft. Die Begriffsdefinition und -interpretation von Social Bots variiert mit dem Einsatzzweck: Der in der Frühphase der Bot-Forschung verwendete Terminus *Socialbot* beschreibt automatisierte Konten, die eine gefälschte Identität vortäuschen, reale Netzwerke von Nutzern infiltrieren und bösartige Links oder Werbung verbreiten (Boshmaf et al. 2011). Oftmals spricht man hier auch von *Sybils* (Alarifi et al. 2016, S. 1). Der Begriff *Social Bot* (mit zwei Wörtern) ist wiederum ein breiteres und flexibleres Konzept und umfasst Programme, die automatisch Inhalte produzieren, mit Menschen in sozialen Medien interagieren und versuchen, ihr Verhalten nachzuahmen und möglicherweise zu ändern (Ferrara et al. 2016, S. 2). Die Nachahmung menschlicher Identitäten und Aktivitäten spielt dabei eine zunehmende Rolle (Abokhodair et al. 2015, S. 13; Hegelich und Janetzko 2016, S. 582; Stieglitz et al. 2017a, S. 381), auch um aktiv auf die öffentliche Gewichtung und Bewertung von Themen Einfluss zu nehmen (Graber und Lindemann 2018).

Der Begriff *Sockpuppet* (oder Sockenpuppe) ist ein weiterer, ähnlicher Terminus im Kontext von Bots. Er wird häufig bei Accounts mit gefälschter Identität verwendet, die unter dieser Tarnung mit anderen Nutzern in sozialen

Netzwerken agieren. Dabei umfasst der Begriff sowohl automatisierte als auch durch Menschen gesteuerte Accounts (Bastos und Mercea 2019, S. 2). Bei politisch motivierten Sockpuppets spricht man häufig von *Trollen,* insbesondere, wenn sie von der Politik oder miteinander verbundenen Akteuren koordiniert werden (Gorwa und Guilbeault 2018, S. 233).

Insgesamt verschwimmen die Grenzen zwischen Bots und von Menschen gesteuerten Accounts zunehmend. Begriffe wie *Cyborgs* und *Hybrid Accounts* umschreiben Nutzerkonten, die sich im Kontinuum zwischen *bot-assisted humans* und *human-assisted bots* (Chu et al. 2012, S. 811) bewegen. Von bot-assisted humans spricht man bei Menschen, die ihren Twitter-Account mit Unterstützung von Programmen nutzen, um beispielsweise Nachrichten eines bestimmten RSS-Feeds automatisch zu teilen. Human-assisted bots sind wiederum Algorithmen, die durch Menschen gesteuert werden und von regelmäßigen Befehlen oder Parametereingaben abhängig sind. Ein Beispiel solcher Teil-Automatisierung wäre ein Bot, der populäre Hashtags identifiziert, aus denen eine steuernde Person geeignete auswählt und der Bot diese wiederum als Grundlage für automatisch generierte und verbreitete Botschaften nutzt. Oftmals ist eine genaue Klassifizierung von Accounts schwer: Die Frage, ob beispielsweise 500 Tweets eines Accounts pro Tag auf eine Automatisierung hinweisen oder ob sich dahinter ein sehr aktiver User verbirgt oder dieser Mensch Hilfsprogramme zur Steuerung des Accounts nutzt, kann selten auf den ersten Blick beantwortet werden. Dies wirkt sich offensichtlich auch negativ auf die Qualität und Zuverlässigkeit von automatischen Bot-Erkennungsmethoden aus.

Die Multidimensionalität von Bots, nicht nur in ihren Funktionen, sondern auch im Grad der Automatisierung, schlägt sich ebenso in der Vielzahl und Heterogenität der Ansätze zur Bot-Identifikation nieder. Dabei reicht die Bandbreite von einfachen, quantitativen Betrachtungen auf Basis einzelner Indikatoren bis hin zu komplexen, auf *Machine Learning* basierenden Klassifikationsalgorithmen. Ein Beispiel simpler, quantitativer Bot-Klassifizierung ist die so populäre wie umstrittene „Oxford-Regel“, die Accounts mit mehr als 50 Tweets am Tag pauschal als Bots definiert (Howard und Kollanyi 2016, S. 4). Andere Studien basieren wiederum auf komplexeren Algorithmen, die anhand verschiedener Parameter (Social) Bots (Ahmed und Abulaish 2013; Loyola-Gonzalez et al. 2019; Miller et al. 2014) oder Sockpuppets (Bu et al. 2013) klassifizieren. Eine Vielzahl der Ansätze nutzt jedoch Machine-Learning-Algorithmen (Alarifi et al. 2016; Cai et al. 2017; Chu et al. 2012; Davis et al. 2016; Daya et al. 2019; Ratkiewicz et al. 2011; van der Walt und Eloff 2018; Varol et al. 2017; Yang et al. 2019), die zumeist mehrere Merkmale („Features“) von Accounts sowie deren Netzwerk und Tweets analysieren und gewichten, in einen zeitlichen

Kontext setzen und darauf basierend beispielsweise die Bot-Wahrscheinlichkeit ermitteln.

Der überwiegende Teil der Ansätze zur Bot-Erkennnung stützt sich (rein) auf parameterbasierte Analysen wogegen der tatsächliche Inhalt eines Tweets oftmals außen vor bleibt beziehungsweise die Analyse selten über die Häufigkeit bestimmter Begriffe (zum Beispiel auf Basis von Part-of-Speech-Tags), die Tweet-Länge oder den Informationsgehalt von Nachrichten hinausgeht (Alarifi et al. 2016; Varol et al. 2017). Eine Fokussierung auf rein quantitative Merkmale kann vor allem bei sehr großen oder Ad-hoc-Analysen sehr hilfreich sein, da diese meist schneller und ohne menschliche Unterstützung erfolgen können. Dennoch kann vor allem der geteilte Inhalt von besonderem Interesse sein – möchte man nicht nur Bots erkennen, sondern auch deren verbreitete Botschaften analysieren. Zudem besteht über den Tweet-Inhalt eine weitere Möglichkeit/Dimension der Bot-Identifikation: Während viele Ansätze Accounts nur anhand ihrer individuellen Statistiken (z. B. Tweet-Häufigkeiten, Follower-Anzahl) und somit einzeln klassifizieren, ermöglichen inhaltsbasierte Analysen die Betrachtung ganzer Account-Gruppen, die zum Beispiel ähnliche Inhalte teilen und bei einer Einzelbetrachtung unauffällig gewesen wären. So lassen sich auch Accounts aufspüren, die als koordiniertes Netzwerk/Cluster agieren und erst dadurch als Bots in Erscheinung treten. Denn allein durch das ständige Wiederholen bestimmter Botschaften können einerseits Gegenmeinungen in der schieren Masse untergehen oder deren Urheber eingeschüchtert werden (Lypp 2017). Andererseits können Account-Netzwerke auch als Verstärker für Botschaften agieren und eine Graswurzelbewegung vortäuschen (Woolley 2016).

Die hier vorliegende Untersuchung basiert daher auf einem korpuslinguistischen Ansatz von Schäfer et al. (2017) zur Identifikation auffälliger Accounts. Die zugrundeliegende Methodik stützt sich auf die Ermittlung von Nahduplikaten (*Near Duplicates,* Tweets mit identischem oder nahezu identischem Inhalt), verknüpft mit der Annahme, dass Nutzer, die häufig oder überwiegend ähnliche Inhalte verbreiten, dies mit hoher Wahrscheinlichkeit automatisiert machen. Dieser Ansatz wurde bereits in einer früheren Studie zur Bundestagswahl 2017 eingesetzt und durch eine manuelle Analyse der auffälligen Accounts erweitert (Pfaffenberger et al. 2019). Die damalige Untersuchung fokussierte sich jedoch nur auf Accounts, die einen hohen Anteil an Nahduplikaten aufweisen. Um auch Gruppen von Accounts zu ermitteln, die zwar einen geringen Nahduplikatanteil aufweisen und bei individueller Betrachtung unverdächtig erscheinen, jedoch als Gruppe ähnliche Botschaften teilen, betrachtet die hier vorliegende Untersuchung auch Nahduplikat-Cluster.

3 Methodik

Bei der Identifikation und Analyse von (Social) Bots auf Grundlage eines Twitter-Datensatzes ist die Methode der Datenerhebung entscheidend. Es stehen grundsätzliche zwei kostenlose Schnittstellen zur Verfügung: Die *Streaming API* und die *REST APIs.* In diesem Fall eignet sich nur die Echtzeit-Datensammlung mittels der *Twitter Streaming API* (statuses/filter), auch wenn hier das Risiko der Deckelung des Datenstroms durch Rate Limits besteht. Bei der Ex-post-Datensammlung über die REST APIs wäre es jedoch wahrscheinlich, dass in der Zwischenzeit durch Twitter gelöschte oder gesperrte Tweets/Accounts im Datensatz fehlen würden. Da diese Studie Bots identifizieren und analysieren möchte, sind genau diese Accounts/Tweets von besonderem Interesse.

Der Analysedatensatz besteht aus Tweets, die mittels eines Python-Skriptes zwischen dem 18. März und dem 30. Mai 2019 über die Streaming API (statuses/filter) gesammelt wurden[1]. Als Filter dienten gängige im Kontext der Europawahl verwendet Begriffe, wie zum Beispiel *EP2019* und *Europawahl2019.* Dabei beschränkte sich die Suche nicht nur auf rein deutsche, sondern umfasste auch englische, spanische und französische Begriffe, um die Europawahl-Debatte auf Twitter möglichst umfangreich zu erfassen. Die vollständige Liste der Suchterme befindet sich im Anhang. Nach Abschluss der Datensammlung wurden die gesammelten Tweets anhand ihrer Sprache[2] in separate Datensätze aufgeteilt, um Vergleiche zwischen Sprachen zu ermöglichen. Die Studie untersucht Tweets deutscher und englischer Sprache. Der bereinigte deutschsprachige Datensatz umfasst 345.543 Tweets von 85.389 Accounts, der englischsprachige 677.562 englischsprachige Tweets von 203.793 Accounts.

Zudem gab es eine nachträgliche Anreicherung der erfassten Accounts *(hydration)* um weitere Informationen: Mittels der Python-Bibliothek *twtoolbox* (hhromic 2016) wurden im Dezember 2019 aktuelle Account-Daten wie beispielsweise Aktivitätswerte und Account-Status gesammelt. Der große zeitliche Abstand der Datenanreicherung ermöglichte somit Erkenntnisse über mittlerweile gesperrte oder gelöschte Accounts. Diese Information ist ein wichtiger Anhaltspunkt für die das Scoring eines Accounts: Ein Löschen oder Sperren seitens Twitter deutet darauf hin, dass diese Nutzerkonten Spam oder beleidigenden

[1]Aufgrund eines Verbindungsfehlers zur API konnten zwischen dem 11. und 13. April nicht alle Tweets erfasst werden.

[2]Twitter erkennt automatisch die (dominierende) Sprache eines Tweets und gibt diese als Tweet-Entity *lang* aus.

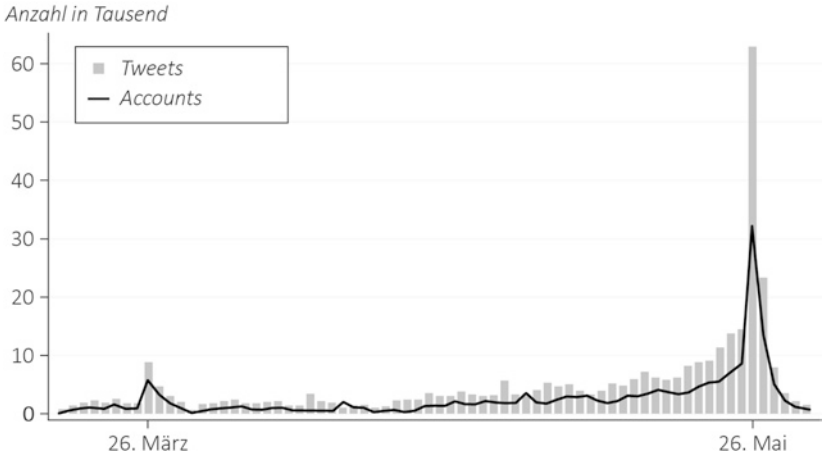

Abb. 1 Anzahl der erfassten deutschen Tweets und deren Urheber-Accounts im Zeitverlauf. (Quelle: Eigene Darstellung)

Tweets verbreiteten oder ein anderes missbräuchlichem Verhalten[3] aufzeigten (Twitter Inc. 2020).

