Evolution einer Desinformationskampagne

Twitter-Aktivitäten der Internet Research Agency zwischen 2009 und 2018

Erstgutachter:

Prof. Dr. Raphael Heiko Heiberger
Institut für Sozialwissenschaften
Abteilung für Computational Social Science

vorgelegt von:

Johannes Engels
Waiblinger Str. 15
71364 Winnenden
post@johannesengels.de
Mat.-Nr.: 3235908

Zweitgutachter:

Prof. Dr. André Bächtiger
Institut für Sozialwissenschaften
Abteilung für Politische Theorie und Empirische Demokratieforschung

Abgabedatum:

20.04.2021

Gliederung

Eigenständigkeitserklärung 3
1. Einleitung 4
2. Bisherige Untersuchungen und Forschungsfokus 6
3. Aufbau des Datensets und Struktur-Analysen 9
3.1 Account- und Tweet-Sprachen 9
3.2 Ortsangaben 12
3.3 Account-Erstelldaten 13
3.4 Account-Aktivitäten 13
3.5 Following und Follower 14
3.6 Tweet-Zeiten 16
3.7 Tweet- und Retweet-Zahl 17
4. Grundlagen und Aufbau des Structural Topic Models 19
4.1 STM 20
4.2 Datenvorbereitung 21
4.3 Suche nach K 23
5. STM-Analysen und Erkenntnisse 24
5.1 Topic-Themengruppen und -Inhaltscodierung 26
5.2 Anteilsstarke News-Topics 28
5.3 Gezielte Verbreitung von Falschnachrichten 29
5.4 Black Lives Matter, Black Representation und Polizeigewalt 31
5.5 Hillary Clinton: Emails und Benghazi 34
6. Rückschlüsse und Einordnung 35
7. Limitationen 38
8. Fazit 39
Literaturverzeichnis 41
Appendix A – Topics und Topic-Inhalte 44
Appendix B – Ausgewählte Topics im zeitlichen Verlauf 51

Eigenständigkeitserklärung

Ich erkläre,

1. dass diese Arbeit selbständig verfasst wurde,
2. dass keine anderen als die angegebenen Quellen benutzt und alle wörtlich oder sinngemäß aus anderen Werken übernommenen Aussagen als solche gekennzeichnet wurden,
3. dass die eingereichte Arbeit weder vollständig noch in wesentlichen Teilen Gegenstand eines anderen Prüfungsverfahrens gewesen ist,
4. dass die Arbeit weder vollständig noch in Teilen bereits veröffentlicht wurde und
5. dass das elektronische Exemplar mit den gedruckten Exemplaren übereinstimmt.

Winnenden, den 23.03.2021

Johannes Engels

1. Einleitung

Die Amtszeit von Ex-US-Präsident Donald Trump mag laut einigen politischen Kommentatoren die politische Felder Amerikas zwar deutlich gespalten haben (so sagen beispielsweise nur 1/5 der Anhänger beider Parteien, dass die Gegenpartei grundlegend ähnliche Werte und Ziele verfolgt, vgl. Pew Research Center 2020: o.S.), eine Kernidee scheint jedoch beide Seiten zu vereinen: Ein Wahlverlust hängt mit einer angeblichen Manipulation der Wahl durch die andere Seite zusammen.

Rechte Theorien zu den Verlusten in den Midterms 2018 und den Wahlen 2020 stützen sich dabei hauptsächlich auf die Idee, dass Demokraten Wahlmaschinen physisch beeinflussten und Stimmen fälschten[1]. Diese Theorien werden dabei meist – insbesondere im Fall der Wahlen 2020 – von Anhängern von Verschwörungstheorien wie QAnon geteilt[2], von offiziellen Staatsorganen jedoch abgelehnt und diskreditiert (vgl. AP News 2020: o.S.).

Linke Theorien zu den Verlusten 2016 drehen sich im Gegensatz dazu insbesondere um die Beeinflussung der öffentlichen Meinung im Vorfeld der Wahl, sei es durch russische Desinformationskampagnen oder Firmen wie Cambridge Analytica, oder gezielte Veröffentlichungen gehackter Materialien via Seiten wie Wikileaks. Während einige dieser Thesen zwar ebenfalls eher dem Feld der Verschwörungstheorien zuzuordnen sind, haben sich seit den Wahlen 2016 durchaus Beweise für diese Beeinflussung gefunden. Die Sonderermittlungen Robert Muellers zogen Anklagen gegen russische Akteure und Organisationen sowie Mitglieder des Trump-Kampagnenteams nach sich (vgl. Time 2019: o.S.) und Experten und Forscher kamen zu den Ergebnissen, dass „[t]he Russian government and its proxies have infiltrated and utilized nearly every social media and online information platform – including Instagram, Reddit, YouTube, Tumblr, 4chan, 9GAG, and Pinterest." (Rosenberger, zit. nach Select Commitee on Intelligence o.J.: 16).

Viele der von Laura Rosenberger angesprochenen *Social Media*-Seiten haben seit den Wahlen 2016 Anstrengungen unternommen, russische Desinformation und

[1] Für ein Beispiel zu rechten Theorien über die Midterms 2020 vgl. exemplarisch Freedom Outpost 2018: o.S., für 2020 siehe Beschwerdeschriften in 60+ Gerichtsverfahren auf Staaten- und Landesebene

[2] Für Belege der Verwurzelung von QAnon in Theorien der Präsidentschaftswahl 2020 siehe die Verbindung von Trump-Kampagnen-Anwälten zu diesen Kreisen, bspw. Newsweek 2020: o.S.

Propaganda zu finden und von ihren Diensten zu entfernen. Twitter, die Plattform der Wahl des Ex-US-Präsidenten Donald Trump bis zu seiner Sperrung, weitete diesen Fokus sogar jenseits russischer Beeinflussung auf generelle staatliche Beeinflussung aus und veröffentlicht seitdem in unregelmäßigen Abständen Berichte zu versuchter Beeinflussung durch Russland, aber auch durch andere Akteure und im Umfeld anderen Wahlen, wie beispielweise in der EU und ihren Mitgliedsstaaten (vgl. Twitter o.J.: o.S.).
Aufgrund der Tatsache, dass Twitter in so weitem Rahmen eigene Nachforschungen anstellt, und ihre Ergebnisse in Datensets der Öffentlichkeit verfügbar macht, ist es wohl die beste Plattform, um Desinformationskampagnen im Rahmen nationaler Wahlen zu untersuchen und unterschiedliche Staaten und Kampagnen in ihren Ansätzen miteinander zu vergleichen.

Die hier vorliegende Schrift befasst sich mit der Untersuchung des ersten dieser Twitter-Datensätze, der Beeinflussungsversuche der US-Wahl 2016 durch die russische *Internet Research Agency* (IRA), die auch von Robert Mueller angeklagt wurde (vgl. Time 2019: o.S.), zeigt. Mithilfe von *Machine Learning*-Ansätzen des *Structural Topic Modellings* (STM) sowie qualitativer Ergebnisinterpretation und basierend auf bereits vorhandener Forschung anderer Autoren wird die Twitter-Kampagne der IRA in ihrer Struktur und ihren Inhalten untersucht, um die versuchte Beeinflussung der US-Amerikanischen Öffentlichkeit sowie die Evolution der IRA-Kampagne über beinahe 9 Jahre an Twitter-Aktivität besser nachvollziehen zu können.
Dabei zeigt sich, dass sich die Methoden der IRA über ihre Twitter-Aktivitäten deutlich verändert haben: Zu Beginn wurde insbesondere versucht, über große Mengen inhaltlich sehr ähnlicher Tweets bestimmte Themen in Twitters Trends zu bekommen. Dabei wurden insbesondere Falschnachrichten zu katastrophalem Versagen öffentlicher Einrichtungen und Institutionen verbreitet. Nachdem keine dieser Kampagnen nennenswerten Erfolg erreichte, verschob sich der Fokus im weiteren Verlauf auf längerfristig angelegte Unterwanderung bestimmter Themenkomplexe, um beispielsweise die Lebenswelt schwarzer Amerikaner „von innen heraus“ zu beeinflussen. Zeitgleich zu diesen beiden Ansätzen entwickelten sich „Nachrichten“-Accounts, die über lange Zeiträume konstant in sachlicher Sprache zu aktuellen Geschehnissen berichteten. Da der größte Fokus dieser Nachrichtenaccounts auf lokalen Straf- und Gewalttaten lag, lässt sich vermuten, dass den Lesenden ein unterschwelliges Gefühl der Unsicherheit vermittelt werden sollte. Den größten Erfolg erzielte die IRA jedoch mit spät etablierten „persönlichen“ Accounts, die Meldungen zu Geschehnissen in den USA und der Welt teilten und mit ihrer „eigenen Meinung“ kommentierten.

Über alle behandelten Themen hinweg zeigt sich um August 2016 ein erkennbarer struktureller Umbruch, der im Folgenden mit einem deutlichen Anstieg der erhaltenen Interaktionen im Verlauf des Jahres 2017 einherging. Aufgrund dieser Entwicklung lässt sich zwar der Einfluss der IRA auf Twitter-Nutzer generell nicht abstreiten, inwiefern diese strukturelle Veränderung die öffentliche Meinungen im Hinblick auf die wenige Wochen später stattfindende US-Präsidentschaftswahl beeinflussen konnte, bleibt jedoch offen.

2. Bisherige Untersuchungen und Forschungsfokus

Aufgrund der Tragweite einer möglichen politischen Beeinflussung durch außerstaatliche Akteure haben sich bereits viele Forscherteams mit der Frage russischer Manipulation im Umfeld der US-Wahl 2016 beschäftigt. Untersuchungen stützen sich dabei auf eigens erhobene oder – wie in diesem Fall – von den Unternehmen veröffentlichte Datensets unterschiedlicher sozialer Medien, dominant jedoch auf Facebook- und Twitterdaten.

Während sich in Medienberichten und aus den Schlüssen der Mueller- und Kongress-Untersuchungen ein Bild der IRA-Manipulation als hochentwickeltes, komplexes System abzeichnet, sind einige Forscher anderer Meinung. Sie bewerten die IRA-Kampagnen eher als simplistische Versuche, durch schiere Mengen veröffentlichter Inhalte ihre Ziele zu erreichen, ohne ihre Herkunft zu verbergen. So zeigt sich insbesondere in den von der IRA aufgegebenen Facebook-Werbungen eine deutliche zeitliche Überlappung der Einreichung mit westrussischen Arbeitsstunden (vgl. Boyd et al. 2018: 2), mit derselben, aber weniger dominanten Tendenz für IRA-Tweets in einem proprietären Twitter-Datenset (vgl. ebd: 5).
Untersuchungen zu der Beeinflussung politischer Meinungen von US-Amerikanern kommen ebenfalls teilweise zu kritischeren Schlüssen: Ein Vergleich der Reichweite von IRA-Posts in 2017 mit einer zeitgleich ablaufenden Panelstudie politisch engagierter Twitter-Nutzer kommt zu dem Schluss, dass „we cannot determine if Russian trolls influenced candidate or media behavior or if they shaped public opinion in other ways“ (Bail et al 2020: 249), und dass die tatsächliche Reichweite der IRA möglicherweise geringer ist, als die puren Zahlen vermuten lassen, denn „even though active partisan Twitter users engaged with trolls at a substantially higher rate than reported by Twitter, the vast majority (80%) did not interact with an IRA account. And for those who did, these interactions represented a minuscule share of their Twitter activity – on average, just 0.1% of their liking, mentioning, and retweeting on Twitter“ (ebd.: 250.).

Das „Problem“ beider eben angesprochenen Untersuchungen ist jedoch, dass sie sich nur mit einem kleinen Teil der Daten befassen: Boyd et al. mit 1.200 Accounts im Gegensatz zu den 3.608 Accounts in den von Twitter veröffentlichten Daten, und Bail et al. mit einem Zeitfenster im späten 2017, zeitlich also relativ weit entfernt von politisch wichtigen Ereignissen wie der Präsidentschaftswahl 2016 und den Midterm-Wahlen 2018.

Forscher, die sich mit der Gesamtheit der IRA-Kampagnen über spezifische Themen- und Zeitbegrenzungen hinaus beschäftigen, teilen die Bewertung der IRA-Kampagne als simpel und ineffektiv jedoch kaum. Kriel und Pavliuc, die neben den englischsprachigen auch Teile der russischsprachigen Tweets untersuchten, sprechen sogar von einer „rigorous methodology of practice at work in Russia’s online interference in foreign democracies“ (Kriel/ Pavliuc 2019: 199). Ihnen zufolge liegt den IRA-Tweets von Anfang an eine tieferliegende Struktur zugrunde, mit deutlich wechselnden Aufgaben und Methoden je nach Erstelldatum des Accounts (vgl. ebd.: 222f.). Laut ihren Ergebnissen wurden viele der Accounts zuerst automatisiert betrieben, um mit „banalen“ Inhalten (bspw. Hashtags wie „#ifgooglewasagirl“) und Lokalnachrichten Follower zu gewinnen, bevor ein Mensch übernahm (vgl. ebd.: 206). Zusätzlich werfen sie auf Basis ihrer Ergebnisse die Frage auf, ob der von Twitter veröffentlichte Datensatz tatsächlich die Gesamtheit russischer Aktivitäten darstellt (vgl. ebd.: 212, 218f.).

Die erste koordinierte Aktivität des IRA-Netzwerkes laut Kriel und Pavliuc ist demnach der Versuch, im September 2014 die Falschmeldung zu verbreiten, dass im US-Bundesstaat Louisiana die Terrorgruppe ISIS ein Chemiewerk habe explodieren lassen (vgl. Kriel/ Pavliuc 2019: 208ff.). Das Verbreiten von Verschwörungstheorien war laut Analysen des Unternehmens yonder.ai auch zu späteren Zeitpunkten Teil der Aktivitäten der IRA. So verbreitete sie beispielsweise auf Facebook und Twitter Theorien über Hillary Clintons Gesundheit und brachten den Tod Seth Richs mit den von Wikileaks veröffentlichten DNC-Emails in Verbindung (vgl. Yonder 2018: 103)[3].

Kriel und Pavliuc stellen außerdem fest, dass neben dem Fokus auf rechte Pro-Trump-Inhalte auch linke Gruppierungen wie Black Lives Matter Ziel von IRA-Desinformation wurden (vgl. Kriel/ Pavliuc 2019: 223) – ein Schluss, zu dem andere Untersuchungen ebenfalls kommen. Forschung zu IRA-Aktivitäten auf Tumblr kommt beispielsweise zu dem Ergebnis, dass „[t]he IRA-linked Lagonegirl account on Tumblr

[3] Für eine Übersicht über die Verbindung zwischen Seth Rich und dem Democratic National Committee (DNC), sowie die Verbreitung dieser Verschwörungstheorie auf klassischen Medien wie Fox News siehe Daily Beast 2020: o.S.

included several performance elements consistent with IRA operations on Twitter […]. While there are similarities between Twitter performances and the Lagonegirl Tumblr account, there are also Tumblr-specific platform conventions and social norms that shaped Lagonegirl's performance" (Neill Hoch 2020: 10), ein weiteres Indiz für die Komplexität der IRA-Strategien in der Beeinflussung öffentlicher Meinungen.
Keyword-Analysen von Twitterposts zeigen zusätzlich, dass IRA-Accounts mit strukturell unterschiedlichen „Identitäten" (schwarz, weiblich oder muslimisch) aufgebaut wurden, die auf jeweils eigene Art und Weise auf die Kandidaten Trump und Clinton reagierten (vgl. Atkinson 2018: 8ff.), mit dem Ziel „to express and influence people based on manufactured identities" „[a]s partisanship and party sorting based on qualitative aspects of identity become increasingly common" (beide Atkinson 2018: 19). Interessanterweise sind dabei keinesfalls die Gruppen klassisch überwiegend demokratischer Wähler (Schwarze, Frauen, …) als pro-Hillary anzusehen – über alle „Identitäten" der IRA-Accounts hinweg schneidet Clinton schlechter ab als Trump, auch wenn Trump in vielen dieser Gruppen nicht nur als positiv angesehen wird (vgl. Atkinson 2018: 8ff., insb. 10). Einige dieser Identitäten wurden dabei mit der Zeit zu Marken weiterentwickelt, die ihr eigenes Ökosystem und Design mit sich brachten (vgl. Yonder 2018: 5).

Aus der hier präsentierten Literatur sowie der Möglichkeiten, die Twitter seinen Nutzern bietet, lassen sich folgende Hypothesen über Struktur und Entwicklung der IRA-Twitterkampagne ableiten:

1. Trotz der Menge an Postings und Accounts handelt es sich hierbei um eine nicht aufwändig versteckte Kampagne, sodass sich aus bspw. Posting-Zeiten und anderen Umständen Rückschlüsse auf einen russischen Ursprung ziehen lassen können.
2. Accounts nutzten mit wenig Aufwand verbundene Möglichkeiten Twitters, um für normale Nutzer als US-Amerikaner zu erscheinen, vorzugsweise aus politisch umkämpften Staaten
3. Je nach Erstelldatum des Accounts bestehen strukturelle Unterschiede in den Aktivitätszeiten sowie den geteilten Inhalten
4. Je nach anstehenden politischen Ereignissen in den USA (Wahlen, …) verändern sich Anzahl und Fokus der IRA-Aktivitäten, mit hoher Einheitlichkeit in verbreiteter Rhetorik für bestimmte Kandidaten bzw. diverse Verbreitung von auf die jeweilige Zielgruppe zugeschnittene Inhalte

5. Zu bestimmten Zeitpunkten immer wieder koordinierte Propagation von Verschwörungstheorien, um Vertrauen in politisches System (z.B. Ermordung Seth Richs/ Hillary Clintons Skandale) oder öffentliche Infrastruktur (z.B. Explosion eines Chemiewerks) zu schwächen
6. Es finden sich dominant rechte Inhalte, aber auch linke Accounts, die auf bestimmte Zielgruppen (schwarz/weiblich/muslimisch) zugeschnittene Inhalte verbreiten

3. Aufbau des Datensets und Struktur-Analysen

Der von Twitter veröffentlichte Datensatz in der Version vom 11.02.2019 umfasst 3.608 einzigartige Accounts, die insgesamt knapp 8,8 Millionen Posts (eigene Tweets sowie Retweets anderer Nutzer) über ihre Zeit auf Twitter veröffentlicht haben. In seiner ursprünglichen Version im Oktober 2018 umfasste der Datensatz 3.841 Accounts, aber weitere Untersuchungen kamen laut Twitter zu dem Ergebnis, dass 228 der Accounts nicht russischem, sondern venezolanischem Ursprung waren (vgl. Twitter 2018: o.S.) – was mit den fehlenden 5 Accounts geschehen ist, kann nicht nachvollzogen werden. Die veröffentlichten Datensets umfassen sämtliche öffentlich einsehbaren Posts der Nutzer zum Zeitpunkt ihrer Entfernung durch Twitter – von den Nutzern selbst gelöschte Inhalte und Medien finden sich jedoch nicht in den Daten (vgl. Twitter o.J.: o.S.). Dieser Umstand ist möglicherweise bedeutsam für die Bewertung der russischen Twitter-Kampagne, da ein ähnliches Problem wie bei der Analyse terroristischer Accounts in sozialen Netzen auftritt: „Researchers do not see what terrorists post. Rather, they see what is left after platform countermeasures are employed“ (Fishman 2019: 99). Es ist durchaus möglich, dass eine größere Menge an Postings veröffentlicht wurde, als sich hier in den Daten finden – und dass diese durch Nutzermeldungen oder andere Umstände bereits vor dieser Identifikation und Katalogisierung von IRA-Strukturen von Twitter gelöscht wurden.
Neben den Posting-Texten und -Metadaten finden sich alle mit diesen Postings und Accounts verbundenen Medien (Bilder und Profilbilder, Videos, Livestreams auf Twitters Periscope-Plattform, …), die in dieser Analyse jedoch nicht berücksichtigt werden. Die Namen und Twitter-IDs von Nutzern mit unter 5.000 Followern liegen zusätzlich in anonymisierter Form vor, was für die hier vorgenommenen Analysen jedoch nicht von Belang ist.

3.1 Account- und Tweet-Sprachen

Für jeden Account findet sich eine eingestellte Sprache in den Daten. Wenig überra-

schend sind die beiden häufigsten Einstellungen „Englisch“ und „Russisch“, aber auch andere Sprachen, insbesondere aus dem europäischen Raum (Deutsch, Britisch, Französisch, Spanisch, Italienisch und Ukrainisch) und in arabischer Sprache sind in kleineren Mengen vertreten. Es ist dabei anzumerken, dass die gewählte Account-Sprache sowohl von Twitter auf Basis der jeweiligen Aktivitäten automatisch gewählt, als auch von den Nutzern selbst eingestellt werden kann (vgl. Twitter Help o.S.). Eine Übersicht über die Mengenverteilung der Accountsprachen findet sich in *Tabelle 1*.

Sprache	**Anteil**
Englisch	66,02 %
Russisch	28,80 %
Deutsch	3,08 %
Andere	2,1 %

Tab. 1: Eingestellte Sprachen aller IRA-Twitteraccounts

Neben den automatisch bzw. manuell bestimmten Account-Sprachen gibt Twitter für jeden Tweet eine ermittelte Sprache an. Während die Account-Sprachen 11 unterschiedliche Kategorien aufweisen, finden sich 58 einzigartige Tweet-Sprachenlabels, mit einmaligem Auftreten der Sprachen Uighurisch, Pashto und Kiluba. Es ist jedoch zu erwähnen, dass es sich dabei ebenso gut um einen Tweet in diesen Sprachen wie um eine Misidentifikation durch den Twitter-Algorithmus handeln kann.

Bemerkenswert ist, dass – obwohl nur 28,8% der Accounts Russisch als Sprache angeben – mehr als die Hälfte aller Posts (55,35%) in russischer Sprache sind, mit Englisch bei knapp über einem Drittel aller Tweets (34,18%) auf Platz 2. Die Desinformationskampagnen der IRA scheinen sich demnach mindestens genau so sehr auf innerstaatliche Beeinflussung wie auf Beeinflussung des englischsprachigen Raums zu konzentrieren. Dieser Umstand wirft die Frage auf, ob die beiden Kampagnen (Russisch und Englisch) voneinander getrennt vollzogen wurden, oder ob dieselben Accounts in beiden Kampagnen zum Einsatz kamen. Ein solcher Umstand würde für die Hypothese einer simpel aufgebauten und leicht zu identifizierenden Aktion sprechen.

Eine genauere Betrachtung der Daten zeigt jedoch, dass es kaum Accounts gibt, die sowohl auf Englisch als auch auf Russisch tweeten – und dass selbst in diesen Fällen eine der beiden Sprachen mit meist über 90% der jeweiligen Tweets dominiert (vgl. *Abb. 1*). Die Unterschiede in Account- und Tweet-Sprachen scheinen sich demnach systematisch darauf zurückführen zu lassen, dass Accounts mit dominant russischsprachigen Posts Englisch als Account-Sprache eingestellt haben. Weitere Analysen der jeweiligen Posts bestätigen dies: Von 2,1 Mio. Posts mit Unterschieden zwischen Tweet- und Account-Sprache haben 1,73 Mio. englische Account- und 1,45 Mio. rus sische Tweet-Sprache eingestellt.

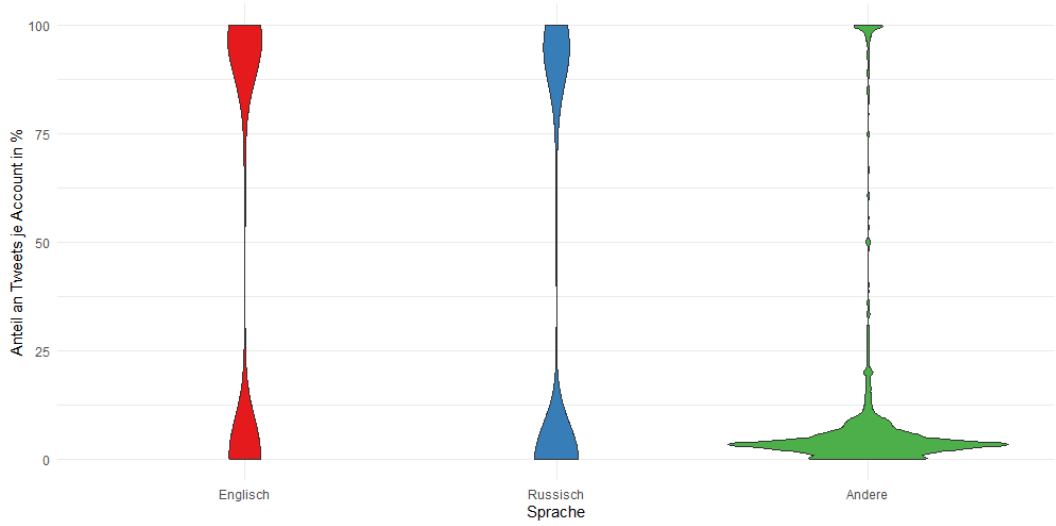


Abb. 1: Verteilung englisch-, russisch- und anderssprachiger Tweets als Anteile an allen Tweets des jeweiligen Accounts über alle 3.608 Accounts

Filtert man die Daten nach von Twitter als Englisch identifizierten Tweets, so verbleiben 3 Mio. Posts. Die Anzahl an Accounts mit mindestens einem englischsprachigen Tweet verändert sich dagegen wenig, von 3.608 auf 3.077. Dieser Umstand spricht dafür, dass sich eine große Menge Accounts in der eben identifizierten „1%-25% englischsprachige Posts"-Gruppe befinden. Ein weiterer Filter nach Accounts mit über 50% englischsprachigen Postings bestätigt dies: Die Anzahl an Nutzern sinkt auf 1.854 ab, während sich die Posting-Zahl nur um knapp 83.000 verringert. Da die Accounts mit unter 50% englischsprachigen Postings jedoch koordiniert in einem bestimmten Zeitintervall Ende 2014 einen großen Teil ihrer Postings absetzen (vgl. *Abbildung. 2*), werden sie weiter in den Daten behalten und analysiert, und die 417

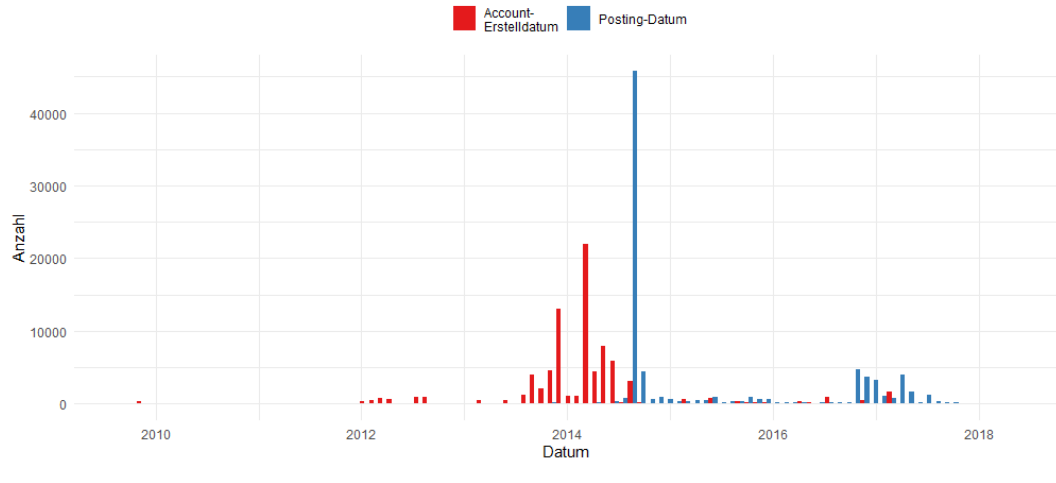


Abb. 2: Histogramm der Erstelldaten sowie Posting-Zeiten aller Tweets von Accounts mit mind. einem englischsprachigen Tweet und englischsprachigem Anteil von unter 50% an allen Tweets des Accounts

Accounts mit Tweets zwischen Juli und September 2014 gesondert für die Untersuchung eventueller Themen in diesem Zeitraum protokolliert.

3.2 Ortsangaben

Twitter-Nutzer haben die optionale Möglichkeit, selbst einen Standort anzugeben. Der angegebene Ort wird auf ihrer Profilseite zusammen mit anderen Informationen jedem Nutzer angezeigt. Diese Ortsangaben müssen dabei nicht unbedingt real existent sein: Der Twitter-Account von Joe Biden (@JoeBiden) gibt beispielsweise seinen Wohnsitz „Wilmington, DA“ als Ort an, während die Twitter-Seite des Videospiels *DOOM* (@DOOM) „Hell“, also Hölle, als Standort angibt (beide Angaben Stand Februar 2021). Obwohl die Ortsangabe freiwillig ist, haben nur 480 der 3.077 IRA-Accounts keinen Standort angegeben. Manuelles Vereinheitlichen und Filtern der uneinheitlichen Angaben zeigt folgende Trends:

1. 1.501 Accounts geben die USA bzw. einen Ort in den USA als Standort an, während 825 Accounts Russland bzw. einen Ort in Russland angeben. Aus 44,6% der US-Angaben lässt sich ein Bundesstaat ableiten
2. Während sich genauere Ortsangaben auf russischer Seite auf Moskau (318) und St. Petersburg (148) konzentrieren, verteilen sich genauere US-Angaben weit auf unterschiedlichste Staaten.
3. In den angegebenen US-Bundesstaaten lässt sich eine leichte Annäherung an die Bevölkerungsverteilung mit den Ausreißern Louisiana und Massachusetts erkennen. Insbesondere im Jahr 2016 ist zudem außer Florida und möglicherweise Georgia keiner der Top-Bundesstaaten politisch als „Swing State“ einzuordnen (vgl. *Tabelle 2*).

US-Staat	Accounts	Bev.-Rang	Wahl 2016	Wahl 2020
New York	150	4	Clinton +21,3	Biden +23,1
Georgia	100	8	Trump +5,7	Biden +0,2
Kalifornien	59	1	Clinton +28,3	Biden +29,2
Louisiana	39	25	Trump +19,7	Trump +18,6
Texas	35	2	Trump +9,2	Trump +5,6
Florida	28	3	Trump +1,3	Trump + 3,3
Massachusetts	25	16	Clinton +27,3	Biden +33,6

Tab. 2: Am häufigsten angegebene US-Bundesstaaten (N≥25) im Vergleich mit ihrem tatsächlichen Bevölkerungsrang und dem jeweiligen Wahlsieger 2016/2020 mit seinem Vorsprung auf Platz 2 in Prozent (Daten: World Population Review 2021: o.S., Politico 2016: o.S., Politico 2021: o.S.)

3.3 Account-Erstelldaten

Der älteste Account in den Daten wurde am 24.04.2009 erstellt, der jüngste Account am 03.04.2018. Das Datenset umfasst somit beinahe neun Jahre an IRA-Aktivitäten auf Twitter. Eine Betrachtung der Account-Erstelldaten gestaffelt nach Quartalen zeigt, dass der größte Anteil der IRA-Accounts (60,97%) im Zeitraum zwischen dem 01.07.2013 und dem 30.06.2014 erstellt wurden (vgl. *Abbildung 3*). Zusätzlich lässt sich festhalten, dass – zumindest bei Accounts mit mindestens einem englischsprachigen Tweet – bis Ende 2015 beinahe nur Englisch und Russisch als Account-Sprachen auftauchen, während ab dem vierten Quartal 2015 auch zunehmend andere Spracheinstellungen, insbesondere aus dem europäischen Raum, auftauchen.