Bei der Betrachtung der Tweet-Häufigkeit im Zeitverlauf zeigt sich wie erwartet eine deutliche Datenhäufung um das Wahlwochenende. Die tägliche Zahl deutschsprachiger Tweets im Datensatz (Abb. 1) steigt ab Mitte April zur Wahlphase hin nahezu konstant an, wobei mit 62.933 Tweets der Großteil am deutschen Wahltag (26. Mai) verfasst wurde. Direkt nach der Wahl sinkt die Tweet-Anzahl wieder deutlich ab. Die kleine Spitze am 26. März 2019 lässt sich auf die Debatte im Europäischen Parlament in Straßburg mit ihren Entscheidungen zur Abschaffung der Zeitumstellung und der Verabschiedung der Urheberrechtsreform zurückführen. Relativ analog verläuft die Anzahl der Urheber-Accounts im Datensatz, also die Zahl unterschiedlicher Nutzerkonten pro Tag.

Um zu untersuchen, ob und in welchem Ausmaß Social Bots in diesem Auszug der globalen Twitter-Aktivität der Europawahl aktiv waren, bedarf es detaillierter Analysen auf Tweet- und Account-Ebene. Wie in Kap. 2 beschrieben,

[3]Missbräuchliches Verhalten widersprich laut Twitter den Verhaltensregeln: https://help.twitter.com/en/rules-and-policies/twitter-rules.

existieren mittlerweile zahlreiche, teils sehr unterschiedliche Ansätze zur Identifikation und Analyse von (Social) Bots. Die folgende Analyse basiert auf dem Ansatz von Schäfer et al. (2017), Tweets mittels Nahduplikat-Erkennung anhand ihrer Ähnlichkeit in Cluster zu gruppieren, welcher bereits zur Untersuchung von Social Bot-Aktivitäten während des Bundestagswahlkampfs 2017 Anwendung fand (Pfaffenberger et al. 2019). Die Verteilung von Nahduplikaten gibt einerseits Auskunft über den Anteil ähnlicher Tweets eines Accounts (also dessen inhaltliche Variation), andererseits lassen sich auch Aussagen darüber treffen, ob es Account-Netzwerke im Datensatz gibt, die jeweils den gleichen oder ähnlichen Tweet verbreiten. Dieser computerlinguistische Ansatz über Nahduplikate nutzt zur Bot-Erkennung keine quantitativen Metriken, wie die Tweet-Aktivität, Frequenz oder die Account-Vernetzung, sondern basiert zunächst auf einer reinen Textanalyse.

Die zu untersuchenden Tweets wurden in drei Schritten für die Analyse vorverarbeitet: Zunächst wurde der Nachrichtentext jedes Tweets mit SoMaJo[4] tokenisiert, also in einzelne Wörter, Satzzeichen und Emoticons zerlegt (z. B. „Hello World! ☺“ in „Hello“ – „World“ –„!“ – „☺“). Und anschließend normalisiert. Dieser Verarbeitungsschritt entfernt alle im Tweet enthaltenen Interpunktions-, Leer- und Sonderzeichen (inklusive @ und #) sowie URLs, Mentions und Retweet-Marker. Aufgrund der Tatsache, dass viele simple Bots den Tweet-Text nur gering variieren oder beispielsweise nur eine angehängte URL oder den adressierten User (Mention) abwandeln, ist eine Betrachtung der vereinfachten, bereinigten Tweets sinnvoll. Im letzten Schritt erfolgte schließlich die Duplikat-Erkennung. Dafür wird jedem normalisierten Tweet ein Hashwert zugewiesen, der sich aus dem generierten Wortbündel eines Tweets ableitet. Tweets mit einem identischen Hashwert (also Hashwert-Duplikate) werden als *Nah*duplikate bezeichnet, da die Tweets nicht mehr in ihrem Rohzustand, sondern in einem sehr vereinfachten Wortbündel vorliegen. Tab. 2 veranschaulicht die zugrundeliegende Systematik. Die Nahduplikate wurden jeweils einem eigenen Cluster zugeordnet und anhand ihres Duplikat-Status klassifiziert: *unique* (kein Duplikat), *first* (zeitlich erster Tweet in einer Reihe von Nahduplikaten), *nduplicate* (alle weiteren ähnlichen Tweets innerhalb der Clusters).

Tab. 1 schlüsselt die Tweet-Häufigkeiten im Datensatz nach Duplikat-Status auf. Es zeigt sich bereits hier, dass der Datensatz einen eher geringen Umfang

[4]SoMaJo (Proisl und Uhrig 2016) ist ein Softwarepaket zur Tokenisierung und Satztrennung, das speziell für deutsch- und englischsprachige Internet- und Social Media-Texte entwickelt wurde.

Tab. 1 Systematik der Nahduplikat-Klassifizierung

Tweet-Zeitpunkt	Text	Status vor Bereinigung	Bereinigter Text	Status nach Bereinigung
Tag X, 12:00 Uhr	@userX nur Lügenpresse		nur lügenpresse	first
Tag X, 12:10 Uhr	alles Lügenpresse!!		alles lügenpresse	unique
Tag X, 12:30 Uhr	Nur Lügenpresse! ☹	duplicate	nur lügenpresse	nduplicate
Tag X, 12:31 Uhr	Nur Lügenpresse! ☹		nur lügenpresse	nduplicate
Tag Y, 11:00 Uhr	@userY Nur Lügenpresse!		nur lügenpresse	nduplicate

Tab. 2 Häufigkeitsverteilung deutschsprachiger Tweets nach Duplikat-Status und Accounts

Tweet-Status	Tweet-Anzahl (in %)	Urheber-Accounts
0 Kein Duplikat (unique)	305.543 (88)	80.191
1 Erster Tweet eines Nahduplikat-Clusters (first)	11.685 (3)	6078
2 Weiterer Tweet eines Nahduplikat-Clusters (nduplicate)	28.724 (8)	11.327
Gesamt	345.543 (100)	85.389

Analysezeitraum: 18. März bis 30. Mai 2019

an Nahduplikaten aufweist: Rund 88 % der Tweets sind originäre Tweets ohne Nahduplikate. Dem gegenüber stehen 40.409 Tweets, die sich auf 11.685 Cluster von 6078 unterschiedlichen Accounts verteilen. Im Folgenden sollen auf Basis der Nahduplikate mögliche Social Bots identifiziert und analysiert werden. Die Untersuchung bezieht sich im ersten Schritt auf die Nutzer-Ebene (Bot-Identifikation über Account-Eigenschaften) und in einem zweiten Schritt auf die Tweet-Ebene (Analyse von Nahduplikat-Clustern).

4 Analyse auf Account-Ebene

Eine manuelle Sichtung aller 85.389 im Datensatz vertretenen Accounts wäre weder wissenschaftlich sinnvoll noch zeitökonomisch vertretbar. Daher empfiehlt sich eine Vorselektion potenziell interessanter Accounts. Ein Blick auf das Histogramm zeigt, dass der Großteil der Accounts mit nur einem oder sehr wenigen Tweets im Datensatz vertreten ist. Durchschnittlich verfasst jeder Account 4,05 Tweets. Vereinzelte Accounts weisen Häufigkeiten größer 200 auf. Da es sich bei diesen Accounts jedoch auch um sehr aktive Menschen (also tatsächliche „User“) handeln kann und die reine Tweet-Anzahl nichts über die Heterogenität der Inhalte aussagt, wurden, wie bereits bei der Analyse zur Bundestagswahl 2017, je Nutzer die Anzahl an Nahduplikaten ermittelt und mit der jeweiligen Tweet-Anzahl im Datensatz verglichen. Mithilfe dieses Ansatzes lassen sich leicht Ausreißer mit auffälligen Werten visuell ermitteln.

Im Histogramm von Abb. 2 erkennt man deutlich, dass der Datensatz überwiegend aus Accounts mit nur wenigen Tweets (weniger als 50 innerhalb von 10 Wochen) besteht. Es gibt jedoch zwei Bereiche, die von besonderem Interesse sind: Eine Gruppe vereinzelter Accounts, die mit sehr vielen Tweets im Datensatz vertreten sind, aber nur einen geringen Duplikat-Anteil aufweisen – im Folgenden *High Performer* genannt – sowie ein Bereich, der aus einer Vielzahl an geringfügig bis stark aktiven Accounts mit einem Duplikat-Anteil von 80 bis

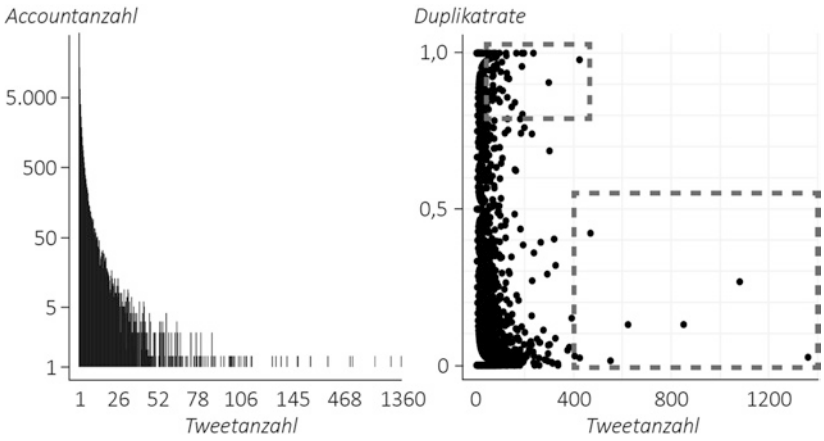

Abb. 2 Nutzerverteilung und Duplikatrate nach Tweet-Häufigkeit. (Quelle: Eigene Darstellung)

nahezu 100 % besteht. Die Analyse auf Account-Ebene betrachtet von diesen 5449 Accounts nur diejenigen 48, die mit mehr als 50 Tweets im Datensatz vertreten sind – im Folgenden *Duplikatoren* genannt. Die Begründung für den Ausschluss von Accounts mit weniger als 50 Tweets liegt in der Annahme, dass von einzelnen Usern mit sehr wenigen Tweets nur eine geringe Gefahr der Beeinflussung ausgeht. Sollten diese Accounts jedoch vernetzt agieren und gemeinsam eine ähnliche Botschaft verbreiten, so würden diese in der späteren Analyse der Nahduplikat-Cluster hervorstechen. Es liegt gibt somit nur ein geringes Risiko der Nichtbetrachtung relevanter Accounts.