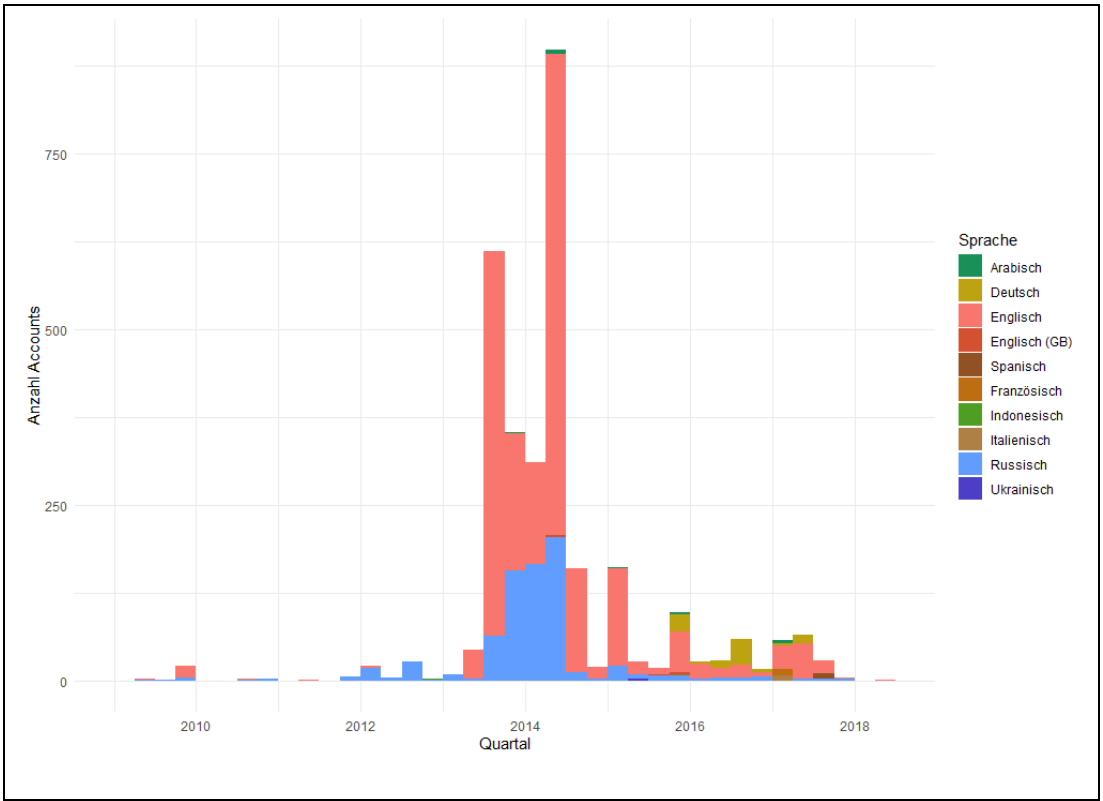


Abb. 3: Erstellzeiten aller IRA-Accounts mit mind. einem englischsprachigen Tweet, gegliedert nach Erstell-Quartal und eingestellter Account-Sprache

3.4 Account-Aktivitäten

Wie bereits von Kriel und Pavliuc festgestellt, gibt es Accounts mit langen Inaktivitätszeiten vor ihren ersten Posts – insbesondere Accounts, die im Jahr 2013 erstellt wurden (vgl. Kriel/ Pavliuc 2019: 222f.). *Abbildung 4* zeigt die einzelnen Accounts aufgeschlüsselt nach Erstelldatum, Tagen an Aktivität (Distanz zwischen erstem Post und letzten Post), und Inaktivität (Distanz zwischen Erstelldatum und erstem Post). Betrachtet man die ersten erstellten Accounts aus den Jahren 2009/2010, so fällt auf, dass diese entweder über sehr lange Zeit nicht genutzt wurden, und/oder seit ihrer

ersten Nutzung über lange Zeiträume immer wieder aktiv waren. Während eine große Anzahl an Accounts aus allen Erstelljahren Aktivitätszeiten von über einem Jahr aufweisen, zeigt sich bei der Inaktivitätszeit zwischen Accounterstellung und erstem Post ein klarer Trend: Accounts aus 2013 weisen in großen Mengen über 500 Tage Inaktivität auf (was sich mit den Erkenntnissen von Kriel und Pavliuc deckt), Accounts vor 2015 weisen dominant mindestens 70 Tage an Inaktivität auf, und Accounts insbesondere ab Mitte 2016 werden meist direkt wenige Tage nach Erstellung aktiv.

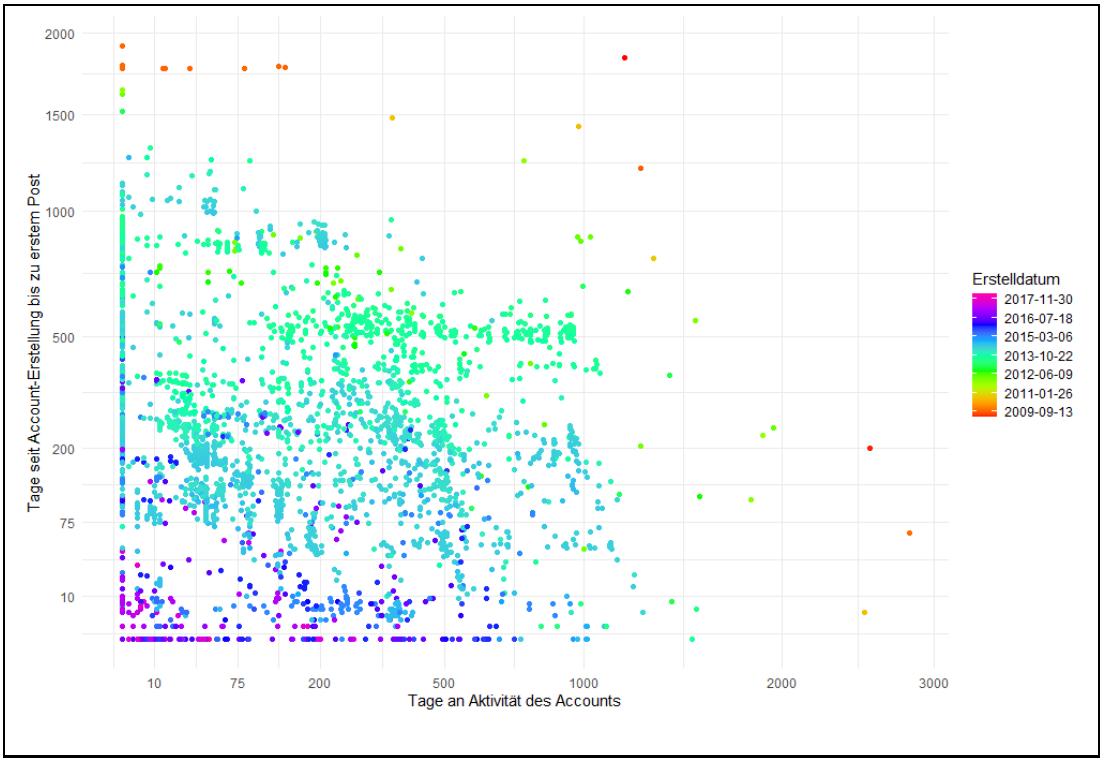


Abb. 4: Gesamt-Aktivitätszeit, Inaktivität zwischen Account-Erstellung und erstem Post, sowie Erstelldatum für alle 3.077 IRA-Acounts.

3.5 Following und Follower

Das Folgen oder „Following“ anderer Nutzer ist die dominante Form langfristiger Interaktion auf Twitter. Folgt man einem Nutzer, so erscheinen alle seine Tweets sowie Retweets auf der eigenen Startseite. Aus den eigenen *Follows* lassen sich demnach Rückschlüsse auf persönliche Ansichten und Vorlieben schließen, während die Anzahl eigener *Follower* etwas über die persönliche Reichweite aussagt. Während die Anzahl an angesammelten Followern theoretisch unbegrenzt steigen kann, ohne die eigene Nutzung von Twitter zu beeinflussen, ist für die eigenen Follows eine Art weiche Obergrenze („weich“ aufgrund der Tatsache, dass nicht jeder Account mit derselben Frequenz Posts veröffentlicht) anzunehmen, ab der zentrale Twitter-Funktionen wie die Startseite nicht mehr wirklich nutzbar sind, da die neu veröffentlichten Postings von unter Umständen mehreren tausend Accounts dort angezeigt werden

könnten. Diese Beobachtung zeigt sich auch in den Twitter-Daten. Während jedoch die meisten Accounts mehreren hundert anderen Accounts folgen (Median: 266 Follows, Quartile: 122 bzw. 583 Follows), gibt es vereinzelte Accounts mit deutlich höheren Follow-Zahlen – der IRA-Account mit den meisten Follows folgt 74.664 anderen Accounts.

Die Menge an gefolgten Accounts lässt vermuten, dass die IRA für einige ihrer Accounts auf Follow-Bots oder ähnlich funktionierende Programme gesetzt hat, um ihre eigenen Follower-Zahlen zu erhöhen. Diese Bots funktionieren nach einem simplen Prinzip: Basierend auf festgelegten Input-Variablen (Themenpräferenzen, Posting-Verhalten, …) werden Accounts identifiziert, die der jeweilige Nutzer als Follower seines Accounts gewinnen will. Der Bot folgt den so identifizierten Accounts mit dem eigenen Account des Nutzers, um sie auf sich aufmerksam zu machen. Followen die identifizierten Accounts nicht zurück, so entfolgt der Bot diesen nach einem gewissen Zeitraum wieder (vgl. Postfity 2020: o.S.). Diese Bots sind bereits seit Jahren auf Twitter im Einsatz, und der Autor kann anekdotisch berichten, dass ihm (vermutlich deswegen) bereits mehrere Accounts ge- und kurz darauf entfolgt sind, unter anderem im Januar 2021 ein Service, der alte Nutzeraccounts sozialer Medien verkauft und vor mehreren Jahren der Twitter-Account des fiktiven Marvel-Superhelden Tony Stark.

Sollten diese IRA-Accounts tatsächlich Bots benutzt haben, um ihre Followerzahlen zu erhöhen, so sollte sich dies mit einer linearen Regression überprüfen lassen, da durch die eben beschriebene Natur dieser Bots die Anzahl eigener Follows eine relativ genaue Vorhersage der jeweiligen Followerzahlen sein sollte. Es stellt sich jedoch die Frage, ab welcher Follower-Zahl die Nutzung von Bots vermutet werden kann. Zu diesem Zweck wurden mehrere Analysen für N eigene Follower durchgeführt, wobei $500 \leq N \leq 5.000$ in 100er-Schritten erhöht wurde. Für jedes N wurden zwei lineare Regressionen berechnet: Regression 1 bestimmt den linearen Zusammenhang zwischen eigenen Follows und Followerzahlen für Accounts unter N Follows, und Regression 2 bestimmt den linearen Zusammenhang für Accounts mit über oder gleich N Follows. Die Ergebnisse in *Abbildung 5* zeigen, dass ab etwa 2.800 Follows davon ausgegangen werden kann, dass Follow-Bots genutzt wurden, da die Varianzaufklärung an diesem Punkt maximal ist, und von dort aus linear nachlässt. Für Accounts über 2.800 Follows erklärt die Anzahl gefolgter Accounts mehr als 71% der Varianz in den jeweiligen Followerzahlen, während die Aufklärung für Accounts unter 2.800 Follows bei knapp über 1% liegt.

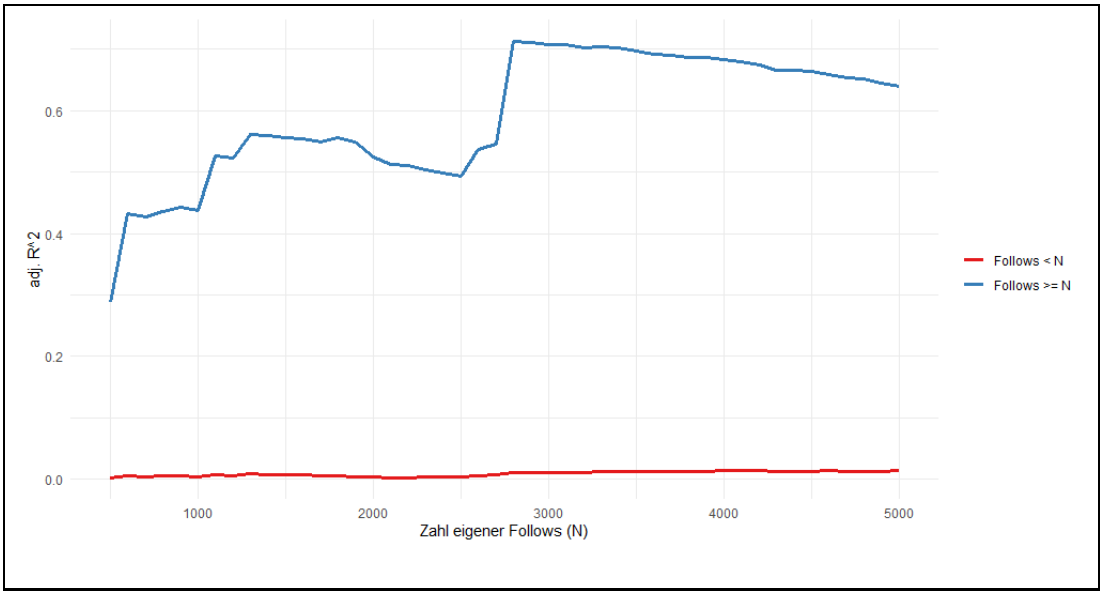


Abb. 5: Aufklärung der Varianz (angepasstes R^2) in den Followerzahlen durch eigene Follows für Accounts mit über bzw. unter N Follows, für N zwischen 500 und 5.000

Diese Strategie scheint dabei eine der dominanten Möglichkeiten gewesen zu sein, mit der IRA-Accounts eigene Follower gewinnen konnten. Während viele der Accounts kaum oder nur in sehr geringem Maß Follower ansammeln konnten – 188 Accounts haben weniger als 10 Follower und 1.091 Accounts weniger als 100 Follower – haben von den Accounts mit über 1.000 Followern 38,76% (169 Accounts) über 2.800 eigene Follows, und von den 300 Accounts mit den meisten Followern sind es sogar über die Hälfte (159 Accounts). Zusätzlich ist zu erwähnen, dass die Follower-Listen der jeweiligen IRA-Accounts nicht einsehbar sind, es besteht also durchaus die Möglichkeit, dass sich einige dieser Accounts untereinander folgten, und somit die Zahl „echter" Follower niedriger als die hier angegebene lag.

3.6 Tweet-Zeiten

Für jeden Post ist auf die Minute genau die jeweilige Veröffentlichungszeit angegeben. Die Zeiten sind dabei in koordinierter Weltzeit (UTC) angegeben, und somit leicht auf andere Zeitzonen übertragbar. Betrachtet man alle Postings nach Stunden gruppiert (12 Uhr ≙ Tweet zwischen 12:00 und 12:59 gepostet), und vergleicht dies mit den von Boyd et al. identifizierten Zeiten für Twitter-Posts (11 bis 21 Uhr Moskau-Zeit, vgl. Boyd et al. 2018: 5) in den Zeitzonen Westrusslands sowie übertragen auf West- und Ostküste der USA, ergibt sich das Bild in *Abbildung 6*.

Es zeigen sich deutliche Trends in den Uhrzeiten. So werden die meisten Tweets zwischen 17 und 20 Uhr Moskau-Zeit (MSK), bzw. 9-12 Uhr an der US-Ostküste (EST) und 6-9 Uhr an der US-Westküste (PST) veröffentlicht, mit Spitzenwerten von etwa dem Doppelten des Medians über alle Uhrzeiten. Die wenigsten Tweets

hingegen wurden zwischen 6 und 9 Uhr MSK, bzw. 23-2 Uhr EST und 20-23 Uhr PST veröffentlicht, das Minimum liegt bei etwa der Hälfte des Medians.

Die Tatsache, dass die identifizierten Maxima und Minima in den Tweetzeiten sich so eindeutig auf bestimmte Intervalle legen lassen, spricht gegen eine Betreibung dieser Accounts aus den USA. Bei angenommener Verteilung der Nutzer über die Fläche der USA (bzw. den tatsächlich existierenden Bevölkerungs-Konzentrationen an Ost- und Westküste) ist zu erwarten, dass sich aufgrund der drei Stunden Zeitunterschied keine so eindeutigen Maxima und Minima identifizieren lassen. Während ein Ursprung der Tweets aus den USA somit unwahrscheinlich erscheint, ist die Zeitverteilung gleichzeitig kein eindeutiger Beweis für einen russischen Ursprung.

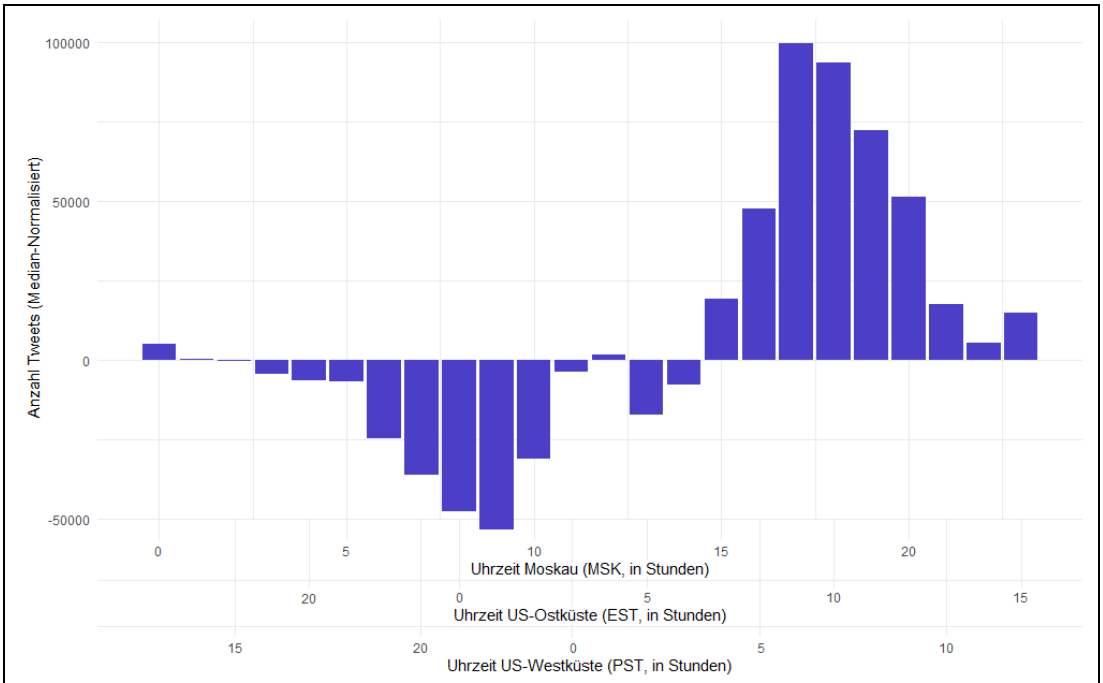


Abb. 6: Uhrzeiten und Menge veröffentlichter Postings in den Zeitzonen Westrusslands sowie der US-Ost- und -Westküsten, normalisiert nach der Medianzahl (Median = 116.902,5) veröffentlichter Posts über alle Uhrzeiten

3.7 Tweet- und Retweet-Zahl

Retweets anderer Nutzer erscheinen ebenso wie eigene Tweets auf den Startseiten der eigenen Follower. Sie sind somit eine gute Methode, bereits vorhandene Aussagen an das eigene „Publikum“ weiter zu verbreiten, was die IRA auch in großen Mengen tat – von den knapp 3 Mio. veröffentlichten Postings sind 1,1 Mio. Retweets bereits vorhandener Inhalte. Während der Anteil an Retweets je Account dabei das gesamte Spektrum abdeckt, scheint es zwei Gruppen zu geben, die jeweils beinahe ausschließlich (>80% aller Postings) auf eigene Tweets bzw. Retweets vorhandener Inhalte setzen.

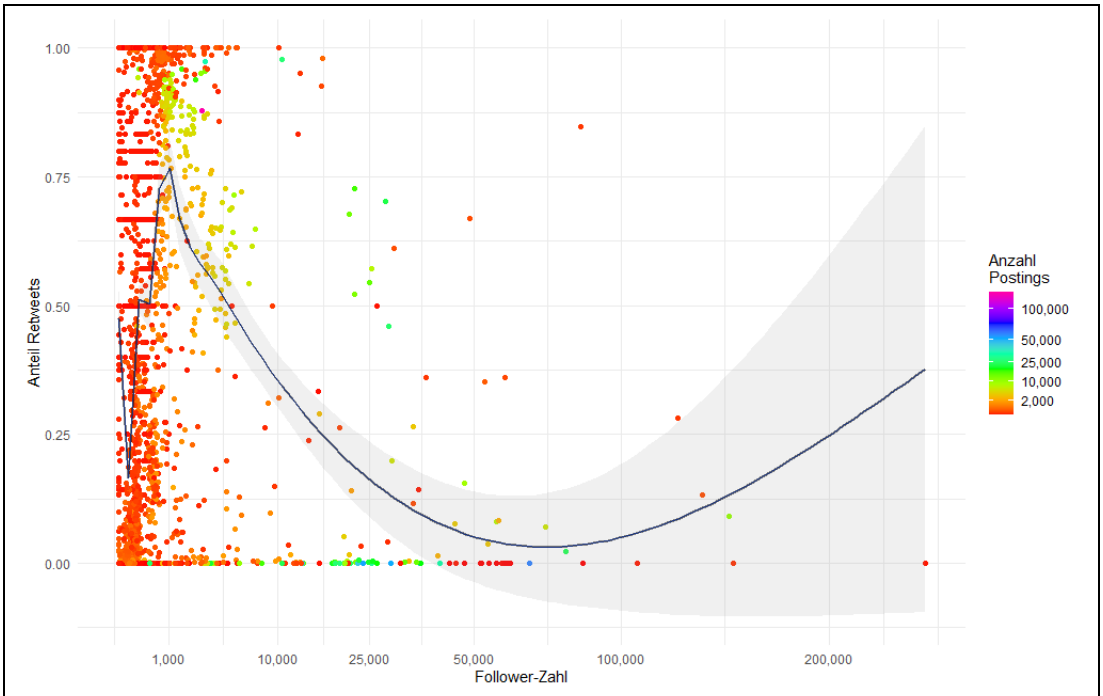


Abb. 7a: Anteil Retweets aller 3.077 Accounts nach Follower-Zahl (Quadratwurzel-transformiert) und Gesamtzahl veröffentlichter Posts des Accounts (log-transformiert), sowie lokale Regression mit Konfidenzintervallen

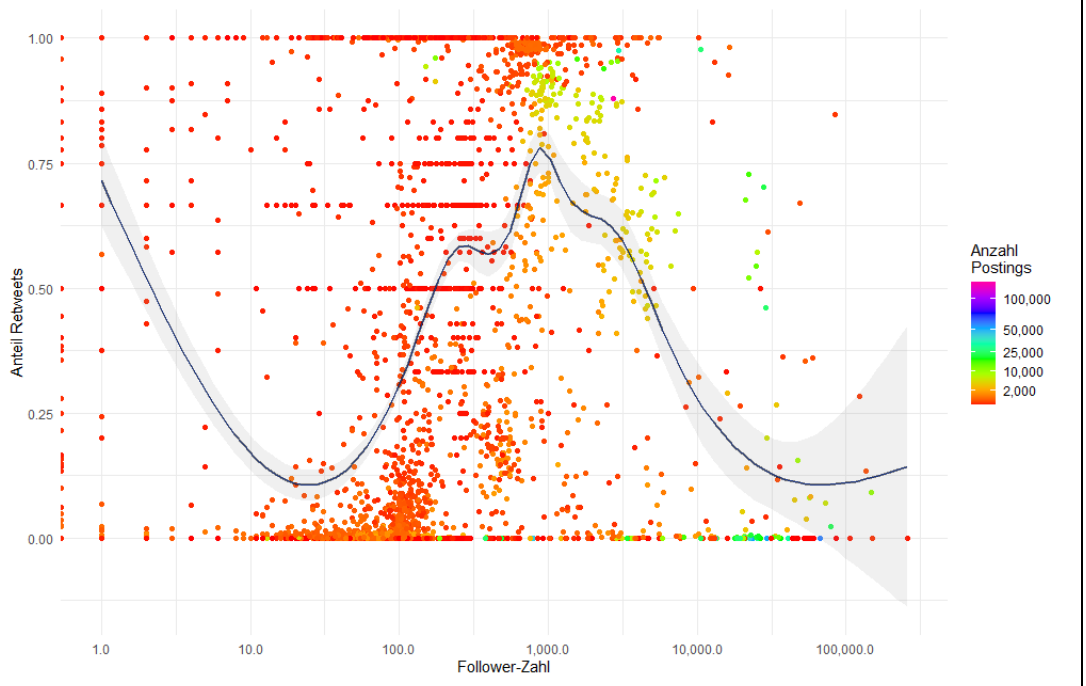


Abb. 7b: Anteil Retweets aller 3.077 Accounts nach Follower-Zahl (log-transformiert) und Gesamtzahl veröffentlichter Posts des Accounts (log-transformiert), sowie lokale Regression mit Konfidenzintervallen

Betrachtet man den Retweet-Anteil in Verbindung mit der jeweiligen Anzahl Follower und der Gesamtzahl veröffentlichter Postings, ergeben sich *Abbildungen 7a* und *b*.

Auffällig ist, dass die beiden identifizierten Gruppen sich um gewisse Followermengen sammeln: Accounts mit weniger als ~15% Retweets finden sich in einer großen Gruppe um die 100 Follower, während Accounts mit über ~85% Retweets sich um die 1.000 Follower sammeln. Zusätzlich findet sich ein großer Teil der Nutzer mit ca. 10.000 veröffentlichten Posts bei über 50% Retweets und zwischen 1.000 und 10.000 Followern wieder. Jenseits der 10.000 Follower-Marke tauchen dominant Accounts mit wenig bis gar keinen Retweets auf.

Es lassen sich mehrere Erkenntnisse ableiten:
1. Accounts, die beinahe ausschließlich auf Retweets als Inhalte setzen, scheinen im Schnitt auch bei geringer Gesamt-Postingzahl beliebter gewesen zu sein als Accounts mit eigenen Tweets (~1000 Follower vs. ~100 Follower). Aus der Tatsache, dass dominant Accounts retweetet wurden, die sich nicht im Datensatz finden (945.475 Retweets, oder 87,3%), lässt sich dabei ableiten, dass Tweets „echter" Nutzer die jeweiligen Zielgruppen mehr überzeugten und gewinnen konnten als von der IRA selbst verfasste Tweet-Texte.
2. Der große Anteil an Accounts mit über ~50.000 Followern hat nur wenig bis gar keine Tweets selbst veröffentlicht. Eine genauere Analyse der Top-20 Accounts zeigt, dass es sich hierbei um Überreste der russischsprachigen Desinformationskampagnen handelt, die aufgrund der getroffenen Sprachfilter-Entscheidungen (siehe S. 10) über vereinzelte als englischsprachig erkannte Tweets in den Daten verbleiben. So finden sich in diesen 20 Accounts neun mit dem Zusatz „Novosti" – russisch für „Nachrichten" – in ihrem Accountnamen und zwischen einem und maximal 36 englischsprachigen Tweets, im Vergleich zu mehreren tausend Tweets der anderen Top-Accounts.

4. Grundlagen und Aufbau des Structural Topic Models

Mit mehreren Millionen Tweets von unterschiedlicher Länge ist dieses Datenset nicht für eine manuelle Kodierung geeignet. Auch ist es aufgrund der erwarteten inhaltlichen und thematischen Vielfalt nicht möglich, ein manuell bearbeitbares und gleichzeitig repräsentatives Sample zu ziehen. Stattdessen wird in dieser Untersuchung mithilfe eines *Structural Topic Models*, einem *Machine Learning*-Ansatz, die Gesamtheit der Tweet-Texte auf vorhandene Strukturen und inhaltliche Konstanten untersucht.

4.1 STM

Das *Structural Topic Model* ist eine Weiterentwicklung probabilistischer Themenmodelle, die die Möglichkeit bietet, Metadaten in der Strukturfindung zu berücksichtigen. Basierend auf den verwendeten Worten, deren Anzahl und Verhältnis, sowie den angegebenen Metadaten, werden eine im Voraus definierte Anzahl an Themengruppen (K) aus Worten mit hoher Koexistenzrate gebildet (vgl. Roberts et al. 2019: 1f.). Aufgrund dieser Strukturfindung lassen sich die entstehenden Themengruppen (im Folgenden Topics genannt) als Abbildungen der thematischen Dimensionen in den Originaldokumenten interpretieren, die aus dem Modell heraus für optimale Beschreibung und Aufteilung der Ausgangsdaten erstellt wurden (vgl. Munoz-Najar Galvez et al. 2020: 622). Diese Eigenschaft der Analysemethode ist es, die STMs nützlich für die Durchführung explorative Analysen inhaltlich diverser Dokumente, sowie die Identifikation vorhandener Strukturen macht: Alle entstandenen Gruppen weisen bestimmte Themen bzw. Wortstrukturen auf, die sich inhaltlich sinnvoll interpretieren und aus denen sich nach Rückbezug auf die Originaldokumente bzw. Metadaten Ableitungen über versteckte Strukturen und Formulierungsgruppen treffen lassen sollten, die sonst möglicherweise unbeachtet oder nicht identifiziert geblieben wären. Aufgrund der Größe des hier verwendeten Textkorpus ist dabei zusätzlich davon auszugehen, dass je nach gewählter Topic-Zahl jedes Topic eine inhaltliche Relevanz aufweist und zufällige, natürlich vorkommende Häufigkeitsunterschiede der einzelnen Worte kaum Einfluss auf die Ergebnisse aufweisen.
Basierend auf den definierten Topics wird jedem Text ein berechneter Übereinstimmungsanteil für jedes Topic (*Theta*) zugewiesen. Diese Anteile geben den jeweiligen Grad der Übereinstimmung mit den für die Topics relevanten Worten und Strukturen auf einer Skala von 0 (keine Übereinstimmung) bis 1 (vollständige und exklusive Übereinstimmung) an. Die Summe der Theta-Werte über alle K Topics beträgt dabei für jedes Dokument 1, sodass sich im Vergleich der Werte Aussagen über die jeweilige Konfidenz in die Zuordnung von Dokumenten zu bestimmten Topics, und strukturelle Ähnlichkeiten unterschiedlicher Topics und Topic-Korrelationen über alle Dokumente hinweg treffen lassen.

In der hier durchgeführten Analyse wurden für das Modell neben den Tweet-Texten als zu untersuchende Dokumente das jeweilige Veröffentlichungsdatum sowie die vier Maße der Interaktion mit den Tweets (jeweilige Anzahl an Likes, Retweets, Antworten und Zitierungen) als Metadaten in der Analyse berücksichtigt. Diese Auswahl der Metadaten stützt sich auf drei zentrale Annahmen:
1. Die IRA verfolgt zu gewissen Zeiten gewisse Ziele, die sich über alle Accounts –

bzw. über alle Accounts mit derselben Zielgruppe – gleichen sollten. Mit Veröffentlichungsdaten als Metadaten sollte das Modell den zeitlichen Verlauf der Postings berücksichtigen und diese veränderlichen Ziele und Methoden voneinander trennen. Zusätzlich lässt sich so eine mögliche Evolution in verwendeter Sprache, Komplexität oder inhaltlichen Zielen nachvollziehen.

2. Da alle diese Accounts der Annahme nach von der IRA aus koordiniert wurden, ist die Unterscheidung in konkrete Accounts anhand von jeweiligen IDs nicht von primärer Relevanz, da alle Erkenntnisse eines Accounts allen anderen Accounts zur Verfügung stehen und alle Inhalte und Themen zentral gesteuert werden. Die einzigen zu erwartenden Unterschiede zwischen den Accounts sind somit die jeweilige Zielgruppe bzw. verwendete Sprachmuster – und diese sollte ein *Structural Topic*-Modell auf Basis der den Berechnungen zugrunde liegenden Methoden von sich aus automatisch voneinander trennen.

3. Es ist zu vermuten, dass die IRA das Ziel verfolgte, so viele Nutzer wie möglich mit ihrer Desinformation zu erreichen. Demnach ist es wichtig, die jeweilige Reichweite der Posts in Verhältnis zueinander zu setzen, um nach Themen zu suchen, die strukturell viel Reichweite erreichten. Aufgrund der unterschiedlichen Natur der vier Twitter-Interaktionsformen[4] ist es dabei wichtig, diese getrennt voneinander zu behandeln, anstatt sie zu einer einzigen Maßzahl aufzusummieren.

4.2 Datenvorbereitung

Für eine maschinelle Auswertung der Daten muss zuerst ein gewisses Maß an Einheitlichkeit in den Input-Dokumenten herrschen. Aus diesem Grund wurden die verwendeten Postings mit mehreren Methoden überarbeitet und vereinheitlicht. Hierbei wurde unter anderem die Entscheidung getroffen, Retweets nicht für die STM-Analyse zu berücksichtigen.

[4] Die vier Interaktionsformen auf Twitter (Likes, Retweets, Antworten und Zitierungen) unterscheiden sich deutlich in Aufwand und Sichtbarkeit. Likes und Retweets sind jeweils nur einen bzw. zwei Knopfdrücke an Aufwand für den Nutzer, während Antworten und Zitate eigene Texte des jeweils reagierenden Nutzers beinhalten.