Wie bereits erwähnt, wurden die Account-Informationen circa ein halbes Jahr nach der Datensammlung mit aktuellen Werten angereichert. Von den 85.389 Accounts sind mittlerweile 4632 (5,4 %) gesperrt oder gelöscht. Um einen Zusammenhang zwischen Sperrung/Löschung und der Account-Aktivität zu prüfen, wurde in Abb. 3 zusätzlich der Account-Status abgebildet. Die gleichmäßige Verteilung gesperrter Accounts hinsichtlich Aktivität und

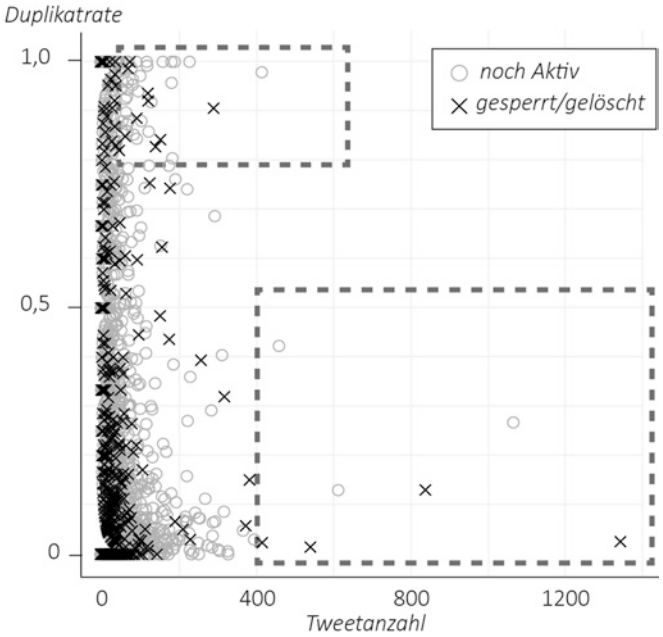

Abb. 3 Account-Verteilung im deutschen Datensatz nach Tweet-Häufigkeit, Duplikatrate und Status. (Quelle: Eigene Darstellung)

Duplikat-Anteil bestätigt die Vermutung, dass die Sperrung/Löschung der Accounts zweifelsohne nicht nur quantitativen Merkmalen, sondern scheinbar auch stark inhaltsbezogenen Kriterien folgt. Ein rein quantitatives Scoring würde zudem auch dazu führen, dass quasi alle Accounts von Nachrichtenmedien, die häufig als reine News-Bots agieren, gesperrt werden.

Die beiden Gruppen auffälliger Accounts werden im Folgenden näher analysiert. Nach einer manuellen Betrachtung der ausgewählten Nutzerkonten anhand ihrer quantitativen Eigenschaften (wie Tweet-Verhalten und -Häufigkeit) folgt eine Untersuchung ihrer geteilten Inhalte. Der aktuelle Account-Status spielte dabei zunächst keine Rolle. Die Verknüpfung einer computergestützten Voranalyse und Selektion mit einer anschließenden manuellen Untersuchung dieser Accounts hat sich bereits bei der Untersuchung von Bot-Aktivität zur Bundestagswahl 2017 behauptet (Pfaffenberger et al. 2019).

4.1 Bereich „High Performer“

Die acht Accounts in diesem Analysecluster weisen allesamt eine relativ hohe Tweet-Anzahl im Datensatz auf: Das Cluster umfasst alle Accounts im Datensatz mit mehr als 400 Tweets und einem Duplikatanteil kleiner 0,6. Nahezu alle Accounts verbreiteten politische Inhalte. Je ein Account teilte AfD-kritische, unionskritische, linke und sozialliberale Botschaften, vier Accounts wiederum rechtspopulistische Inhalte – teils verbunden mit der Verbreitung von Fake News und Verschwörungstheorien. Die Hälfte dieser Accounts wurde mittlerweile von Twitter gesperrt oder eingeschränkt, während es im gesamten Datensatz etwa fünf Prozent sind. Hier besteht also eindeutig ein Zusammenhang zwischen Zugehörigkeit zu diesem Cluster und einer Sperrung.

Nach der inhaltlichen Analyse wiesen jedoch nur zwei Accounts Automatisierungsanzeichen auf: Der News-Bot des linken Nachrichtenportals freiewelt.eu *(@FreieWeltEU)* sowie *@Nanomed8,* der in großem Umfang Fake-News, Prognosen zum schlechtem Abschneiden der gemäßigten politischen Parteien bei der Europawahl und Aufrufe zur „Abwahl“ von Bundeskanzlerin Merkel teilt. *@Nanomed8* weist jedoch keine offensichtlichen Anzeichen einer Automatisierung auf. Nur einzelne Phrasen wurden regelmäßig wiederholt und durch Nebensätze ergänzt. Auch die Aktivitätswerte liefern keinen direkten Nachweis für Bot-Aktivitäten: Es zeichnet sich ein Tages- und Nachtrhythmus ab, und die Tweet-Frequenz pro Tag schwankt (siehe Abb. 4). Es besteht somit auch die Möglichkeit, dass es sich hierbei um einen nicht- oder nur teilautomatisierten Cyborg handelte. *@Nanomed8* wurde mittlerweile von Twitter gesperrt, jedoch

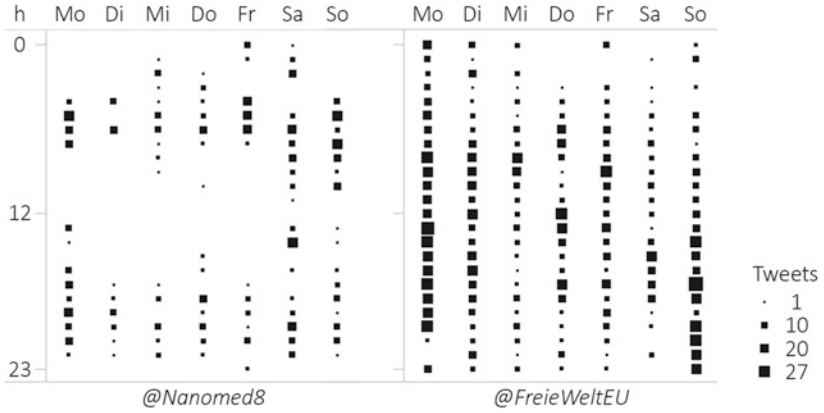

Abb. 4 Aktivitätsmuster der Accounts @Nanomed8 und @FreieWeltEU nach Wochentag und Stunde. (Quelle: Eigene Darstellung)

nicht zwangsläufig aufgrund einer Automatisierung, sondern unter Umständen auch aus inhaltlichen Gründen (siehe oben). Bis auf *@FreieWeltEU* weisen alle Accounts dieser Untersuchungsgruppe Aktivitätsmuster auf, die auf einen Tages-Nachtrhythmus mit Ruhephasen hindeuten. Man kann folglich davon ausgehen, dass die anderen sieben Accounts von Menschen (teilautomatisiert) betrieben werden.

4.2 Bereich „Duplikatoren“

Das Analysecluster umfasst 47 Accounts mit mehr als 50 Tweets im Datensatz und einem sehr hohen Duplikatanteil von 80 bis 100 %. Bis auf drei Accounts weisen alle weniger als 200 Tweets auf, 30 weniger als 100 Tweets. Nutzer mit diesen Eigenschaften sind von Interesse, da sie durch ihre geringfügige Aktivität eher „unter dem Radar“ der Bot-Erkennungs-Algorithmen von Twitter agieren können und auch für andere Nutzer weniger auffällig sind. Operieren mehrere solcher Accounts in einem Netzwerk, teilen also mehrere Accounts die gleiche oder ähnliche Nachricht, steigt zudem die Wahrnehmungswahrscheinlichkeit und Glaubwürdigkeit dieser Botschaft *(Astroturfing)*. Daher besteht Grund zur Annahme, dass es keineswegs nur simple Bots gibt, die die gleiche Botschaft tausendfach twittern, sondern auch Netzwerke/Gruppen von Accounts mit niedrigerer Aktivität.

Von den 48 Accounts in diesem Cluster sind zehn offizielle, zumindest teilautomatisierte Accounts von Nachrichtenmedien, wie beispielsweise der *Süddeutschen Zeitung* oder der *Frankfurter Allgemeinen Zeitung.* Diese teilen (ausgewählte) Schlagzeilen (zum Beispiel *@FOCUS_TopNews*) oder alle Meldungen zu bestimmten Themen eines Mediums *(@FAZ_Top),* weshalb von einer mittleren bis vollständigen Automatisierung ausgegangen wird. Weitere acht Accounts teilen mehr oder weniger systematisch Nachrichtenbeiträge.

Von den anderen 30 Nutzerkonten wiesen nur drei relativ sichere Anzeichen für eine menschliche Aktivität auf: Nur ein Teil des Nachrichtentextes ähnelte sich (beispielsweise durch die Verwendung der gleichen Hashtags), die Aktivität variierte stark nach Stunde, Wochentag und Datum und es gab teilweise Interaktion mit anderen Nutzern. Die restlichen Konten weisen Automatisierungstendenzen auf, wenngleich unterschiedlich stark ausgeprägt. Ein Teil agiert als simple Bots, die – in diesem Fall zumeist rechtspopulistische, fremdenfeindliche und verschwörungstheoretische – Nachrichten verbreiten. Dabei sticht besonders der Account *@ffd365* mit einem Nahduplikat-Anteil von über 98 % bei 423 Tweets hervor, hinter dem sich die rechte Plattform *Fernsehen für Deutschland* verbirgt. *@ffd365* machte ausschließlich Werbung für die Partei *DIE RECHTE* und deren Europawahl-Spitzenkandidatin Ursula Haverbeck. *@GntherSchulze2* ist laut Profilbeschreibung „Überzeugter AfD – Wähler". Seine 187 originären Tweets im Datensatz beschränken sich auf „Stimmt bei der Europawahl ab. #AfD". In seinem Account finden sich aber auch kommentarlose Retweets von IG Metall, FridaysforFuture, Die Linke, SPD-Politikern und der Piratenpartei. Die fehlende inhaltliche Stringenz spricht für einen schlecht programmierten Bot. Der mittlerweile gelöschte Account *@BEBELvsRALLE* teilte in 84 % der Tweets die Botschaft „Das AfD-Programm zur Europawahl in 99 Sekunden!".

Eine Kategorisierung in Bot und Mensch fällt bei einigen dieser Accounts jedoch auch nach manueller Sichtung der Tweets und Account-Informationen schwer. So ist der noch immer aktive Account *@alba_dalai* ausschließlich mit Tweets im Datensatz, die die Botschaft „Die #Erde ist unsere #Wirtszelle sei kein #Virus wähle #Grün #Europawahl 2019" beinhalten. Eine ausführliche Betrachtung des Profils lässt jedoch vermuten, dass sich dahinter eine existierende Person verbirgt, die das Twitter-Profil womöglich mit Unterstützung von Software pflegt. Von den 48 Nutzerkonten in diesem Cluster wurden mittlerweile elf gesperrt und gelöscht (23 %). Auch hier besteht ein Zusammenhang zwischen Zugehörigkeit zum Cluster und der allgemeinen Löschungshäufigkeit (5 %). Es überrascht, dass zwei Accounts in diesem Cluster bereits bei der Analyse zur Bundestagswahl 2017 auffielen und noch immer aktiv sind: *@ARCHITEKTENHAUS,* ein Spam-Account mit Bezug zur rechten politischen

Gruppierung *Rettung für Deutschland (RfD),* der in diesem Datensatz 56 nahezu identische Tweets teilte, sowie *@ShananJanie,* ein sehr simpler Retweet-Bot mit Nachrichten der Zeitschrift *Focus.*

4.3 Vergleich der Analysecluster

Zur weiteren Analyse werden die verschiedenen Analysecluster miteinander verglichen. Die Einteilung der Nutzerkonten erfolgte anhand ihres Account-Status beziehungsweise ihrer Cluster-Zugehörigkeit: *Duplikatoren* (1), *High Performer* (2), gelöschte Konten ohne Cluster-Zugehörigkeit (0) sowie sonstige, unauffällige Accounts (9). Tab. 3 fasst die durchschnittlichen Aktivitätswerte nach Account-Status zusammen, Abb. 5 zeigt die tägliche nach Account-Status aggregierte Tweet-Anzahl im Zeitverlauf, wobei aus Darstellungsgründen die Tweets der sonstigen unauffälligen Konten ausgeblendet sind.