Zusätzlich unterscheiden sich die vier stark in ihrer Sichtbarkeit für eigene Follower: Likes sind – solange man sie nicht über das Profil aktiv aufsucht – für eigene Follower nicht sichtbar, und Antworten erscheinen den eigenen Followern im Normalfall nur, wenn sie von sich aus den beantworteten Post öffnen (über Algorithmus-Gewichtung der Antworten auf diesen Post), oder sowohl dem antwortenden als auch dem beantworteten Nutzer folgen. Retweets und Zitate erscheinen dagegen wie eigene Tweets auf der Startseite der Follower, und sind somit deutlich leichter auffindbar und besser zur Verbreitung von Posts geeignet.

Zwar würde die Betrachtung aller Postings (Tweets sowie Retweets) ein vollständigeres Bild der IRA-Kampagne liefern – insbesondere im Hinblick auf die scheinbar strukturellen Unterschiede in den Followerzahlen je nach Retweetanteil, die bereits identifiziert wurden – aber gleichzeitig können über Retweets einzelne Tweets häufiger in den Daten auftauchen. Sollte sich beispielsweise herausstellen, dass mehrere Accounts denselben Ursprungstweet retweetet haben, so wäre dieser Tweet für jeden Account in den Daten als einzelner Fall hinterlegt. Die Struktursuche des STM würde in diesem Fall aus einem einzelnen, mehrfach auftauchenden Tweet eine Themenstruktur ableiten, die so in der Wirklichkeit nicht gegeben ist. Aus diesem Grund werden die rund eine Million Retweets in dieser STM-Analyse nicht beachtet, und nur von den IRA-Accounts selbst geposteten Tweets analysiert.

Ein ähnliches Interpretations-„Problem“ stellen wortgleiche Tweets dar. Aufgrund der perfekten Übereinstimmung und somit hohen semantischen Kohärenz wortgleicher Tweets behandelt der STM-Algorithmus diese schon bei geringen Mengen als eigenes Topic. Je nach Anzahl gesuchter Topics (K) kann es also passieren, dass wortgleiche Tweets die einzigen Faktoren der Topic-Unterteilung darstellen, und der Erfahrungsgewinn eines solchen Modells mit simpleren Duplikats-Suchen identisch ist. Mit knapp 600.000 Tweets, die zumindest einem anderen Tweet im Wortlaut exakt gleichen, ist die Menge an Duplikaten dabei so groß, dass auch bei Tests mit über 100 Topics beinahe nur Duplikate die Topic-Unterteilung definierten.
Um das Problem der Duplikat-Dominanz zu lösen, ohne die in diesen Duplikaten vertretenen Themen sowie deren Reichweite zu sehr zu verfälschen, wurden alle Duplikate entfernt, deren Text wortgleich in einem Tweet mit höherer Interaktions-Zahl (gemessen als Summe aus Zitaten, Antworten, Likes und Retweets) vorkommt und die selbst eine Interaktions-Zahl von unter 100 vorweisen. Auf diese Weise verbleibt mindestens ein Exemplar aller Tweets in den Daten, und es werden nur Duplikate entfernt, die aufgrund geringer Reichweite kaum Einfluss auf „echte“ Twitter-Nutzer ausübten. Aus den knapp 600.000 Duplikaten verbleiben so nur ca. 94.000 Tweets als erfolgreichste Version des jeweiligen Textes sowie ca. 2.000 Tweets mit weniger Erfolg aber einer Interaktionszahl von 100 oder mehr. Es sei hierbei festzuhalten, dass die Wahl der Interaktionszahl-Grenze nicht auf Basis faktischer Daten erfolgt, und es keinen objektiv besten Wert gibt – eine dreistellige Zahl erschien dem Autor lediglich als annehmbarer Kompromiss zwischen Beeinflussung des STM und erreichter Reichweite.

Weitere Datenvorbereitungsmethoden entfernen inhaltlich für die hier vorgenommene Analyse wenig relevante Textpassagen wie die Accountnamen anderer Twitter-

Nutzer und Links zu externen Inhalten, um Beeinflussungen der Textmustersuche zu minimieren, und vereinfachen die verwendete Sprache durch Reduzierung der Worte auf ihre Wortstämme, um die Suche nach Wortähnlichkeiten zu verbessern. Ebenso wurden verwendete Emoji recodiert, um Fehler zu verhindern und Verzerrungen des Algorithmus‘ entgegenzuwirken[5].

4.3 Suche nach K

Wie bereits angesprochen, ist ein STM von einer im Voraus festgelegten Topic-Menge K abhängig, auf die die Texte aufgeteilt werden. Es gibt dabei jedoch keine „perfekte“ Anzahl an K, da die Kernqualitäten der Topics in Konflikt miteinander stehen: So ist beispielsweise die Exklusivität der Topics – also der Umstand, dass Worte eines Topics nicht bzw. nur selten in anderen Topics vorkommen – theoretisch maximal, wenn jedes Wort sein eigenes Topic ist, während die semantische Kohärenz eines Topics – also die Wahrscheinlichkeit, mit der Worte aus demselben Topic in einem Tweet auftauchen – theoretisch bei wenigen Topic maximal ist. Die „Grenze“ zwischen Kohärenz- und Exklusivitätseinfluss ist dabei eine der dominanten Richtlinien für die Entscheidung für oder gegen ein bestimmtes Modell (vgl. Roberts et al. 2014: 1070).

Die Auswahl eines bestimmten K-Wertes ist dabei im Kern eine subjektive Abwägung zwischen mehreren konträren Faktoren je nach Untersuchungskontext, und keine definitive Aussage über die inhaltlichen Strukturen des jeweiligen Feldes (vgl. Munoz-Najar Galvez et al. Supplemental Material: 1) – Wie die Autoren des *Structural Topic Models* selbst sagen: „There is not a "right” answer to the number of topics that are appropriate for a given corpus“ (Roberts et al. 2019: 11). Aus diesem Grund beinhaltet das STM-R-Pakets eine Funktion, mit der sich mehrere unterschiedliche K-Werte für dasselbe Modell berechnen und in ihren Qualitäten vergleichen lassen. Für die hier vorgenommene Analyse wurden Modelle zwischen 20 und 120 Topics in Zehnerschritten miteinander verglichen, um ein robustes K zu finden. Die Veränderungen

[5] Es ist im normalen Sprachgebrauch auf sozialen Medien meist üblich, Emoji im Gegensatz zu Worten nicht durch Leerzeichen zu trennen, sondern direkt aneinander zu reihen. Dies kann bei der Sprachanalyse zu Problemen führen, da bspw. „😀“ und „😀😀“ aufgrund der fehlenden Leerzeichen als zwei grundverschiedene Worte aufgefasst werden können.

Zusätzlich liegen Emoji in codierter Form vor – das Emoji „😀“ ist zum Beispiel als „\u0001f600“ in den Daten angegeben, was aufgrund der Zahlen und Sonderzeichen ohne Recodierung nicht von der Sprachanalyse berücksichtigt wird.

bestimmter Kennzahlen für die Qualitäten eines STM über diese K-Werte finden sich in *Abbildung 8*.

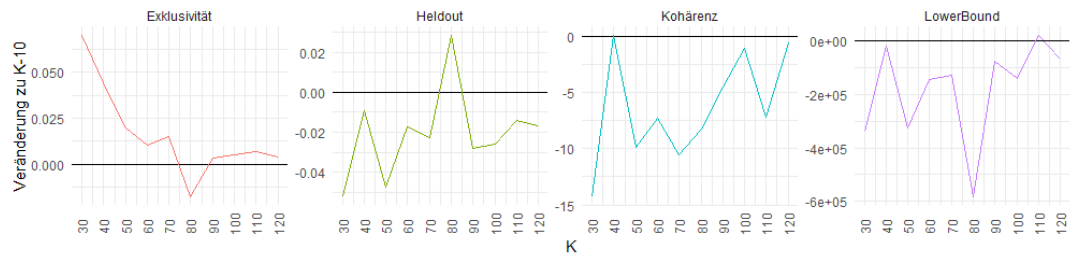


Abb. 8: Exklusivität, Heldout-Wahrscheinlichkeit, semantische Kohärenz und Lower-Bound-Wert für alle getesteten Modelle im Vergleich zum jw. Vormodell

Bei der Suche nach K und der Interpretation dieser Messzahlen ist insbesondere die Identifikation von Plateau-ähnlichen Strukturen ohne große Schwankungen von Bedeutung, um den Einfluss der jeweiligen Messqualitäten zu minimieren und ein robustes Modell zu berechnen. Aus den vier präsentierten Qualitäten lässt sich jeweils bei etwa 90 Topics der Anfang von Plateau-ähnlichen Strukturen erkennen, weshalb ein K von 90 für die folgenden Analysen gewählt wurde.

Es ist bei diesen Kennwerten anzumerken, dass sie im Gegensatz zu den Modellen von z.B. Munoz-Najar Galvez et al. eine deutliche Inter-Modell-Varianz in beide Richtungen anstatt eines klar erkennbaren, kohärenten Trends aufzeigen (vgl. Munoz-Najar Galvez et al. Supplemental Material: 2, 4f.). Ob dies an inhärenten Qualitäten der hier verwendeten Daten, oder der gewählten Laufzeit-Obergrenze in den Berechnungseinstellungen liegt, lässt sich dabei nicht feststellen, es sei jedoch aus Transparenzgründen hier angemerkt.

5. STM-Analysen und Erkenntnisse

Wie in *Abbildung 9* zu erkennen ist, zeigt sich große Varianz in der Menge an Tweets und deren zeitliche Unterteilung je nach Topic. Während zu einigen Topics nur in wenigen Wochen getweetet wurde (siehe Ende 2014), zeigen sich einige Topics über lange Zeiträume dominant.

Anhand der Menge an veröffentlichten Tweets lässt sich das Datenset in drei Zeiträume unterteilen. Diese drei Zeiträume decken sich dabei beinahe exakt mit Zeiträumen der drei dominanten Topics:

1. Eine Phase geringer Aktivität und kurzer Aktivitäts-Spitzen vom Beginn der Daten bis zum Jahresübergang 2014 zu 2015. In den hier erfassten 125 Wochen ist Topic 13 in 59 Wochen das Topic mit dem größten Anteil.

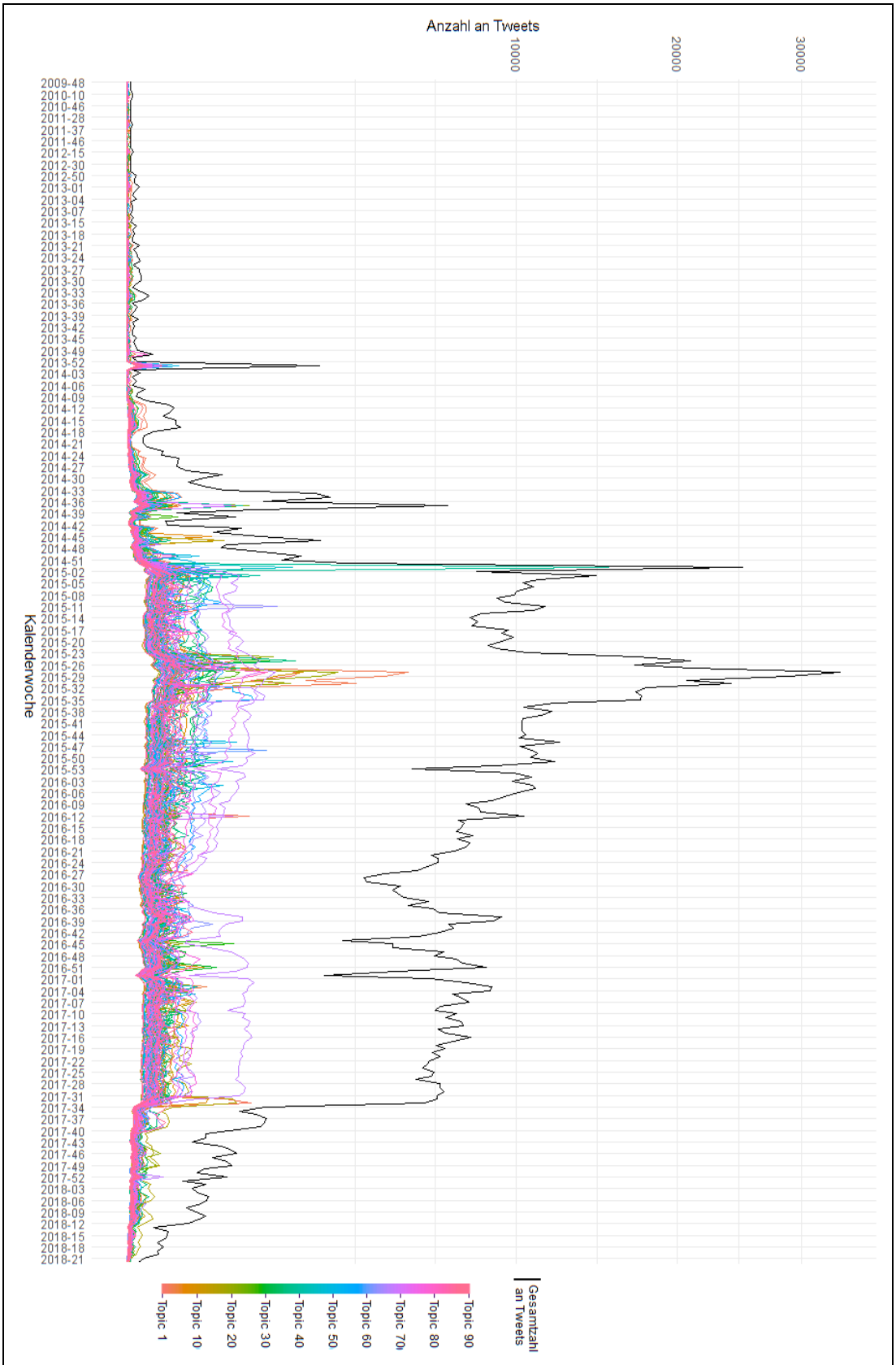


Abb. 9: Anzahl an veröffentlichten Tweets nach Topic (Farbe), gruppiert nach Kalenderwoche der Veröffentlichung. Schwarz: Gesamtzahl an Tweets in jw. Kalenderwoche – Wochen ohne Tweets nicht angezeigt

2. Eine Phase konstanter Aktivität mit mehrwöchigen Schwankungen von Beginn 2015 bis Ende Juli 2017. In den hier erfassten 137 Wochen ist Topic 68 in 111 Wochen das Topic mit dem größten Anteil.
3. Eine Phase nachlassender Aktivität ab August 2017. In den hier erfassten 43 Wochen ist Topic 17 in 33 Wochen das Topic mit dem größten Anteil.

5.1 Topic-Themengruppen und -Inhaltscodierung

Jedes der 90 Topics wurde manuell anhand der jeweils mit der höchsten Wahrscheinlichkeit vorkommenden Worte, der exklusivsten (also hauptsächlich in diesem Topic vorkommenden) Worte sowie der Tweets mit der größten Übereinstimmung mit dem jeweiligen Topic codiert und mit einer kurzen Beschreibung gekennzeichnet. Die Ergebnisse dieser Codierung finden sich in *Appendix A*. Zusätzlich wurden sie anhand der 20 Tweets mit der jeweils größten Übereinstimmung in drei Gruppen unterteilt:

- „News“: Für Tweets, die in sachlicher Formulierung aktuelle Geschehnisse präsentieren und in diesem Stil auch von dem Account einer seriösen Nachrichtenorganisation stammen könnten
- „Person“: Für Tweets ohne Nachrichtenwert, zu trendenden Themen oder mit eigener Wertung, wie sie ein normaler Twitter-Nutzer auf seinem eigenen Account posten könnte und
- „Spam“: Für Tweets, die selbst durch übermäßige Nutzung von Emoji, Hashtags oder das Taggen anderer Nutzer auffallen. Alternativ: Tweets, die zwar selbst „normal“ erscheinen, aber sich in den Top-Tweets mehrfach fast wortgleich wiederfinden, sodass eine zugrundeliegende Formulierung und deren strukturierte Verbreitung angenommen werden kann.

Um als eine dieser Kategorien codiert zu werden, mussten mindestens 6 der Top 20 Tweets dem jeweiligen Schema entsprechen. Diese Grenze wurde dabei gewählt, um die Möglichkeit zu erhalten, ein Topic als alle drei Themenblöcke gleichzeitig zu deklarieren, sollte dieser Fall eintreten.

Wie bereits von Kriel und Pavliuc festgestellt, sind viele der Tweets, und die meisten Topics, zu Nachrichtenthemen. Neben den von ihnen bemerkten und in mehreren Topics vertretenen Themenkomplexen Sport- (7 Topics, ~10% Anteil an allen Topics) bzw. Lokalnachrichten (6 Topics, 14% Anteil an allen Topics) finden sich jedoch auch internationale Nachrichten zu den Konflikten im Nahen Osten sowie zwischen den USA und China bzw. Russland, und nationale Nachrichten zu politischen Ereignissen in eigenen Topics wieder. Diese Struktur lässt darauf schließen, dass Nachrichtenaccounts nicht nur zur automatisierten Ansammlung von Followern (vgl. Kriel/Pavliuc

2019: 206), sondern auch zur Konfrontation der Follower mit bestimmten weltpolitischen Ereignissen genutzt wurden.

Der Anteil an Tweets, die aus Nachrichtentopics kommen, verhält sich dabei proportional zur Gesamtzahl an Tweets in den jeweiligen Wochen. Wie *Abbildung 10* zeigt, steigt zeitgleich mit der Gesamtzahl an Tweets pro Woche Anfang 2015 (vgl. *Abb. 9*) auch der Anteil an Tweets aus Nachrichtenthemen an, um im August 2017 gemeinsam mit der Gesamtzahl an Tweets zu fallen.

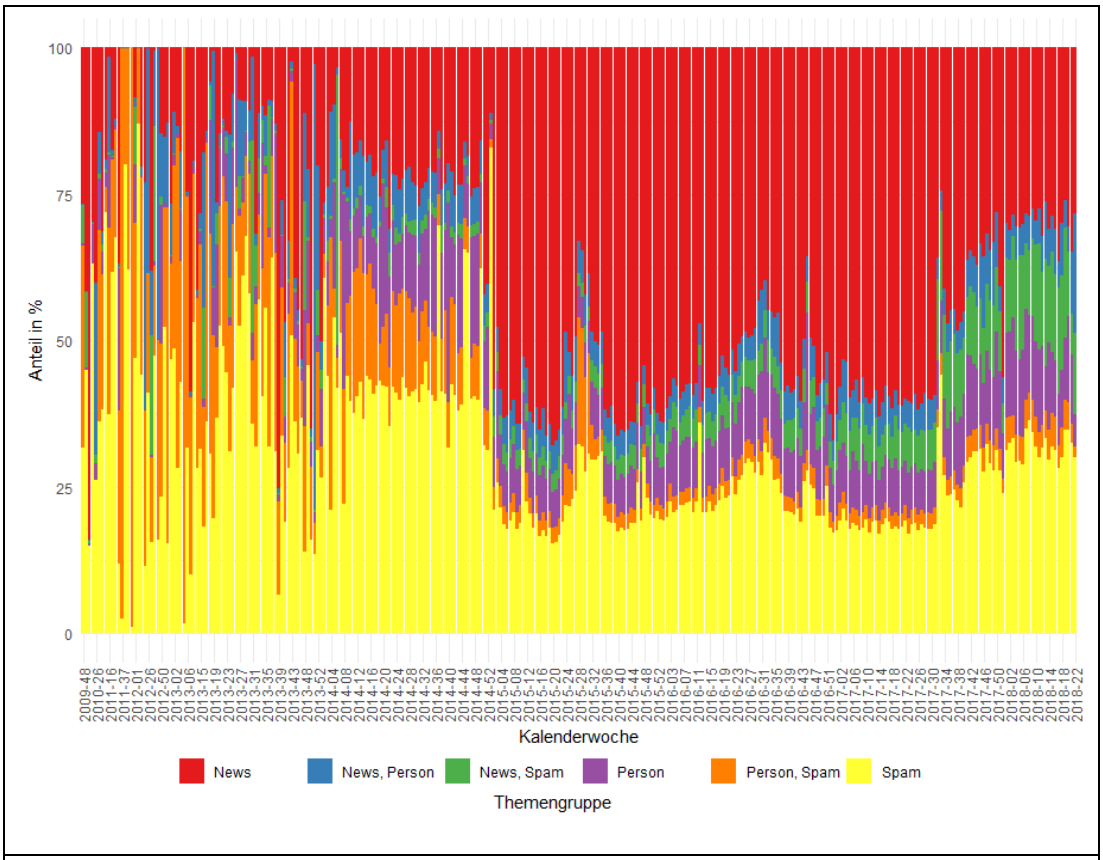


Abb. 10: Anteile der Topicgruppen an allen Tweets im jeweiligen Zeitraum

Abbildung 10 zeigt zusätzlich, dass sich insbesondere in den ersten Jahren der IRA-Aktivitäten starke und kurzfristige Unterschiede in den Inhaltsmengen der Topic-Gruppen zeigten. Erst ab etwa März 2014 lassen sich festere und länger beständige Strukturen in der inhaltlichen Mengenverteilung feststellen – selbst in Wochen mit ähnlich geringer Tweet-Anzahl wie vor März 2014.

Über das gesamte Jahr 2014 hinweg sind Topics, die teilweise oder ausschließlich Spam beinhalten, beinahe durchgängig über 50% Anteil an allen Tweets, während zwischen Januar 2015 und Juli 2017 News-Topics dominieren. Die letzte identifizierte Phase ab August 2017 weist hingegen keine dominante Topic-Gruppe auf, was einerseits an einem relativen Gleichgewicht zwischen Spam- und News-Topics und andererseits an einem Anstieg an Hybrid-Topics liegt.

Anhand der Ergebnisse des STM sowie der vorgenommenen Topic-Codierungen lassen sich neben der bereits beschriebenen Aufteilung in Tweet-Typen mehrere Inhaltsgruppen identifizieren, die häufig genug auftauchten, um ein oder mehrere Topics zu prägen.

5.2 Anteilsstarke News-Topics

Das Topic mit dem mit Abstand größten Anteil ist Topic 68. Mit knapp 8% Topic-Anteil in der STM-Auswertung liegt es dabei weit vor dem zweithäufigsten Topic 72 mit ~4,5% Anteil. Ordnet man jeden Tweet nach dem jeweils maximalen Theta-Wert einem einzigen Topic zu, so entfallen sogar 10% des Datensatzes auf Topic 68.
Wie in der Codierung in *Appendix A* zu sehen, handelt dieses Topic inhaltlich hauptsächlich von kriminellen Handlungen wie Diebstähle und Morde, sowie von Unfällen mit teilweise tödlichen Folgen und Vermisstenanzeigen.
Es zeigen sich dabei wenige Accounts für diese Tweets verantwortlich: Die 15 Accounts mit den meisten Tweets zu diesem Topic posteten gut 2/3 der Topic-Tweets. Eine genauere Betrachtung zeigt, dass die Nutzernamen von 14 der 15 Accounts nicht in gehashter Form vorliegen – und 13 der 14 namentlich vorhandenen Accounts sich als Nachrichtenorganisationen präsentieren, mit Namen wie „Baltimore Online", „Chicago Daily News" oder „New York City Today".
Topic 68 weist zwar über die komplette Laufzeit des Datensatzes zugeordnete Tweets auf, es lässt sich jedoch eine deutliche Hochphase an Aktivität ab Januar 2015 bis August 2017 erkennen. Bis auf einen mehrmonatigen Einbruch in der Tweetmenge in den Monaten um Juli 2016 verzeichnet dieses Topic im angegebenen Zeitraum eine deutliche und konstante Aktivität mit steigenden Interaktionszahlen (vgl. *Appendix B: 1*). Während sich inhaltlich über eine Stichprobenanalyse keine Veränderung in den Inhalten der Tweets vor bzw. nach diesem Aktivitätseinbruch erkennen lässt, findet zu dieser Zeit ein Wechsel in den postenden Accounts statt: Nur 8 der 20 aktivsten Accounts ab August 2016 waren bereits vor August 2016 unter den 20 aktivsten Accounts. Zusätzlich lässt sich in Abbildung 9 erkennen, dass ab dieser zweiten Welle an Aktivität mit Ausnahme einzelner Spitzen kein anderes Topic auch nur annähernd in die Nähe des Mengenanteils von Topic 68 kommt.

Topic 72, das zweitstärkste Topic, hält mit 4,5% Topic- bzw. 5,9% Tweetzuordnungs-Anteil bereits einen deutlich geringeren Anteil des Datensatzes als Topic 68. Wie aus der Codierung hervorgeht, ist Topic 72 ebenfalls ein „News"-Topic, nur dass es dieses Mal hauptsächlich um Sport-Themen geht.
Wie in dem bereits untersuchten Topic 68 zeigt sich auch hier eine deutliche inhaltliche Verschiebung im zeitlichen Verlauf und ein Zuwachs an Interaktionen (vgl.

Appendix B: 1). Während ab Beginn der Aktivitäten im Januar 2015 durchgängig mehrere tausend Tweets pro Woche diesem Topic zugeordnet wurden, bricht der Anteil im späteren Verlauf immer weiter ein. Ab etwa Juli 2016 bis zum Ende der Topic-Aktivitäten im August 2017 fällt die Anzahl wöchentlicher Tweets deutlich unter die bisherigen Werte und verbleibt im dreistelligen Bereich. Auch in diesem Topic finden sich beinahe ausschließlich Twitter-Accounts, die sich als Nachrichtenredaktionen präsentieren – 11 Nachrichten-Accounts finden sich sogar bei beiden untersuchten Topics in den 20 Accounts mit den jeweils meisten Tweets.

Mit dem drittgrößten Topic-Anteil (3,21%) und einem Tweetzuordnungs-Anteil von 4,29% findet sich mit Topic 66 das dritte News-Topic auf Platz 3 der Häufigkeitsverteilung. Neben dem in der Codierung erkennbaren Themenkomplex der gleichgeschlechtlichen Ehe finden sich in diesem Topic auch generelle Nachrichten zu nationalpolitischen und juristischen Entscheidungen sowie Ankündigungen zu Kandidaturen und Reden im Umfeld der Präsidentschaftswahl 2016.
Auch dieses Topic zeigt wieder zwei Aktivitätsphasen, mit einem Einbruch in der Anzahl geposteter Tweets zwischen Juli und September 2016, und einem Anstieg der erhaltenen Interaktionen im Verlauf (vgl. *Appendix B: 1*). Im Gegensatz zu den beiden bisher betrachteten Topics enden die Tweets zu diesem Topic jedoch nicht mit einem klar erkennbaren Schnitt im August 2017, sondern bleiben in deutlich geringerer Form noch bis November 2017 präsent. Dieses Topic zeichnen ebenfalls die bereits identifizierten News-Accounts aus, mit Top-20-Überschneidungen von 12 (Topic 68) bzw. 15 (Topic 72).

Der Aufbau einer Identität als Nachrichten-Accounts scheinen demnach einen großen Teil der IRA-Aktivitäten auf Twitter ausgemacht zu haben. Die beschriebenen drei Topics allein zeichnen sich je nach Berechnungsart für 15-20% der Tweets verantwortlich. Interessant ist dabei, dass alle drei Topics trotz ihrer großen inhaltlichen Unterschiede dieselbe grobe Struktur aufweisen: Eine Phase starker Aktivität zwischen Januar 2015 und Juli 2016, ein ein- bis zweimonatiger Aktivitätseinbruch, dann eine Wiederaufnahme der Aktivitäten mit veränderten Accounts und geringerer Tweetzahl.

5.3 Gezielte Verbreitung von Falschnachrichten

9 der 90 Topics behandeln angeblich reelle Inhalte, die nicht bzw. nicht annähernd so wie von der IRA beschrieben stattfanden. Anhand der in den Topics vertretenen Tweets und Worte lassen sich die dabei benutzten Hashtags ableiten. Die Gesamtmenge an Tweets zu diesen Hashtags in den unbereinigten Daten wurde im Folgenden untersucht.

Neben der von Kriel und Pavliuc als „first solid ‘burst of activity’“ bezeichneten Tweets am 11. September 2014 zu der angeblichen Explosion eines Chemiewerks in Louisiana (Kriel/Pavliuc 2019: 208) finden sich ab diesem Zeitpunkt mehrere Versuche, diese und ähnliche ausgedachte Katastrophen auf Twitter zu verbreiten:

- Die angebliche Explosion des Chemiewerks in Louisiana am 11. und 12. September 2014. Während diese Meldung laut Berichten über mehrere Medien (Textnachrichten, gefälschte Wikipedia-Einträge, …) verbreitet wurde (vgl. Borthwick 2015: o.S.), findet sie sich unter Hashtags wie „#ColumbianChemicals“ oder „#LouisianaExplosion“ auch in diesem Datensatz, gemeinsam mit einer zweiten, deutlich kleineren, Tweet-Gruppe am 16. September mit demselben Thema. Zeitlich passt diese Aktivität beinahe perfekt auf den zu Beginn identifizierten sprunghaften Anstieg an Aktivität von Accounts mit hauptsächlich nicht-englischsprachigen Tweets (vgl. *Abb. 2*). Tatsächlich sind 41,3% der zu diesem Thema tweetenden Accounts in der dort identifizierten Gruppe von 417 Accounts vertreten.
- Ein angeblicher Ausbruch von Ebola in Atlanta am 13. und 14. Dezember 2014. Aus Tweet-Texten lässt sich rekonstruieren, dass es sich hierbei neben angeblichen 911-Anrufen zu Ebola um eine vermutete Ebola-Patientin handelte, die nach ihrer Ankunft aus Libyen frei in Atlanta herumlief. Mehrere der Tweets bezeichneten sie als „Ebola zombie“ und verwenden ihren Namen und die Stadt als Hashtags: „#EbolaInAtlanta“ und „YattaQuirre“. In dieser Accountgruppe sowie in allen folgenden findet sich im Gegensatz zu der ersten untersuchten Falschmeldung jedoch kein einziger der zu Beginn identifizierten Accounts mit geringem Anteil englischsprachiger Tweets.
- Die angebliche Gefährdung eines Kernkraftwerks in der Ukraine am 3. und 4. Januar 2015 unter Hashtags wie „#Fukushima2015“, „#Chernobyl2015“ und ab dem 4. Januar „#Nukraine“. Was von offizieller Seite als ungefährlicher Kurzschluss präsentiert wurde (vgl. Huffington Post 2014: o.S.), wurde in IRA-Tweets als „radiation leak“ und „possible nuclear disaster“ beschrieben, das entweder als Konsequenz eines Konfliktes zwischen Obama und Putin, oder aufgrund des Versagens ukrainischer Politik geschah. Dies ist die umfangreichste Falschnachrichten-Kampagne, mit über 20.000 Tweets der IRA-Accounts und gleichzeitig der geringsten Reichweite.
- Angebliche Phosphorverunreinigungen im US-Bundesstaat Idaho am 10. März 2015. Unter dem Hashtag „#PhosphorousDisaster“ sammeln sich Tweets zu angeblich verunreinigtem Trinkwasser in mehreren Städten Idahos. Es lässt sich keine Verbindung zu tatsächlichen Ereignissen in dieser Gegend zu dieser Zeit finden.
- Angeblich vergiftete Thanksgiving-Truthähne. Am 26. und 27. November, sowie vereinzelt am 3. Dezember 2015 finden sich unter Hashtags wie „#KochFarms“,

„#Walmart“ oder „#Foodpoisoning“ Gerüchte über angeblich vergiftete Truthäne, die in Walmart-Läden in New York verkauft wurden. Neben bloßer Entrüstung finden sich auch mehrere Tweets zu vergifteten Müttern, sowie zu angeblichem Versagen der verantwortlichen staatlichen Stellen.

Alle diese Falschnachrichten richten sich dabei mehr oder weniger direkt gegen öffentliche Strukturen, die auf bestimmte Art und Weise in ihrer eigentlichen Rolle versagt haben. Es kann demnach vermutet werden, dass das Erschüttern von Vertrauen in Staaten und öffentliche Institutionen ein Kernziel der IRA war.
Zusätzlich ist zu erwähnen, dass sich bei den jeweiligen Tweets in allen Falschnachrichten-Kampagnen kaum exakte Duplikate, aber beinahe jeder Tweet in mehreren ähnlichen Formulierungen findet, weshalb alle Topics als „Spam“ deklariert wurden. Es ist aufgrund dieser Duplikate und Ähnlichkeiten zu vermuten, dass die strukturelle Verbreitung der Falschnachrichten einem normalen Twitternutzer – trotz des betriebenen Aufwandes bezüglich einzigartiger Formulierungen – bei genauerer Betrachtung der jeweiligen Hashtags aufgefallen wäre – insbesondere aufgrund der Tatsache, dass sich die große Menge an Tweets in jeder Kampagne auf eine bedeutend kleinere Zahl einzigartiger Accounts verteilte. Dies geschah in den meisten Fällen jedoch vermutlich nicht, da die Tweets kaum Interaktionen anderer Nutzer sahen (vgl. *Tabelle 3*).