Die aggregierten Werte nach Account-Status zeigen erste Unterschiede zwischen den Gruppen. Unbeachtet der beiden Ausreißergruppen der *Duplikatoren* und *High Performer* mit sehr hohen Aktivitätswerten weisen die sonstigen bereits gelöschten Nutzerkonten eine moderat höhere Tweet-Häufigkeit (5,95 Tweets pro Account) und Tweet-Frequenz (0,08 Tweets pro Account am Tag) als die sonstigen unauffälligen Accounts auf (3,8 und 0,05). Jedoch kann erst eine detaillierte Betrachtung auf Account-Ebene direkte Unterschiede zwischen den Gruppen sichtbar machen, indem auf die einzelnen Aktivitätsmuster am Tag und im Zeitverlauf eingegangen wird.

Tab. 3 Deskriptive Statistiken nach Account-Status

Account-Status	Tweets (%)	Accounts (%)	$\text{ØTweets}_{\text{Account}}$	$\text{ØTweets}_{\text{Account/Tag}}$
0 Gelöscht und ohne Cluster	27.477 (8)	4618 (5)	5,95	0,08
1 Duplikatoren	5292 (1)	48 (0)	110,25	1,58
2 High Performer	5757 (2)	8 (0)	719,63	10,28
9 Sonstige Accounts	307.027 (89)	80.715 (95)	3,8	0,05
Gesamt	345.543 (100)	85.389 (100)	4,05	0,06

Analysezeitraum: 18. März bis 30. Mai 2019
(Quelle: Eigene Darstellung)

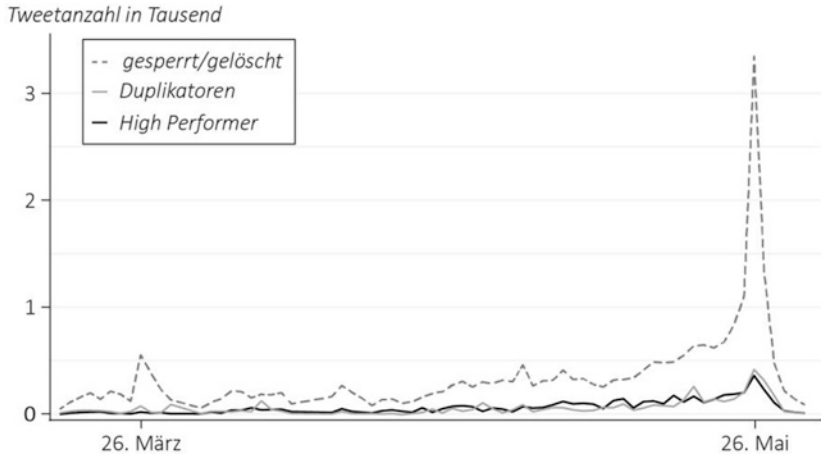

Abb. 5 Anzahl deutscher Tweets nach Account-Status im Zeitverlauf (ohne sonstige Accounts). (Quelle: Eigene Darstellung)

Abb. 6 vergleicht die Aktivitätsmuster der beiden definierten Analysecluster (Duplikatoren und High Performer) mit einer Zufallsstichprobe aus den sonstigen unauffälligen Accounts im deutschen Datensatz. Die Stichprobengröße richtet sich nach der Gruppengröße der Duplikatoren (48). Die Konten wurden jeweils anhand ihrer ID aufsteigend sortiert, wobei die Nutzerkonten mit der niedrigsten ID oben stehen. Da die durch Twitter vergebenen Account-IDs im Zeitverlauf ansteigen, ließen sich durch die Sortierung nach ID auch Aussagen auf Basis des relativen Account-Alters tätigen. Der erste Block bildet die Aktivität nach Stunde binär ab (aktiv/nicht aktiv), darunter befindet sich die aggregierte Tweet-Häufigkeit aller im Cluster enthaltenen Accounts. Danach erfolgt eine analoge Darstellung auf Datumsbasis.

Mithilfe dieser vereinfachten Darstellung erkennt man teils deutliche Unterschiede in der Account-Aktivität zwischen den beiden Analyseclustern und der Stichprobe, während innerhalb der drei Gruppen ein relativ einheitliches Muster vorherrscht. Bei den bereits betrachteten High Performern verteilt sich die Aktivität nahezu auf den gesamten Tag, wobei die aggregierte Tweet-Frequenz in den frühen Morgenstunden stark reduziert. Auch im Zeitverlauf zeigt sich eine relativ konstante und durchgehende Aktivität.

Im Vergleich dazu erkennt man bei der Tagesaktivität innerhalb der Duplikatoren stärkere Schwankungen und bei einigen Accounts deutliche

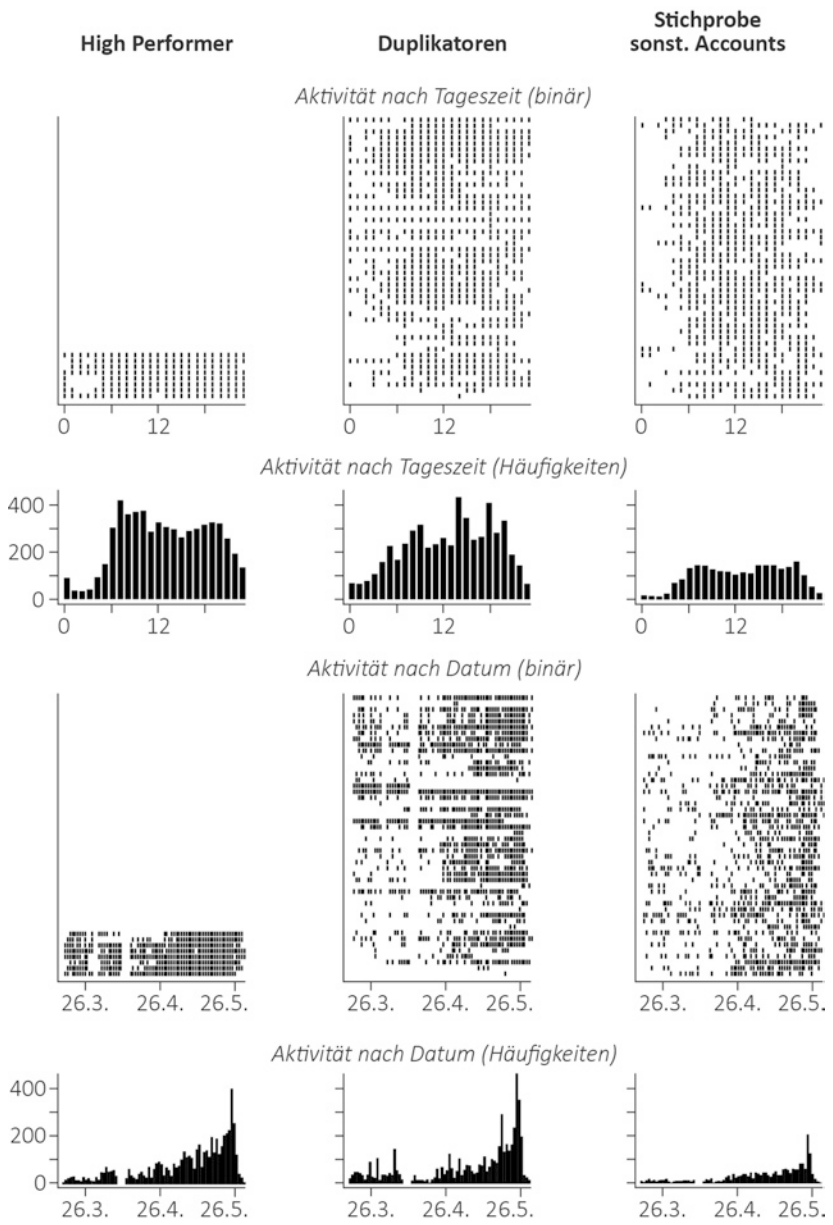

Abb. 6 Aktivitätsmuster ausgewählter Account-Cluster nach Tageszeit und Datum. (Quelle: Eigene Darstellung)

Phasen der nächtlichen Inaktivität. Zudem stechen die automatisierten Accounts der Nachrichtenmedien mit ihrer nahezu ganztägigen Aktivität hervor. Manche Nutzerkonten weisen im Tages- und Datumsverlauf nur einen oder sehr wenige Striche auf. Das heißt, sie tauchen nur zu einer bestimmten Uhrzeit beziehungsweise an einem oder vereinzelten Tagen mit Tweets im Datensatz auf. So wurden alle 187 erfassten Tweets von *@GntherSchulze2* am 20. März zwischen 14 und 15 Uhr gepostet sowie alle 75 Tweets von *@DonPajon* am 29. März zwischen 18 und 19 Uhr und alle 69 Tweets von *@ChristianH_Roth* datumsunabhängig um 18 Uhr. Während *@GntherSchulze2* vielfach Wahlaufrufe für die AfD teilte und somit auch für diese Untersuchung von Bot-Aktivität im Europawahlkampf interessant ist, verbreiteten *@ChristianH_Roth* und *@DonPajon* nur (wissenschaftlich belanglose) Werbung für eigene Blogbeiträge oder Musikalben. Eine konzentrierte Tweet-Aktivität an einzelnen Tagen oder zu bestimmten Uhrzeit spricht wie die eine ganztägig hohe Aktivität für eine Automatisierung der Accounts. Auch wenn die Accountanalyse nur auf Basis der mittels Suchtermen erfassten Kommunikation erfolgte, also keinesfalls der Anspruch an Datenvollständigkeit erhoben wird, lassen sich Rückschlüsse auf die Automatisierung ziehen. Denn folgt die Tweet-Aktivität bereits bei dieser systematischen Stichprobe einem bestimmten Muster, dann vermutlich auch bei allen anderen Tweets dieser Konten.

Die zufällige Stichprobe von Accounts, die in keinem der beiden Analysecluster vertreten sind (dritte Spalte), unterscheidet sich dagegen deutlich von den beiden vorigen Gruppen mit höchstwahrscheinlich automatisierten Accounts. Man erkennt einen Tag-Nacht-Rhythmus (keine nächtliche Aktivität/keine Striche an den Rändern), die absolute Tweet-Häufigkeit ist geringer. Im Zeitverlauf gibt es jeweils einige Tage der Inaktivität – mit einer Häufung der Aktivität rund um die Wahltage. Die deutlichen Unterschiede zwischen der Stichprobe und der systematisch selektierten Accounts bekräftigen das gewählte methodische Vorgehen: Durch die computergestützte Vorselektion anhand von Nahduplikaten lassen sich einfach auffällige Accounts mit atypischer Aktivität identifizieren.

Die Betrachtung auf Account-Ebene lieferte bereits erste Ansatzpunkte für eine Bot-ähnliche Twitter-Aktivität. Zumeist sind diese wahrscheinlich automatisierten Twitter-Konten jedoch simple Spam- oder News-Bots, während nur eine geringe Zahl tatsächlich politische Botschaften teilten. Stattdessen weisen einige Accounts, die politische Inhalte verbreiteten, Anzeichen überwiegend menschlicher Nutzung auf.

5 Analyse auf Tweet-Ebene

Die Nahduplikat-Analyse ermöglicht auch Aussagen über die Verbreitung ähnlicher Tweets durch mehrere Accounts. Diese Nahduplikat-Cluster sind folglich Gruppen von Nutzerkonten, die jeweils das gleiche Nahduplikat verbreiten. Wie bereits weiter oben argumentiert, können Gruppen von jeweils unauffälligen Accounts, die alle die gleiche Botschaft teilen, einfacher und unbeachteter Falschmeldungen streuen oder Debatten beeinflussen.

Auf Basis des ermittelten Duplikat-Status *(unique, first, nduplicate)* jedes Tweets im Datensatz wurden diese Tweets gruppiert: Jedem ersten Nahduplikat (nach Zeitstempel) wurden die weiteren Nahduplikate zugeordnet. Anschließend wurde die absolute Tweet-Zahl (Cluster-Größe) und die Anzahl eindeutiger Accounts in diesem Cluster ermittelt. Die durchschnittliche Clustergröße ist mit 3,46 Tweets (*SD:* 8.82, *min:* 2, *max:* 448) von 2,37 Accounts (*SD:* 6.95, *min:* 1, *max:* 406) klein. Kleine Cluster können auch durch das zufällige Verbreiten eines ähnlichen Inhaltes entstehen. Demgegenüber lassen große Cluster eine systematische Verbreitung von Inhalten vermuten – sei es durch offizielle Kampagnen, populäre Hashtags oder eben durch Bot-Netzwerke. Es fällt auf, dass der tägliche Anteil an Nahduplikaten insgesamt mit 0,13 ($SD = 0{,}02$, *min:* 0,07, *max:* 0,16) und der Anteil an ersten Nahduplikaten *(first)* beziehungsweise Clustern mit 0,04 ($SD = 0{,}01$, *min:* 0,02, *max:* 0,08) relativ konstant bleiben, wenngleich beide Werte im Zeitverlauf leicht abnehmen. Dies kann durch die Konstruktion der Datensammlung bedingt sein, da der Untersuchungszeitraum zu beiden Seiten beschränkt ist: Je später die Veröffentlichung eines Tweets im Untersuchungszeitraum, desto geringer ist aufgrund des immer kürzeren Zeitfensters die Wahrscheinlichkeit, dass ein Nahduplikat folgt.

Um auffällige Tweet-Cluster zu identifizieren, wurden Cluster-Größe (Tweet-Anzahl) und Account-Anzahl analog zum Vorgehen bei der Account-Analyse in einem Streudiagramm gegenübergestellt (siehe Abb. 7). Dabei erkennt man die deutliche Fallhäufung auf der 45 Grad-Linie (jedes Nahduplikat wurde durch einen anderen Account gepostet) und auf der X-Achse (alle Nahduplikate wurden durch den gleichen Account gepostet) – jeweils vor allem Nahduplikat-Cluster mit geringerer Tweet-Anzahl (kleiner 50). Letztere Fälle sind bereits durch die Analyse auf Account-Ebene gut abgedeckt, da hier Accounts mit hohem Nahduplikat-Anteil betrachtet wurden. Von größerem Interesse sind wieder die Ausreißer in der Grafik, also hier 13 Nahduplikat-Cluster mit mehr als 100 Tweets, deren nähere Betrachtung.

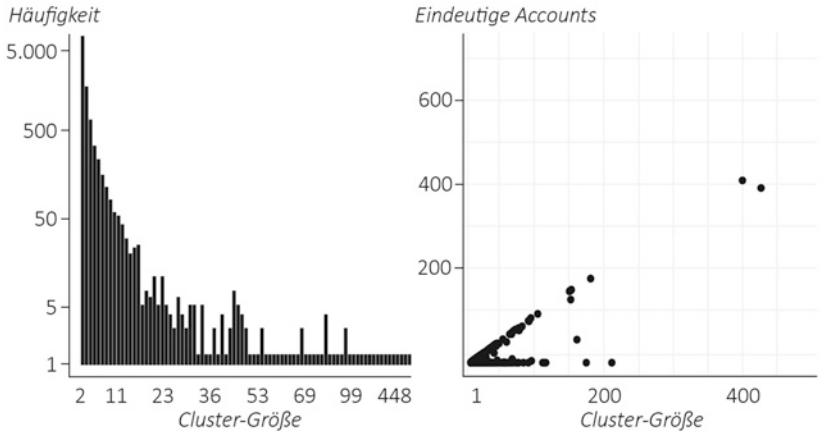

Abb. 7 Tweet-Verteilung der deutschen Nahduplikat-Cluster nach Tweet- und Account-Anzahl. (Quelle: Eigene Darstellung)

Die fünf Cluster auf der X-Achse gehören alle zu den bereits im vorigen Kapitel identifizierten, bot-ähnlichen Accounts, wie *@ffd365, @GntherSchulze2* oder*@CrazyIvan1979,* die vor allem Wahlaufrufe für die AfD verbreiteten. Da diese Cluster jeweils nur aus Tweets eines Accounts bestehen, sind sie für die weitere Analyse irrelevant.

Die Tweet-Cluster auf der 45 Grad-Achse haben ein Tweet-Account-Verhältnis gleich 1 (jeder Tweet von einem anderen Account) und repräsentieren Tweets zu Hashtag-Aktionen wie *#thistimeimevoing* und *#1europafüralle* (Ein Europa für Alle 2019). Dort finden sich auch Online-Aktionen des Kampagnen-Netzwerks *Avaaz,* bei der man Kampagnen der Webseite – beispielsweise gegen die Verbreitung von Falschmeldungen im Internet (Avaaz 2019) – über den einen Button teilen konnte (siehe Abb. 8). Eine solche geteilte Kampagne richtete sich gegen die Verbreitung von Falschinformationen während der Europawahl. Ein weiteres Tweet-Cluster beinhaltete eine für diese Untersuchung relevante Wahlkampf-Botschaft: 52 User teilten insgesamt 172-mal „Das AfD-Wahlprogramm in 99 Sekunden“, wovon allein 78 Nahduplikate der Account *@BEBELvsRALLE* verbreitete, der bereits zuvor mit hoher Wahrscheinlichkeit als Bot eingestuft wurde. Eine genauere Untersuchung ergab jedoch, dass es sich hier weitestgehend um Tweets handelt, die – analog zu den Avaaz-Kampagnen – über einen Teilen-Button mit vordefiniertem Text verbreitet wurden.

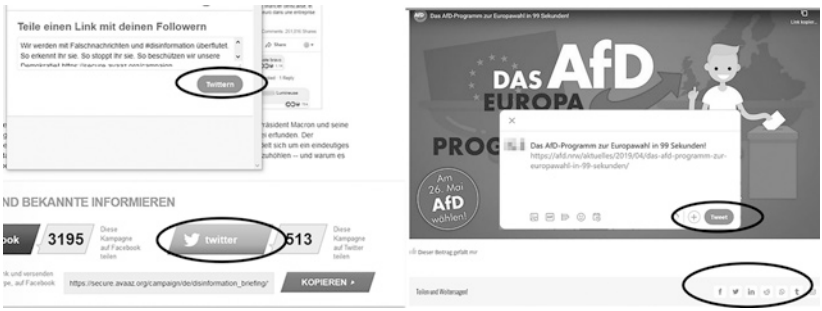

Abb. 8 Teilen-Funktion mit Text-Vorschlag bei Avaaz-Kampagnen (Avaaz 2019) und AfD-Videos (Alternative für Deutschland 2019). (Quelle: Eigene Darstellung)

Somit zeigt sich zwar, dass es mehrere Nahduplikat-Cluster gibt, die jeweils gleiche/ähnliche politische Botschaften während des Europawahlkampfes teilten, jedoch keines, das eindeutig Bot-ähnliche Aktivitäten aufweist. Die hier analysierten statistischen Ausreißer bezogen sich alle auf Online-Kampagnen, die wohl überwiegend bewusst durch Menschen geteilt wurden.

Insgesamt gibt es auf Basis der gesammelten deutschsprachigen Tweets kaum Anhaltspunkte für Bot- oder Bot-ähnliche Aktivitäten während des Europawahlkampfes auf Twitter. Vereinzelte Accounts wiesen deutliche Charakteristika einer Automatisierung auf, deren Ausmaß beziehungsweise Einfluss im Vergleich zur Gesamtzahl an Tweets/Accounts im Datensatz allerdings nur gering sein dürfte. Hinter den wenigen verdächtigen Nahduplikat-Clustern verbargen sich vor allem die Tweets einzelner Bots oder massenhaft geteilte Internet-Kampagnen. Es konnte keine auffälligen und aktiven Bot-Netzwerke im Datensatz entdeckt werden.

6 Vergleich mit englischsprachigen Tweets

Die Analysen im vorigen Kapitel gaben bereits Hinweise auf (teil-)automatisierte Account-Aktivität, wenn gleich in sehr geringem Ausmaß. Um sprachenbezogene Effekte auszuschließen, folgt eine zusätzliche Betrachtung englischsprachiger Tweets. Neben den bereits untersuchten 345.543 deutschsprachigen Tweets wurden im gleichen Zeitraum (18. März bis 30. Mai 2019) auch 677.562 englischsprachige Tweets von 203.793 unterschiedlichen Accounts erfasst.

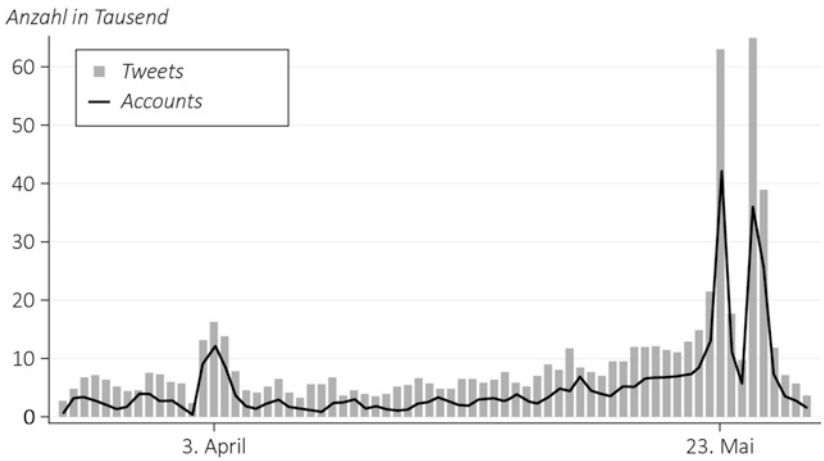

Abb. 9 Anzahl der erfassten englischen Tweets und deren Urheber-Accounts im Zeitverlauf. (Quelle: Eigene Darstellung)

Bei der Anzahl der täglich erfassten Tweets im Zeitverlauf in Abb. 9 erkennt man deutlich drei Spitzen: Eine um den 2. bis 4. April sowie je eine am 23. und am 26./27. Mai. Die erste Spitze lässt sich auf die Mehrdeutigkeit des Hashtags „EP" zurückführen: *#EP* wird unter anderem als Abkürzung für *European Parliament* verwendet und diente deshalb als Suchterm bei der Datensammlung. Zugleich ist EP aber auch eine Abkürzung für einen Tonträger. Ein Großteil der Tweets Anfang April befasste sich mit dem neu erschienenen Album der südkoreanischen Girlgroup „Blackpink" (Billboard Magazine 2019). Nahezu alle Tweets im Datensatz zu diesem Thema wurden zwischen dem 2. und 5. April verfasst. Die zweite Spitze am 23. Mai fällt auf den Beginn der Wahlphase und den Wahltermin in Großbritannien. Den rapiden Abfall der Tweet-Anzahl in den beiden Folgetagen lässt vermuten, dass direkt nach dem Wahltag das Interesse für die Europawahl im englischsprachigen Raum zunächst abnimmt, um zum Ende der Wahlphase und der Bekanntgabe der Ergebnisse am 26. Mai wieder deutlich zuzunehmen.

Wie Tab. 4 zeigt, ist hier der Anteil von Nahduplikaten um 4 Prozentpunkte höher – bei einem annähernd gleichen Anteil an Accounts im Datensatz, die diese Nahduplikate verbreiten. Dies kann ein Indiz für eine erhöhte Bot-Aktivität sein, deren Existenz beziehungsweise Ausmaß in den folgenden beiden Unterkapiteln auf Account- und Tweet-Ebene ermittelt wird.