Meldung	Chemie	Ebola	Ukraine	Phosphor	Truthahn
Gesamt-Tweets	6.107	636	23.819	2.661	3.031
Einzigartige Tweets	4.610	636	22.583	2.427	1.447
Einzigartige Nutzer	349	18	327	63	91
Ø Interaktionen	0,1873	0,1824	0.0134	0.0248	0.2408
Max. Interaktionen	379	85	9	15	21

Tab. 3: Gesamtzahl an IRA-Tweets, Anzahl einzigartiger Tweets in bereinigten Daten und Anzahl postender Accounts unter den analysierten Hashtags sowie deren durchschnittliche und maximale Interaktionszahl für alle 5 identifizierten Falschmeldungen

5.4 Black Lives Matter, Black Representation und Polizeigewalt

Aus den Studien anderer Autoren geht hervor, dass die IRA auch die Lebenswelt schwarzer Amerikaner als Ziel hatte und Teile ihrer Aktivitäten beispielsweise unter dem Banner der *Black Lives Matter*-Bewegung verbreitete. Diese „Zielgruppe“ findet sich innerhalb der vorliegenden Daten in drei thematischen Gruppen angesprochen: Direkt über Tweets, die die *Black Lives Matter*-Bewegung ansprechen, über Nachrichtentweets, die die Beteiligung schwarzer Amerikaner an politischen,

gesellschaftlichen und wissenschaftlichen Entwicklungen hervorheben und schwarze Rollenmodelle präsentieren, und indirekt über Tweets zu übermäßiger Polizeigewalt, insbesondere gegen schwarze Mitbürger.

Die Tweets, die sich mit dem Themenkomplex der Polizeigewalt befassen, zeigen sich strukturell als Mischung aus den beiden in vorherigen Schritten identifizierten Ansätzen. Zu Beginn der Aktivitäten zu diesem Themenbereich sind beinahe die kompletten Aktivitäten gebündelt an jeweils einem Tag in März und April des Jahres 2015, während sich ab Ende Juli 2015 eine konstantere Aktivität, mit Tweets zu diesen Hashtags in beinahe jeder Woche, einstellt.

Die beiden Spitzen am 14.03.2015 unter den Hashtags #CopsWillBeCops und #FergusonShooting und 28.04.2015 unter den Hashtags #BaltimoreVsRacism und #BaltimoreRiots stehen dabei vermutlich im Zusammenhang mit damals aktuellen Geschehnissen:
Während Michael Brown Jr. bereits im August 2014 von einem weißen Polizisten in Ferguson erschossen wurde, wurde am 12. März 2015 auf zwei Polizisten in Ferguson geschossen, mit der Bekanntgebung der Festnahme eines schwarzen Verdächtigen zwei Tage später (vgl. Reuters 2015: o.S.). Die IRA-Tweets zu diesem Thema verfolgen dabei kein einheitliches Motiv, sondern scheinen Dissens säen zu wollen. So finden sich Tweets, die die Seite der Polizei ergreifen („Violence against police is unacceptable!!! They do everything to protect us! #CopsWillBeCops“), aber auch Tweets, die das Geschehen als Verschwörung sehen, um Protestierende zu verunglimpfen („It's an attempt to make #Ferguson citizens look like criminas [sic] and broaden police power #CopsWillBeCops“).
Die Tweet-Aktivitäten Ende April 2015 sind vermutlich auf die Umstände des Todes des schwarzen Freddie Gray, der während der Festnahme schwer verletzt wurde und im Krankenhaus starb, bezogen. Nach seiner Beerdigung am 27. April 2015 kam es zu Ausschreitungen in Baltimore, bei denen Polizisten verletzt wurden und Sachschäden entstanden. Interessanterweise findet sich in diesem Fall trotz der gesteigerten Aggressivität der durchschnittlichen Protestierenden eindeutige Unterstützung der Proteste in den IRA-Tweets. So werden Protestierende weiterhin als Opfer dargestellt („Can you see the irony? Protesters are against #policebrutality and officials are getting more brutal to stop them #BaltimoreVsRacism“), und offizielle Reaktionen der Stadt sowie von Seiten Obamas kritisiert („#Obama wins the #Noble [sic] #peaceprize but does nothing to stop police violence#BaltimoreRiots #BaltimoreVsRacism“).
Während an beiden Tagen jedoch mehrere hundert Tweets der IRA gepostet wurden, konnte mit maximal 3 Interaktionen mit einem Tweet keine bedeutende Reichweite

erzielt werden. Erst im weiteren Verlauf der Kampagne und während der Zeit konstanter Postings erreichte dieses Thema Twitter-Nutzer – insbesondere ab Mitte 2016 finden sich dabei Tweets, die in größeren Mengen die 500-Interaktionen-Marke knackten. Ab dem Jahreswechsel 2016/2017 – wenige Wochen nach dem reichweitestärksten Tweet – finden sich kaum noch Tweets zu diesen Hashtags (vgl. *Appendix B: 2*).

Tweets, die die Worte „black lives" bzw. die Hashtags „BlackLivesMatter" oder „BlackTwitter" beinhalten, weisen ein ähnliches Muster auf. Zwar fallen hier die eben beschriebenen Aktivitätsspitzen zu den Ereignissen in März und April 2015 deutlich kleiner aus, und die erste Aktivität findet sich bereits im Dezember 2014, die generelle Struktur ist jedoch sehr ähnlich: Nach diesen ersten vereinzelten Spitzen stellt sich eine konstante Aktivität im Verlauf von 2015/2016 zu den untersuchten Phrasen ein, oft mit mehreren hundert Tweets pro Woche. Ab Ende 2016 fällt die Aktivitätszahl dann jedoch deutlich unter die bisherigen Maße zurück, und Ende 2017 fällt die Aktivität noch einmal deutlich ab, auch wenn sich bis zum Ende des Datensatzes weitere Tweets finden (vgl. *Appendix B: 2*).
Die Zahl erreichter Interaktionen ist dabei dem Themenkomplex „Polizeigewalt" sehr ähnlich: Während die ersten Tweets kaum Nutzer erreichten, erzielen spätere Tweets immer höhere Interaktionswerte. So wird Ende 2015 erstmals die 1.000-Interaktions-Marke geknackt und im weiteren Verlauf des Jahres 2016 erhalten mehrere Tweets über 10.000 Interaktionen. Auch nach den Reduktionen der Tweet-Anzahl erzielen die geposteten Tweets weiter in großen Mengen mehr als dreistellige Interaktionen.

Wenig überraschend, finden sich Tweets zu schwarzer Repräsentation und den Errungenschaften schwarzer Amerikaner insbesondere im „Black History Month" Februar sowie in der Zeit um den Geburtstag Martin Luther King Jr.s, dem „MLK Day" Mitte/Ende Januar. Während sich zum MLK Day bereits 2015 Tweets finden, taucht der Hashtag „#BlackHistoryMonth" erst im Februar 2016 auf (vgl. *Appendix B: 2*). Wie bereits bei anderen Topics festgestellt, finden sich auch hier die meisten Tweets im Jahr 2016, mit deutlichen Nachlässen in den folgenden zwei Jahren. Gleichzeitig steigt auch hier die Anzahl erhaltener Interaktionen beinahe exponentiell an: Während Tweets ab 2016 vierstellige Interaktionen verzeichnen und im Februar vereinzelt über die 10.000er Marke kommen, springen die Interaktionszahlen Ende 2016 jenseits der sechsstelligen Werte – ein Erfolg, der im Verlauf 2017 mehrmals wiederholt wird, insbesondere mit dem Tweet „Daily reminder that the most educated First Lady in American history is a black woman with two Ivy League degrees from Harvard and Princeton", der am 26.06.2017 gepostet wurde und 325.826 Likes und 123.617 Retweets

erhielt.
Eine weitere erwähnenswerte Tatsache ist die semantische Komplexität dieser Themengruppe. Während sich für alle bisher untersuchten Gruppen ein klar identifiziertes Topic bzw. klar identifizierte Hashtags finden, ist das Thema schwarzer Repräsentation abseits der Hashtags zu Black History Month oder MLK Day inhaltlich divers gefüllt mit Geschichten „persönlicher Erfolge“ angeblich schwarzer Amerikaner.

5.5 Hillary Clinton: Emails und Benghazi

Insgesamt finden sich 1623 Tweets in den Daten, die die Worte „Clinton“ und „Mail“, die Worte „Podesta“ und „Mail“ oder „DNCLeaks“ beinhalten und auf den E-Mail-Skandal Hillary Clintons abzielen. Trotz dieser vergleichsweise geringen Tweet-Zahl findet sich in den STM-Auswertungen ein eigenes Topic für dieses Thema (Topic 55). Gemeinsam mit 275 Tweets, die die Worte „Clinton“ und „Benghazi“ beinhalten, sollte sich aus der Strukturanalyse dieser Gruppe ein guter Überblick über Umfang und Erfolg der IRA-Angriffe auf Hillary Clinton in zwei Kern-Themengruppen untersuchen lassen.

Es finden sich insgesamt 1.898 Tweets mit den entsprechenden Worten. Während erwartbarerweise ein großer Teil dieser Tweets im November 2016, also dem Monat der Präsidentschaftswahl, gepostet wurden, finden sich bereits seit Januar 2015 Clinton-kritische Tweets. Diese sind dabei eine Mischung aus Nachrichten-Tweets zu Geschehnissen um Benghazi-Ermittlungen und Hillary Clintons E-Mails einerseits, und Kampagnen mit festen Hashtags im Stil bereits identifizierter Kampagnen andererseits. Die erste Spitze in der Anzahl an Posts im März 2015 ist beispielsweise auf eine Tweetkampagne am 8.3.2015 unter dem Hashtag #WhatClintonWrites zurückzuführen, die jedoch beinahe keine Interaktionen erlangte.
Die nächsten Aktivitätsspitzen im September und Oktober 2015 stützen sich im Gegensatz dazu nicht auf Hashtag-Kampagnen, sondern auf die strukturelle Verbreitung von Nachrichten zu Clintons Aussagen und den angeblichen Inhalten der E-Mails in ähnlichen Formulierungen über unterschiedliche Accounts. Auch hier werden höchstens zweistellige Interaktionszahlen erreicht (vgl. *Appendix B: 3*).
Die Tweets, die im Oktober und November 2016 eine deutliche Menge an Interaktionen erhalten, sind jedoch weder Nachrichten-Tweets noch ausgenutzte Hashtags. Stattdessen finden sich hier Tweets mit persönlichen Wertungen, die Clinton zu ihren Handlungen im Zusammenhang mit Benghazi, ihren E-Mails sowie Donald Trump scharf kritisieren. Die Tweets mit der höchsten Interaktionszahl lauten etwa: „"Hillary Clinton asks judge to toss defamation case filed by parents of #Benghazi" Hell no! She needs to pay for these lives.“ (14.717 Interaktionen) und „This is sickening. Hillary

using the "Mentally Ill" to incite violence at Trump rallies. #FreeJulian #BirdDogging #PodestaEmails10“ (11892 Interaktionen). Viele der weiteren Tweets mit hohen Interaktionszahlen bedienen sich einem ähnlichen Sprachmuster, mit vereinzelten Tweets im Nachrichten-Stil dazwischen. Alle der einflussreichsten Tweets kommen dabei von wenigen Accounts: Die 50 Tweets mit der größten Reichweite stammen alle von den Accounts „SouthLoneStar“, „TEN_GOP“, „Jenn_Abrams“, „Pamela_Moore13“ und „TheFoundingSon“, die alle jeweils über 45.000 Follower aufweisen. Ein großer Teil der reichweitestärksten Tweets wurde dabei interessanterweise erst nach der Präsidentschaftswahl gepostet.

Analysen der IRA-Tweets zu Donald Trump wurden bewusst nicht durchgeführt, da sich aufgrund seiner fehlenden politischen Laufbahn vor Amtsantritt keine so konkreten und untersuchbaren Themenkomplexe wie Hillary Clintons Emails oder Benghazi finden ließen, und viele der Aussagen zu Donald Trump laut STM-Topics entweder deutlich als Spam erkennbar oder inhaltlich diverse persönliche Tweets sind.

6. Rückschlüsse und Einordnung

Die hier präsentierten Analysen und deren Ergebnisse werfen ein gemischtes Licht auf die Methoden und den Erfolg der IRA-Twitterkampagne. Während sich zweifellos bestätigen lässt, dass mithilfe zugrundeliegender Koordination über mehrere tausend Twitter-Accounts und mehrere Millionen Tweets ein Versuch der Manipulation öffentlicher Meinung stattgefunden hat, ist der tatsächliche Erfolg dieser Kampagne und der Einfluss auf einen normalen Twitter-Nutzer fraglich. Grundlegend lassen sich drei miteinander verbundene Vektoren der Entwicklung der IRA-Kampagne ableiten: **Struktur**, **Inhalte** und **Reichweite**.

Strukturell zeigt sich von Beginn an eine deutliche Komplexität in der Twitter-Kampagne der IRA. So wurden weit vor dem tatsächlichen Start der Posting-Aktivitäten Ende 2014/Anfang 2015 (vgl. *Abb.9*) bereits die in dieser Kampagne benutzten Accounts erstellt (siehe 3.3), vermutlich um nicht offensichtlich anhand des Erstelldatums als gefälschter Account zu erscheinen. Auch zeigt die starke Nutzung von freiwilligen Ortsangaben (siehe 3.2) und die weite Streuung der Tweetzeiten (siehe 3.6), dass ein gewisses Mindestmaß an Aufwand in die Erstellung und Betreuung mehrerer tausend Fake-Accounts floss. Auch die in der STM identifizierte thematische Vielfalt und die Koexistenz vieler Themen im zeitlichen Verlauf (vgl. *Abb.9*) zeigt, dass komplexe Strukturen hinter der Betreuung dieser Accounts stecken. Während in den ersten Kampagnen jedoch noch teilweise vermutlich umfunktionierte russische Bots verwendet wurden (vgl. Falschmeldung „Columbian Chemicals“), verschwinden diese

schnell aus den identifizierten Kampagnen zugunsten rein englischsprachiger Accounts.

Während die Strukturen also von Anfang an hohe Komplexität aufweisen, lässt sich bei den **Inhalten** und der erzielten **Reichweite** eine starke Entwicklung nachvollziehen. Lässt man die ersten vermutlichen Testphasen mit wenigen Tweets und stark schwankenden Themenverteilungen (vgl. *Abb.10*) außen vor, lassen sich vier Kernphasen der IRA-Aktivitäten identifizieren:

- Eine erste Phase im Jahr 2014, die deutlich durch die Verbreitung von Spam und wild fluktuierenden Tweetmengen gezeichnet ist.
- Eine Stabilisierung der Tweetmengen auf hohem Niveau Ende 2014, begonnen durch die erste verbreitete Falschmeldung. Ab diesem Zeitpunkt werden die meisten der 2013/2014 erstellten Accounts aktiviert (vgl. *Abb. 4*) und es dominieren Newstopics in der Menge veröffentlichter Tweets (vgl. *Abb. 10*).
- Ein Umbruch in den veröffentlichten Inhalten um August 2016. Neben einem Einbruch an Aktivität in den drei größten Topics sowie der Veränderung an tweetenden Accounts zu diesen Topics (vgl. 5.2) findet sich ab diesem Zeitpunkt in mehreren Topics, insbesondere in Themenkomplexen zu schwarzen Amerikanern, ein deutlicher Anstieg in den erhaltenen Interaktionen. Es lässt sich vermuten, dass an diesem Zeitpunkt wenige Monate vor der Präsidentschaftswahl im November eine Art Umstrukturierung der IRA-Taktiken stattfand, um Einfluss und Reichweite zu maximieren. Insbesondere aufgrund der Tatsache, dass bis zu diesem Zeitpunkt die meisten untersuchten IRA-Accounts mit eigenen Tweets maximal mehrere hundert Interaktionen zu ihren Tweets erhielten, während Accounts, die Inhalte anderer Nutzer retweeteten auf etwa tausend Interaktionen kamen (vgl. *Abb.7*), lässt sich vermuten, dass die IRA von den erfolgreichen Inhalten anderer Nutzer lernte, um ihre eigenen Inhalte zu verbessern und mehr Reichweite zu erhalten
- Ein Einbruch an Aktivitäten ab etwa August 2017. An diesem Punkt verschwinden insbesondere Nachrichten-Accounts beinahe von einer Woche auf die nächste, aber auch in anderen Topicgruppen findet sich ein deutlicher Einschnitt in der Anzahl veröffentlichter Tweets (vgl. 5.2, 5.4). Während die Anzahl der Tweets sich über die IRA-Kampagne als Ganzes deutlich verringerte, erzielten viele der Tweets weiterhin große Mengen an Reichweite und Interaktionen, sodass davon ausgegangen werden kann, dass dieser Verringerung keine Reduktion des von Seiten der IRA betriebenen Aufwands für die einzelnen Tweets zugrunde lag.