Tab. 4 Häufigkeitsverteilung englischsprachiger Tweets nach Duplikat-Status und Accounts

Status	Tweet-Anzahl (in %)	Urheber-Accounts
0 Kein Duplikat (unique)	573.576 (85)	186.123
1 Erster Tweet eines Nahduplikat-Clusters (first)	22.244 (3)	12.267
2 Weiterer Tweet eines Nahduplikat-Clusters (nduplicate)	81.742 (12)	27.836
Gesamt	677.562 (100)	203.793

Analysezeitraum: 18. März bis 30. Mai 2019. Aufgrund eines Verbindungsfehlers zur API konnten zwischen dem 11. und 13. April nicht alle Tweets erfasst werden

6.1 Account-Ebene

Analog zum Vorgehen im deutschen Datensatz wurden zunächst Tweet-Anzahl und Duplikat-Rate auf Account-Ebene gegenübergestellt, wodurch sich, ähnlich zum deutschen Datensatz, wieder zwei Gruppen verdächtiger Accounts zeigten: Eine Gruppe von Accounts mit mehr als 50 Tweets und einer Duplikat-Rate größer 0,8 („englische Duplikatoren“) und vereinzelte Accounts mit mehr als 400 Tweets im Datensatz und einem Duplikat-Anteil unter 50 % („englische High Performer“). Abb. 10 stellt die Verteilung grafisch dar.

Insgesamt wurden im englischen Datensatz etwa 7,4 % der Accounts bereits gelöscht, also etwa 2 Prozentpunkte mehr als im deutschen Datensatz. Auch hierbetrifft dies wieder vor allem Accounts mit einer kleinen Anzahl erfasster Tweets (siehe Abb. 11). Die durchschnittliche Tweet-Häufigkeit liegt bei 3,32, was einer Tweet-Frequenz von 0,05 Tweets pro Tag und Account entspricht (siehe Tab. 5). Mit Blick auf das Histogramm und das Streudiagramm wird eine deutlich größere Streuung der Tweet-Aktivität sichtbar. Dies zeigt sich auch in den Unterschieden zwischen den Analyseclustern beider Datensätze: während die Aktivität der sonstigen unauffälligen Accounts mit einer durchschnittlichen Tweet-Häufigkeit von 3,03 signifikant niedriger als im deutschen Datensatz (3,8) ist, sind die Aktivitätswerte der Duplikatoren und High Performer deutlich höher.

Eine erste inhaltliche Analyse ergibt, dass sich die Inhalte deutlich vom deutschen Datensatz unterscheiden: Während die (verdächtigen) Accounts der deutschen Analysecluster überwiegend Nachrichten und teilweise rechtspopulistische Inhalte teilten, verknüpfen die Tweets innerhalb der englischsprachigen Analysecluster vor allem die Europawahl mit dem nahenden Brexit (vgl. Adrian et al. in diesem Band).

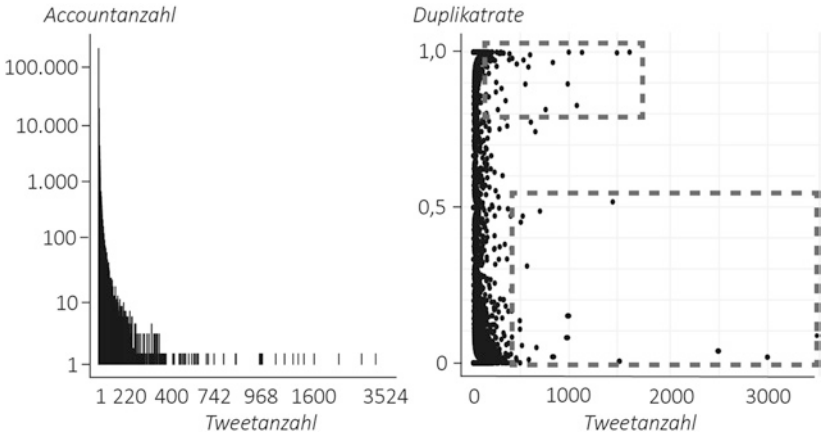

Abb. 10 Nutzerverteilung nach Tweet-Häufigkeit und Duplikatrate im englischen Datensatz. (Quelle: Eigene Darstellung)

Von den 147 Accounts der englischsprachigen Duplikatoren wurden mittlerweile 39 gelöscht oder gesperrt. Das Analysecluster umfasst vor allem Spam-Accounts (46) und einfache Retweet-Bots (37). Nur wenige Accounts verbreiteten überwiegend politische Inhalte. Neben *@drjdvalentin,* einem Retweet-Bot für AfD-Tweets, waren dies vor allem 19 Accounts, die ausschließlich Brexit-bezogene Inhalte teilten – 11 davon konnten dem *Leave*-, 8 dem *Remain*-Lager zugeordnet werden. Nahezu alle Accounts dieses Analyseclusters weisen aufgrund ihres kaum variierenden Nachrichtentextes starke Automatisierungstendenzen auf. Neben sehr einfachen Spam-Bots wie *@posts_from_Asia,* der täglich und zu jeder Uhrzeit 15 bis 25 nahezu identische Tweets teilt, gibt es auch Accounts wie *@SirBrianClough1,* der zwar überwiegend identische „Remain-Aufrufe" postet, jedoch einem Tag-Nacht-Rhythmus folgt und eine schwankende Tagesaktivität aufweist. Aufgrund der großen Zahl an Accounts wird auf eine grafische Darstellung der Account-Aktivität wie auf Seite 13 verzichtet.

Von den 20 High Performern wurden mittlerweile 5 gelöscht oder gesperrt. Insgesamt weist die Hälfte starke Indizien einer Automatisierung auf. Wie bereits bei den Duplikatoren befasst sich die überwiegende Zahl der Accounts mit dem Brexit, zehn teilten ausschließlich Brexit-bezogene Inhalte, äußerten sich meinungsstark und versuchten die Debatte durch Meldungen und Kommentare für oder gegen den Brexit zu beeinflussen. Hierbei fallen vor allem zwei mittlerweile gelöschte Accounts auf, die vermutlich gemeinsam gesteuert wurden: *@Bevanite2017* und

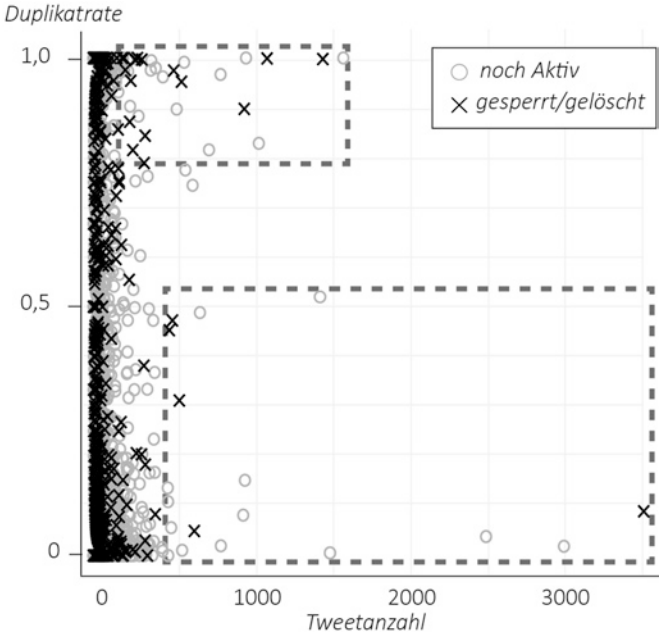

Abb. 11 Account-Verteilung im englischen Datensatz nach Tweet-Häufigkeit, Duplikatrate und Status. (Quelle: Eigene Darstellung)

Tab. 5 Deskriptive Statistiken nach Account-Status

Account-Status	Tweets (%)	Accounts (%)	$\text{ØTweets}_{\text{Account}}$	$\text{ØTweets}_{\text{Account/Tag}}$
0 Gelöscht und ohne Cluster	57.288 (8)	15.021 (7)	3,81	0,05
1 Duplikatoren	26.866 (4)	143 (0)	187,87	2,68
2 High Performer	21.021 (3)	20 (0)	1051,05	15,02
9 Sonstige Accounts	572.387 (85)	188.609 (93)	3,03	0,04
Gesamt	677.562 (100)	203.793 (100)	3,32	0,05

Analysezeitraum: 18. März bis 30. Mai 2019
(Quelle: Eigene Darstellung)

@Bevanite2019 – wohl in Anlehnung an den englischen Bevanismus, dem linken Flügel der Labour Partei (Gelman 1954) – hatten deutliche Automatisierungs-Merkmale und teilten in großen Mengen „Remain-Aufrufe". *@Bevanite2019* startete die Aktivität nach der Sperrung/der Löschung von *@Bevanite2017*. Bei den anderen acht Brexit-bezogenen Accounts in diesem Cluster kann man aufgrund ihres Tweet-Verhaltens (Aktivitätsmuster, Interaktionen, geteilte Inhalte) davon ausgehen, dass es sich um Menschen/User handelt. Von den 10 mit hoher Wahrscheinlichkeit automatisierten Accounts teilten die meisten wiederum Nachrichtenmeldungen mit Verlinkung zur jeweiligen Nachrichten-/Blogseite.

Zusammenfassend bestätigen die Ergebnisse im englischen Datensatz zwar die Vermutung, dass eine Beeinflussung politischer Debatten aufgrund der Durchschaubarkeit mittlerweile kaum noch über hochaktive automatisierte Accounts erfolgt. Es überrascht jedoch, dass diese Debatte nicht nur von mäßig aktiven, automatisierten Accounts (bei den Duplikatoren) beeinflusst wird, die „unter dem Radar" agieren, sondern auch von hochaktiven menschlichen Nutzern (bei den High Performern). Dies widerspricht der üblichen Annahme, dass gerade hochaktive Accounts eher automatisiert werden.

Es konnten im englischen Datensatz insgesamt und auch relativ mehr Accounts gefunden werden, die mit hoher Wahrscheinlichkeit (teil-)automatisiert sind und politische Inhalte teilen. Zumeist handelte es sich jedoch um Brexit-bezogene Tweets aus dem Remain- und Leave-Lager und nicht um Inhalte, die sich direkt auf die Europawahl bezogen. Um diese Beobachtung zu bekräftigen, erfolgte wie beim deutschen Datensatz nach der Analyse auf Account-Ebene eine Untersuchung der Nahduplikat-Cluster.

6.2 Tweet-Ebene

Wie im deutschen Datensatz wurden Tweets anhand ihres Nahduplikat-Status in Cluster gruppiert. Die Clustergrößen sind dabei ähnlich verteilt wie im deutschen Datensatz, aufgrund der größeren Datenmenge sind die Cluster jedoch entsprechend größer. Das Streudiagramm (Abb. 12) verdeutlicht mehrere Ausreißer, also Cluster mit besonders hoher Tweet- oder Nutzer-Anzahl, die im Folgenden näher betrachtet werden.

Auch im englischen Datensatz beinhalten die Cluster entlang der 45-Grad-Achse Tweets von Online-Kampagnen (wie von Avaaz) und populären Hashtags (wie *#thistimeimvoting, #letskillthislove)* oder typische Tweets mit einfachen Reaktionen oder Antworten (wie „this: …"). Entlang der X-Achse sind die Cluster abgebildet, deren Nahduplikate von einem oder wenigen Accounts

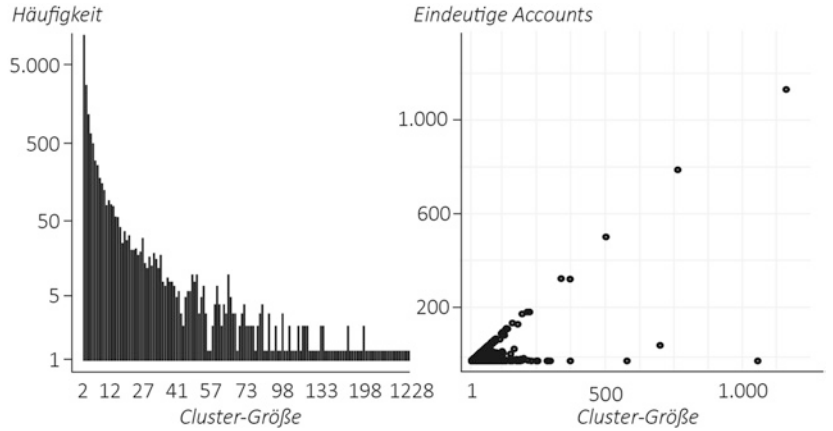

Abb. 12 Tweet-Verteilung der englischen Nahduplikat-Cluster nach Tweet- und Account-Anzahl. (Quelle: Eigene Darstellung)

verbreitet wurden. Dies sind zumeist Accounts der *High Performer,* wie @*posts_from_Asia* oder @*atomic_tweets,* jedoch auch ein Cluster von 66 meist automatisierten Accounts, die Nachrichtenmeldungen oder Tweets anderer Parteien teilten („@xyz posted …"). Aufgrund der Tatsache, dass die geteilten URLs bei der Normalisierung vor der Nahduplikat-Analyse entfernt wurden, variierten derartige Retweets nur noch im Namen des zitierten Accounts (Mention), weshalb der Algorithmus diese Tweets unbeabsichtigt zu einem Nahduplikat-Cluster bündelte.

Die Cluster zwischen diesen beiden Achsen bestehen aus Accounts mit heterogener Aktivität: Dort sind sowohl Accounts mit nur einem Nahduplikat als auch Accounts mit vielen Duplikaten vertreten. In diesem Bereich fallen vor allem mehrere Nahduplikat-Cluster auf, die zahlreiche Tweets zu „Iraqui refugees in Turkey" und zu „Iranian refugees in Turkey" verbreiteten. Insgesamt 285 Cluster mit 4293 Tweets von 279 unterschiedlichen Nutzer teilten Hilfsgesuche (zum Beispiel Bitten um ein Resettlement oder eine Asylaufnahme). Zwar ließ die starke inhaltliche Nähe der Tweets eine Automatisierung vermuten, eine nähere Betrachtung der Aktivitätsmuster nach Datum und Uhrzeit konnte dies jedoch nicht bestätigen. Vermutlich bedienten sich die Account-Nutzer (vermutlich die betroffenen Flüchtlinge selbst) bestehender Formulierungen bereits veröffentlichter Tweets anderer Accounts oder retweeteten diese.

7 Die überschätzte Gefahr von Social Bots im politischen Kontext

Insgesamt konnte diese Studie keine Bot-Netzwerke oder Cluster von höchstwahrscheinlich automatisierten Accounts mit nennenswerter Aktivität im Sinne einer Verbreitung politischer Botschaften oder Beeinflussung von Debatten im Kontext der Europawahl erkennen.

Es finden sich im deutschen und englischen Datensatz zur Europawahl mehrere Account-Gruppen, die eine ähnliche politische Botschaft verbreiteten. Einige Cluster verbreiteten Tweets zur Lage von iranischen und irakischen Flüchtlingen in der Türkei. Auch wenn Größe und Inhalte der Cluster zunächst für die Existenz von Bot-Netzwerken sprachen, konnte dies ein Blick auf die Aktivitätsmuster nicht bestätigt werden. Der überwiegende Anteil dieser Cluster ließ sich auf Online-Kampagnen zurückführen, bei denen Nutzer vorformulierte Texte über einen Share-Button über den eigenen Twitter-Account teilen konnten. Eine Automatisierung der Nutzerkonten fand also nicht statt – vielmehr handelte es sich um ein bewusstes Verbreiten durch Menschen. Diese Erkenntnis ist für Ansätze, die auf Nahduplikat-Erkennung beruhen, sehr wichtig. Account-Gruppen, die identische Inhalte verbreiten, sind demnach nicht zwangsläufig ein Anzeichen für Bot-Aktivitäten. Es bedarf also einer genaueren Betrachtung der verbreiteten Inhalte und einer detaillierten Recherche, um mittels Share-Buttons verbreitete Text-Duplikate auszuschließen. Zudem zeigt sich, dass es gar keiner aufwändigen Maßnahmen wie Bot-Netzwerke bedarf, um massenhafte politische Botschaften zu verbreiten. Share-Kampagnen sind mächtige Beispiele, wie Inhalte mittels motivierter Menschen einfach und ohne Koordinationsbedarf geteilt werden können.

Auch wenn die Untersuchung keine Hinweise auf die Aktivität von Bot-*Netzwerken* liefern kann, so konnten dennoch einzelne Accounts identifiziert werden, die mit hoher Wahrscheinlichkeit automatisiert politische Botschaften verbreiten. Verglichen mit der Gesamtzahl an Accounts und Tweets in den Datensätzen ist deren Bedeutung und Einfluss eher gering. Ein Teil der aufgrund ihres Nahduplikat-Anteils und Tweet-Häufigkeit selektierten „asymptomatischen" Nutzerkonten erwiesen sich bei genauerer Betrachtung jedoch als sehr engagierte Menschen. Vor allem im englischsprachigen Datensatz gab es vereinzelte sehr aktive Accounts, die im Umfeld der Brexit-Debatte auf Twitter Meinungen und (Falsch-)Meldungen streuten. Der überwiegende Anteil dieser Accounts weist jedoch starke Anzeichen menschlicher Aktivität auf: Ein ausführliches Nutzerprofil, schwankende Aktivitätsmuster sowie eine Interaktion mit anderen Nutzern

in Form von Replies, Mentions und Reetweets. Die vermutlich ideologisch, mitunter patriotisch motivierten Nutzer verfassten häufig mehr als 50 Tweets am Tag, hatten aber einen geringen Nahduplikat-Anteil. Diese Accounts wären demnach bei rein parametrischer Betrachtung (zum Beispiel anhand ihrer Tweet-Frequenz) leicht als Bots klassifiziert worden, während die Variation der Nachrichten und ihre tatsächliche Interaktion mit anderen (zum Beispiel durch Diskussionen und Gegenargumente) stark für einen menschlichen Nutzer sprechen.

Außerhalb des politischen Kontexts, also abseits von politischen Botschaften finden sich im Datensatz durchaus einige Accounts, die mit hoher Wahrscheinlichkeit automatisiert betrieben werden. Allerdings handelt es sich hierbei zumeist um typische Spam-Bots, die Werbe-Links mit momentan populären Hashtags kombinieren, oder News-Bots von Nachrichtenmedien, die Links zu Schlagzeilen oder themenspezifischen Nachrichtenmeldungen verbreiten. Die Automatisierung dieser Twitter-Konten ist vorwiegend offensichtlich, die zugrundeliegenden Mechanismen beschränken sich auf das reine Retweeten oder das Teilen einfacher, verlinkter Botschaften. Eine Subsumtion dieser Accounts zu *Social* Bots ist aufgrund ihrer sehr simplen Funktionsweise fraglich. Entsprechend der in Kap. 1 genutzt Definition von Social Bots liegt bei Verwendung von Mentions und Hashtags bestenfalls eine automatisierte Interaktion mit anderen Nutzern vor. Das Vortäuschen einer menschlichen Identität und der Versuch der Manipulation liegt hier nicht vor. Twitters Vorstoß zu einer Kennzeichnungspflicht von Bots (Hamblock 2020) ist in diesem Zusammenhang zwar ein erster Ansatz zur Regulierung und Sichtbarmachung gesteuerter Accounts. Die Aktivität bösartiger Accounts wird dies jedoch kaum einschränken, sondern eher nur die bereits offensichtliche Aktivität „guter" Bots (wie von Nachrichtenmedien) offiziell kennzeichnen.

Die Tatsache, dass bei diesem aufwendigen mehrstufigen Verfahren mit parameterbasierter Vorselektion und manueller Analyse nur wenige mit hoher Wahrscheinlichkeit automatisierte Accounts auffielen, dafür einige weitere, die teils auch nach intensiver Betrachtung nicht zweifelsfrei als Mensch oder Bot klassifiziert werden konnten, führt zu folgenden Schlussfolgerungen:

Rein parameterbasierte Klassifikationsverfahren können bei falscher Sensitivität eine große Unsicherheit hinsichtlich ihrer Präzision aufweisen. Wie die oben genannten Beispiele ideologisch motivierter, hochaktiver Menschen zeigen, bergen auf die Tweet-Aktivität basierende Betrachtungen oder anhand typischer Nutzerverhalten trainierte Machine-Learning-Algorithmen die Gefahr einer falschpositiven Bot-Klassifizierung von hochaktiven oder sich atypisch verhaltenden Nutzern.

Auch inhaltsbasierte Ansätze, wie die in dieser Studie angewendete Nahduplikat-Analyse, können anfällig für atypisches Nutzerverhalten sein. So könnten Accounts mit einer hohen Nahduplikat-Rate auch von Menschen betrieben werden, die bestimmte Botschaften duplizieren oder nur leicht abgewandelt teilen. Ein hoher Anteil ähnlicher oder identischer Tweets deutet folglich nicht zwangsläufig auf eine Automatisierung hin. Die über die Nahduplikat-Rate vorselektierten, auffälligen Accounts sollten daher durch eine anschließende manuelle Analyse klassifiziert werden.

Eine große Herausforderung bei der Identifikation von Bots sind Cyborgs, also Accounts, die innerhalb des Kontinuums zwischen real existierenden Personen und vollautomatisierten Accounts agieren. Sowohl bei parameter- als auch inhaltsgestützten Verfahren besteht eine große Gefahr von Klassifikationsfehlern (falschpositive und falschnegative Bot-Einstufungen). Darüber hinaus ist eine binäre Klassifikation bei hybridem Nutzungsverhalten mit phasenweiser automatisierter Steuerung und menschlicher Nutzerinteraktion nicht zielführend. Vielmehr bedarf es einer separaten Betrachtung dieser teilautomatisierten Accounts.

Die Bedeutung und Aktivität von *Social* Bots im politischen Kontext wird zudem überschätzt. Vollautomatisiert und intelligent agierende Accounts, die über die Verbreitung politischer Botschaften Meinung und Verhalten anderer Nutzer beeinflussen, konnten in dieser Untersuchung nicht ermittelt werden. Neben zahlreichen News- und Spam-Bots waren es vielmehr hochaktive, vermutlich intrinsisch motivierte Menschen, die keiner Steuerung durch Dritte bedürfen. Darüber hinaus generierten die wenigen als „hochwahrscheinlich automatisiert" klassifizierten Bots nur etwa zwei Prozent aller im Datensatz erfassten Tweets.

Literatur

Abokhodair, N., Yoo, D., & McDonald, D. W. (2015). Dissecting a social botnet. In D. Cosley, A. Forte, L. Ciolfi & D. McDonald (Hrsg.), *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing – CSCW 15* (S. 839–851). New York: ACM Press.

Ahmed, F., & Abulaish, M. (2013). A generic statistical approach for spam detection in online social networks. *Computer Communications, 36,* 1120–1129. https://doi.org/10.1016/j.comcom.2013.04.004.

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.

Alternative für Deutschland. (2019). Das AfD-Programm zur Europawahl in 99 Sekunden!, Alternative für Deutschland. https://afd.nrw/aktuelles/2019/04/das-afd-programm-zur-europawahl-in-99-sekunden/. Zugegriffen: 1. Mai 2020.

Avaaz. (2019). Briefing: Die akute Gefahr der Desinformation und wie wir uns schützen können, Avaaz. https://secure.avaaz.org/campaign/de/disinformation_briefing/. Zugegriffen: 1. Febr. 2020.

Bastos, M. T., & Mercea, D. (2019). The brexit botnet and user-generated hyperpartisan news. *Social Science Computer Review, 37,* 38–54. https://doi.org/10.1177/0894439317734157.

BBC News Labs. (2019). Bots, BBC News Labs. https://bbcnewslabs.co.uk/projects/bots/. Zugegriffen: 12. Jan. 2020.

Billboard Magazine. (2019, 25. März). BLACKPINK announce new single & EP ‚Kill this love'. https://www.billboard.com/articles/columns/k-town/8503849/blackpink-announce-kill-this-love-single-ep. Zugegriffen: 28. Jan. 2020.

Boshmaf, Y., Muslukhov, I., Beznosov, K., & Ripeanu, M. (2011). The socialbot network. In R. H. Zakon, J. McDermott, & M. Locasto (Hrsg.), *Proceedings of the 27th Annual Computer Security Applications Conference on – ACSAC ,11* (S. 93). New York: ACM Press.

Bu, Z., Xia, Z., & Wang, J. (2013). A sock puppet detection algorithm on virtual spaces. *Knowledge-Based Systems, 37,* 366–377. https://doi.org/10.1016/j.knosys.2012.08.016.

Cai, C., Li, L., & Zengi, D. (2017). Behavior enhanced deep bot detection in social media. In *2017 IEEE International Conference on Intelligence and Security Informatics (ISI)* (S. 128–130). IEEE.

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,* 811–824. https://doi.org/10.1109/TDSC.2012.75.

Davis, C. A., Varol, O., Ferrara, E., Flammini, A., & Menczer, F. (2016). BotOrNot. In J. Bourdeau, J. A. Hendler, R. N. Nkambou, I. Horrocks, & B. Y. Zhao (Hrsg.), *Proceedings of the 25th International Conference Companion on World Wide Web – WWW ,16 Companion* (S. 273–274). New York: ACM Press.

Daya, A. A., Salahuddin, M. A., Limam, N., & Boutaba, R. (2019). A graph-based machine learning approach for bot detection. In *2019 IFIP/IEEE Symposium on Integrated Network and Service Management (IM)* (S. 144–152).

Ein Europa für Alle. (2019). Ein Europa für alle. Deine Stimme gegen Nationalismus, Ein Europa für Alle. https://www.ein-europa-fuer-alle.de/. Zugegriffen: 5. Mai 2020.

Ferrara, E., Varol, O., Davis, C., Menczer, F., & Flammini, A. (2016). The rise of social bots. *Communications of the ACM*, 59(7), 96–104. http://arxiv.org/pdf/1407.5225v4.

Gallwitz, F. (2020a). Die Enquete-Kommission KI hatte mich um meine Einschätzungen zum Themenkomplex „Social Bots" gebeten. Hier meine Stellungnahme. https://twitter.com/FlorianGallwitz/status/1230067574341238784. Zugegriffen: 1. Mai 2020.

Gallwitz, F. (2020b). Stellungnahme zum Themenkomplex „Social Bots". https://www.in.th-nuernberg.de/Professors/Gallwitz/Stellungnahme-Enquete-KI-Social-Bots-Gallwitz.pdf. Zugegriffen: 1. April 2020.

Gelman, N. I. (1954). Bevanism: A philosophy for British Labour? *The Journal of Politics, 16,* 645–663. https://doi.org/10.2307/2126563.

Gensing, P. (2020). Das Problem mit den Social Bots, Tagesschau. https://www.tagesschau.de/faktenfinder/social-bots-111.html. Zugegriffen: 1. Mai 2020.

Gorwa, R., & Guilbeault, D. (2018). Unpacking the social media bot: A typology to guide research and policy. *Policy & Internet*. https://doi.org/10.1002/poi3.184.

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. Zywietz (Hrsg.), *Fake News, Hashtags & Social Bots: Neue Methoden populistischer Propaganda* (S. 51–68). Wiesbaden: Springer Fachmedien.

Graff, B. (2017, 13. Juli). Wenn Computerprogramme Propaganda betreiben, Süddeutsche Zeitung. https://www.sueddeutsche.de/digital/soziale-netzwerke-wenn-computer-programme-propaganda-betreiben-1.3581125. Zugegriffen: 12. Apr. 2020.

Hamblock, D. (2020). Updates to the Twitter Developer Policy, Twitter Inc. https://blog.twitter.com/developer/en_us/topics/community/2020/twitter_developer_policy_update.html. Zugegriffen: 20. Apr. 2020.

Hegelich, S. (2020a). #SocialBots Zwei Twitter-Nutzer gieren nach Aufmerksamkeit. Die Enquete Kommission KI fällt darauf rein. Und das sind jetzt die Suchergebnisse 2 und 3 auf Twitter. Schätze, ich war noch viel zu freundlich in meinem Text. https://twitter.com/SimonHegelich/status/1229028505066917888. Zugegriffen: 1. Mai 2020.

Hegelich, S. (2020b). Argumente zu #SocialBots: #OKBoomer, kein Grund zur Aufregung. Erläuterungen zu den Fragen der Enquete Kommission „Künstliche Intelligenz" zum Thema Social Bots. http://politicaldatascience.blogspot.com/2020/02/argumente-zu-socialbots-okboomer-kein.html. Zugegriffen: 1. Apr. 2020.

Hegelich, S., & Janetzko, D. (2016). Are social bots on Twitter political actors? Empirical evidence from a Ukrainian social botnet. In *Tenth International AAAI Conference on Web and Social Media.* https://www.aaai.org/ocs/index.php/ICWSM/ICWSM16/paper/download/13015/12793.

hhromic. (2016). twttoolbox. https://github.com/hhromic/python-twitter-toolbox. Zugegriffen: 1. Okober 2019.

Howard, P. N., & Kollanyi, B. (2016). Bots, #strongerin, and #brexit: Computational propaganda during the UK-EU referendum. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.2798311.

Kind, S., Jetzke, T., Weide, S., Ehrenberg-Silies, S., & Bovenschulte, M. (2017). *Social Bots. TA-Vorstudie* (Büro für Technikfolgen-Abschätzung beim Deutschen Bundestag (TAB), Hrsg.) (TAB-Horizon-Scanning Nr. 3). https://www.tab-beim-bundestag.de/de/pdf/publikationen/berichte/TAB-Arbeitsbericht-hs003.pdf. Zugegriffen: 10. März 2020.

Klinger, U. (2019, 6. Juni). Social Bots: Realität digitaler Öffentlichkeit. *Tagesspiegel*. https://background.tagesspiegel.de/digitalisierung/social-bots-realitaet-digitaler-oeffentlichkeit. Zugegriffen: 25. Apr. 2020.

Kreil, M. (2019, November). *The army that never existed: the failure of social bots research*. OpenFest Conference, Sofia, Bulgarien. https://github.com/MichaelKreil/openbots/tree/master/slides. Zugegriffen: 1. Mai 2020.

Lokot, T., & Diakopoulos, N. (2016). News Bots. *Digital Journalism, 4,* 682–699. https://doi.org/10.1080/21670811.2015.1081822.

Loyola-Gonzalez, O., Monroy, R., Rodriguez, J., Lopez-Cuevas, A., & Mata-Sanchez, J. I. (2019). Contrast pattern-based classification for bot detection on Twitter. *IEEE Access,* 45800–45817. https://doi.org/10.1109/access.2019.2904220.

Lypp, L. (2017). Wirkung von „Social Bots“ ist unter Sach-verständigen strittig, Deutscher Bundestag. https://www.bundestag.de/dokumente/textarchiv/2017/kw04-pa-bildung-forschung-social-bots-488818. Zugegriffen: 13. Apr. 2020.

Miller, Z., Dickinson, B., Deitrick, W., Hu, W., & Wang, A. H. (2014). Twitter spammer detection using data stream clustering. *Information Sciences, 260,* 64–73. https://doi.org/10.1016/j.ins.2013.11.016.

Pfaffenberger, F., Adrian, C., & Heinrich, P. (2019). Was bin ich – und wenn ja, wie viele? In C. Holtz-Bacha (Hrsg.), *Die (Massen-)Medien im Wahlkampf: Die Bundestagswahl 2017* (S. 97–124). Wiesbaden: Springer Fachmedien.

Pieterson, W., Ebbers, W., & Madsen, C. Ø. (2017). New channels, new possibilities: A typology and classification of social robots and their role in multi-channel public service delivery. In M. Janssen, K. Axelsson, O. Glassey, B. Klievink, R. Krimmer, & I. Lindgren (Hrsg.), *Electronic government. 16th IFIP WG 8.5 International Conference, EGOV 2017, St. Petersburg, Russia, September 4–7, 2017* (Lecture notes in computer science Information systems and applications, incl. internet/web, and HCI, Bd. 10428, S. 47–59). Cham: Springer International Publishing.

Proisl, T., & Uhrig, P. (2016). SoMaJo: State-of-the-art tokenization for German web and social media texts. In P. Cook, S. Evert, R. Schäfer, & E. Stemle (Hrsg.), *Proceedings of the 10th Web as Corpus Workshop* (S. 57–62). Stroudsburg: Association for Computational Linguistics.

Ratkiewicz, J., Conover, M. D., Meiss, M., Gonçalves, B., Flammini, A., & Menczer, F. M. (2011). Detecting and tracking political abuse in social media. *Fifth international AAAI conference on weblogs and social media* (S. 297–304). Barcelona: AAAI Press.

Reuter, M. (2019). Social Bots: Was nicht erkannt werden kann, sollte nicht reguliert werden, netzpolitik.org. https://netzpolitik.org/2019/social-bots-was-nicht-erkannt-werden-kann-sollte-nicht-reguliert-werden/. Zugegriffen: 12. Apr. 2020.

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,* 294–309. https://doi.org/10.1089/big.2017.0049.

Stieglitz, S., Brachten, F., Berthelé, D., Schlaus, M., Venetopoulou, C., & Veutgen, D. (2017a). Do social bots (still) act different to humans? – Comparing metrics of social bots with those of humans. In G. Meiselwitz (Hrsg.), *Social computing and social media. Human behavior* (Lecture Notes in Computer Science, Bd. 10282, S. 379–395). Cham: Springer International Publishing.

Stieglitz, S., Brachten, F., Ross, B., & Jung, A.-K. (2017b, 11. Oktober). *Do social bots dream of electric sheep? A categorisation of social media bot accounts*. http://arxiv.org/pdf/1710.04044v1.

Tagesschau. (2019, 13. Mai). EU-Kommissarin warnt vor Manipulation, Tagesschau. https://www.tagesschau.de/ausland/eu-wahl-justizkommissarin-russland-101.html. Zugegriffen: 14. Apr. 2020.

Twitter Inc. (2020). Über gesperrte Accounts, Twitter Inc. https://help.twitter.com/de/managing-your-account/suspended-twitter-accounts. Zugegriffen: 1. Mai 2020.

Van der Walt, E., & Eloff, J. (2018). Using machine learning to detect fake identities: bots vs humans. *IEEE Access,* 6540–6549. https://doi.org/10.1109/access.2018.2796018.

Varol, O., Ferrara, E., Davis, C. A., Menczer, F., & Flammini, A. (2017, 27. März). *online human-bot interactions: detection, estimation, and characterization*. https://arxiv.org/pdf/1703.03107.

Woolley, S. C. (2016). Automating power: Social bot interference in global politics. *First Monday*, *21*(4). https://doi.org/10.5210/fm.v21i4.6161.

Yang, K.-C., Varol, O., Hui, P.-M., & Menczer, F. (2019, 20. November). *Scalable and Generalizable Social Bot Detection through Data Selection*. http://arxiv.org/pdf/1911.09179v1.

Fabian Pfaffenberger, M.Sc., war wissenschaftlicher Mitarbeiter am Lehrstuhl für Kommunikationswissenschaft an der Friedrich-Alexander-Universität Erlangen-Nürnberg. Er promoviert zurzeit über methodische Herausforderungen der digitalen Kommunikation am Beispiel von Twitter. Seine Forschungsschwerpunkte sind Computational Methods, Social Media und Politische Kommunikation.

Philipp Heinrich, M.Sc., ist wissenschaftlicher Mitarbeiter am Lehrstuhl für Korpus- und Computerlinguistik der Friedrich-Alexander-Universität Erlangen-Nürnberg. Zu seinen Forschungsschwerpunkten zählen die automatische Verarbeitung von Daten aus sozialen Medien sowie korpusbasierte Diskursanalyse. Sein Promotionsprojekt beschäftigt sich mit der Erforschung der transnationalen algorithmischen Öffentlichkeit.