Die Twitter-Kampagne der IRA zeigt sich somit in den untersuchten Daten als ein komplexes System, das sich aus simplen Anfängen weiterentwickelte. Dabei stützte

sie sich hauptsächlich auf zwei Pfeiler: Generelle Nachrichtenaccounts und Accounts mit Fokus auf bestimmte Themen.

Während sich bei den „Fokus-Accounts“ zu Beginn insbesondere der Versuch finden lässt, über Nutzung von Hashtag-Kampagnen bestimmte Themen mithilfe Spam-artiger Tweets trenden zu lassen, lernte die IRA aus den Fehlern dieses Ansatzes. Ab 2016 findet sich keine einzige dieser Hashtag-Kampagnen mehr in den hier angestellten Untersuchungen. Stattdessen finden sich länger angelegte Kampagnen zu bestimmten Themen, die diese konstant mit Inhalten versorgen und vermutlich dem Ziel dienten, bestehende Gruppen und Gemeinschaften zu unterwandern. Dieser Wechsel und der scheinbare Umbau der IRA-Strukturen im Juli und August 2016 führte zu einem deutlichen Anstieg in den erhaltenen Interaktionen in beispielsweise auf schwarze Amerikaner fokussierten Themenkomplexen. Den größten Erfolg erzielte die IRA jedoch mit Accounts, die sich als Privatperson präsentieren und ihre „Privatmeinung“ zu aktuellen politischen und sozialen Geschehnissen in den USA und der Welt äußern, wie die Analysen der Clinton-Tweets zeigen. Diese Accounts erhielten dabei nicht nur in bestimmten Topics mit Abstand die meisten Interaktionen, sondern sammelten auch deutlich mehr Follower an als alle anderen IRA-Accounts.

Aus der Untersuchung der Nachrichtentopics und deren Inhalte lässt sich die Vermutung ableiten, dass zeitgleich zu den ersten Trend-Spams versucht wurde, mithilfe von gefälschten News-Accounts ein komplexes Ökosystem an Nachrichtenaccounts über ganz Amerika hinweg zu erschaffen, bei denen die erreichten Nutzer unter dem Anschein „echter“ Nachrichten eine bestimmte Wirklichkeit, hauptsächlich bestehend aus Sportnachrichten, lokalen Gewalttaten und Unsicherheiten sowie internationalen Konflikten, zu sehen bekamen. Insbesondere im Hinblick auf die Tatsache, dass ein bedeutend großer Teil des Datensatzes sich mit lokaler Kriminalität und Gewalttaten befasst, lässt sich die Vermutung aufstellen, dass ein unterschwelliges Framing von Amerika und insbesondere der lokalen Umgebung als unsicher und gefährlich erreicht werden sollte.

In all diesen Analysen zeigt sich jedoch, dass ein großer Teil der reichweitenstärksten Tweets erst kurz vor oder weit nach der US-Präsidentschaftswahl 2016 gepostet wurde. So finden sich in den untersuchten Topics und Themenkomplexen kaum von der IRA verfasste Tweets, die bereits mehrere Monate vor der Präsidentschaftswahl 2016 US-Amerikaner in ihren jeweiligen Twitter-Umgebungen erreichten, und diese somit in ihrer politischen Meinungsbildung beeinflussen hätten können.
Der Einfluss der IRA und somit Russlands auf die öffentliche Meinung über Hillary Clinton und Donald Trump sowie Amerikas globalpolitische Position im Vorfeld der

US-Wahlen lässt sich somit anzweifeln, auch wenn die Reichweite der IRA-Tweets ab der Wahl und über das Jahr 2017 hinweg ein generelles Vorhandensein dieses Einflusses auf US-Bürger nahelegen.

Warum die Menge an Tweets der IRA-Accounts sich ab August 2017 drastisch reduzierte, bleibt jedoch offen. Mögliche Erklärungen könnten eine erste Welle der Löschung durch Twitter sein, die viele der Accounts -insbesondere solche, die sich als Nachrichtenaccounts präsentieren – im Zuge der Fake-News-Debatte und Berichten zu russischer Beeinflussung auf Twitter entfernte und somit die potenziell mögliche Zahl der Posts reduzierte.
Andere Möglichkeiten könnten das veränderte politische Klima zu dieser Zeit sein: So wurde Anfang August 2017 bekannt, dass Robert Muellers Nachforschungen zu Russlands Beeinflussung der US-Wahlen eine *Grand Jury* nutzte und somit potenzielle Anklagen und Prozesse auf die IRA und russische Offizielle zukommen könnten (vgl. NPR 2017: o.S.). Eine weitere Möglichkeit stellen die ebenfalls Anfang August 2017 beschlossenen Sanktionen gegen Russland dar, die von russischer Seite als „Handelskrieg“ bezeichnet wurden (vgl. BBC 2017: o.S.). In diesem Kontext könnte die Reduzierung der Posts eine Fokusverschiebung russischer Seite von gesellschaftlicher Beeinflussung hin zu wirtschaftlichen Konsequenzen für die USA bedeuten. All diese Möglichkeiten sind jedoch reine Spekulation und lassen sich anhand der vorliegenden Informationen weder bestätigen noch ablehnen.

7. Limitationen

Die hier präsentierten Analysen stoßen an bestimmten Punkten an Grenzen. Einige dieser Grenzen ergeben sich aus den vorhandenen Daten. So lässt sich keine abschließende Aussage über die Qualität der erhaltenen Interaktionen der Tweets treffen, da nicht nachvollzogen werden kann, ob und wie viele der Likes und Retweets von IRA-Accounts, und wie viele von „echten“ Accounts kamen. Auch lassen sich die Followerzahlen nicht zu ihrem jeweiligen Stand nachverfolgen, um beispielsweise die Followerzahl eines Accounts im Jahr 2016 mit ihrem Stand bei Erstellung des Datensets 2018 zu vergleichen, und nachzuverfolgen, ob diese Zahlen langsam im Lauf der Zeit, oder innerhalb kurzer Zeit aufgrund viraler Tweets oder Manipulationen anstiegen.

Die größte Limitation liegt jedoch in der genutzten STM-Analyse. Die als Ausdruck der Zuordnung zu den jeweiligen Topics angegebene Kennzahl Theta zeigt, dass das verwendete Modell nur bedingt seine volle Erklärungskraft entfaltet. Während mit knapp 5,6% aller Tweets ein kleiner Teil beinahe eindeutig einem einzigen Topic

zugeordnet wird (max. Theta-Wert >0,9), ist beinahe exakt ein Viertel der untersuchten Tweets bei einem maximalen Theta-Wert von unter 0,3.
Da Theta in der Summe über alle möglichen Topics für jeden Tweet 1 ist, bedeutet dies für das hier verwendete Modell, dass die erstellten Topics einen kleinen Teil der Tweets beinahe perfekt wiederspiegeln, während ein großer Teil der Tweets kaum bzw. nur schlecht von den jeweiligen Topic-Definitionen abgebildet wird. Während dieser Umstand in den hier durchgeführten Analysen zwar dafür sorgte, dass deutlich semantisch und inhaltlich ähnliche Tweetgruppen wie beispielsweise Spam-Kampagnen mit mehreren Tausend oder sogar nur mehreren hundert Tweets erkannt wurden, bleibt dabei ein großer Teil des Datensets und die dort vertretenen Inhalte von den Analysen unberührt.

Dieses Problem ließe sich theoretisch dadurch lösen, die Anzahl der untersuchten Topics K deutlich auf mehrere hundert oder vielleicht sogar tausend Topics anzuheben, bis eine im Voraus zu definierende Untergrenze für maximale Theta-Werte nicht mehr von einem so großen Anteil des Datensets unterschritten würde. In der Praxis stößt jedoch bereits das hier verwendete Modell an die Grenzen des aktuell technisch durchführbaren. Allein das Testen der unterschiedlichen K-Werte bis zu K=120 führt zu mehreren Tagen ununterbrochener Rechenzeit und benötigt neben längerer Laufzeit bei höherer Topiczahl aufgrund der Komplexität der Modelldimensionen Mindestvoraussetzungen an vorhandenem Arbeitsspeicher, die vermutlich ohne dediziert für diese Modelle gebaute Hardware nicht zu erreichen sind. Somit erweist sich der Ansatz des *Structural Topic Modellings* für ein Datenset dieser Größe zwar als gute Methode, um einen generellen Überblick über die vorhandenen Strukturen und dominanten Themen zu gewinnen. Für einen vollständigen Überblick über die hier vorhandene Menge an Dokumenten müssen jedoch andere Analysemethoden herangezogen, oder das Datenset zuvor in mehrere zuvor identifizierte und inhaltlich verschiedene Gruppen aufgebrochen werden.

8. Fazit

Obwohl sich das verwendete *Structural Topic Model* nicht als ideal für die Untersuchung der hier vorliegenden Daten erweist, lassen sich einige neue Erkenntnisse über die Twitter-Aktivitäten der IRA ab 2014 gewinnen:

So finden sich vier Arten der Veröffentlichung von Twitter-Inhalten:
1. Der strukturelle Aufbau von Nachrichtenaccounts, die durchgängig im Zeitraum Januar 2015 bis mindestens August 2017 aktiv waren, Nachrichten zu unterschiedlichsten Themen verbreiteten, und womöglich die Nachrichtenwahrnehmung der Follower

in bestimmte Richtungen beeinflussen sollten.
2. Spam-Tweets mit hohen inhaltlichen und semantischen Überschneidungen, die bestimmte Themen auf Twitter trenden lassen sollten, um echte Nutzer zu erreichen und zu beeinflussen. Zu Beginn für unterschiedlichste Themen eingesetzt (Falschmeldungen, Polizeigewalt, Clintons Emails), erzielten sie keine große Reichweite und wurden ab 2016 nicht mehr genutzt
3. Die Unterwanderung bestimmter Themenkomplexe und Subgruppen. Insbesondere die Lebenswelt schwarzer Amerikaner zeigt sich hierbei in den Diskussionen um Polizeigewalt, Black Lives Matter und schwarze Repräsentation als Ziel der IRA. Es steht zu vermuten, dass hier wie bei den Nachrichtenaccounts unter dem Deckmantel einer aufgebauten Identität im weiteren Verlauf bestimmte Ideen und Themen in die Diskussionen eingebracht wurden, um die öffentliche Meinung zu verändern.
4. Der Aufbau „persönlicher“ Accounts bestimmter „Privatpersonen“, die politische und gesellschaftliche Ereignisse in den USA und der Welt dokumentieren und kommentieren. Dieser Ansatz erwies sich dabei als deutlich effektiver als die restlichen – „Privat“-Accounts erreichten sowohl die höchsten Follower- als auch die höchsten Interaktionszahlen.

Während die ersten Jahre der IRA-Aktivität in allen untersuchten Themenkomplexen kaum „echte“ Twitter-Nutzer erreichte, fand im Zeitraum Juli/August 2016 scheinbar ein struktureller Umbau in Accounts und Inhalten statt. Ab diesem Zeitpunkt finden sich über alle Themen hinweg deutlich höhere Interaktionszahlen. Während jedoch der potenzielle Einfluss der IRA auf die Twitter-Erfahrung US-amerikanischer Nutzer ab dieser Veränderung nicht bestritten werden kann, bleibt die Frage offen, inwiefern die Wahlentscheidungen amerikanischer Twitter-Nutzer für die nur wenige Wochen später stattfindende Präsidentschaftswahl 2016 von IRA-Twitteraktivitäten beeinflusst wurden.

Literaturverzeichnis

AP News 2020: Disputing Trump, Barr says no widespread election fraud. In: https://apnews.com/article/barr-no-widespread-election-fraud-b1f1488796c9a98c4b1a9061a6c7f49d Zugegriffen am 13.01.2021

Atkinson, Ryan 2018: Content Analysis of Russian Trolls' Tweets Circa the 2016 United States Presidential Election. Undergraduate Research Paper, in: University of Minnesota Digital Conservancy: http://hdl.handle.net/11299/199858

Bail, Christopher A./ Guay, Brian/ Maloney, Emily/ Combs, Aidan/ Hillygus, D. Sunshine/ Merhout, Friedolin/ Freelon, Deen/ Volfovsky Alexander 2020: Assessing the Russian Internet Research Agency's impact on the political attitudes and behaviors of American Twitter users in late 2017. In: Proceedings of the National Academy of Sciences Jan 2020, 117 (1). S. 243-250

BBC 2017: US sanctions are 'trade war' on Russia, says PM Medvedev. In: https://www.bbc.com/news/world-europe-40809715 Zugegriffen am 02.03.2021

Borthwick, John 2015: Media hacking. In: https://render.betaworks.com/media-hacking-3b1e350d619c Zugegriffen am 04.03.2021

Boyd, Ryan/ Spangher, Alexander/ Fourney, Adam/ Nushi, Besmira/ Ranade, Gireeja/ Pennebaker, James W./ Horvitz, Eric 2018. Characterizing the Internet Research Agency's Social Media Operations During the 2016 U.S. Presidential Election using Linguistic Analyses. In: PsyArXiv: https://psyarxiv.com/ajh2q/

Daily Beast 2020: Fox News Editor Thought Their Disastrous Seth Rich Story Would Be 'Vindicated'. In: https://www.thedailybeast.com/fox-news-editor-thought-their-disastrous-seth-rich-story-would-be-vindicated Zugegriffen am 22.01.2021

Fishman, Brian 2019: Crossroads: Counter-terrorism and the Internet. In: Texas National Security Review Vol. 2, Iss. 2. S. 83-100.

Freedom Outpost 2018: Blatant Voter Fraud In Plain Sight In 2018 Mid-Term Elections. In: https://freedomoutpost.com/blatant-voter-fraud-in-plain-sight-in-2018-mid-term-elections/ Zugegriffen am 13.01.2021

Huffington Post 2014: Accident Occurred At Ukraine Nuclear Power Plant, But Poses No Danger: Govt. In: https://www.huffpost.com/entry/ukraine-nuclear-plant-accident_n_6260390 Zugegriffen am 04.03.2021

Kriel, Charles/ Pavliuc, Alexa 2019: Reverse Engineering Russian Internet Research Agency Tactics Through Network Analysis. In: Defence Strategic Communications Spring 2019, 6. S. 199-227

Munoz-Najar Galvez, Sebastian/ Heiberger, Raphael/ McFarland, Daniel 2020: Paradigm Wars Revisited: A Cartography of Graduate Research in the Field of Education (1980–2010). In: American Educational Research Journal Vol. 57, Iss. 2. S. 612-652.

Neil Hoch, Indira 2020: Russian Internet Research Agency Disinformation Activities on Tumblr: Identity, Privacy, and Ambivalence. In: Social Media + Society Vol. 6, Iss. 4. S. 1-12

Newsweek 2020: Sidney Powell's Ties to QAnon Movement Explained. In: https://www.newsweek.com/sidney-powell-qanon-voter-fraud-lawsuits-georgia-1552050 Zugegriffen am 13.01.2021

NPR 2017: Source: Mueller Using D.C. Grand Jury In Russia Probe. In: https://www.npr.org/2017/08/03/541432868/source-mueller-using-d-c-grand-jury-in-russia-probe?t=1614693560401 Zugegriffen am 02.03.2021

Pew Research Center 2020: Amid Campaign Turmoil, Biden Holds Wide Leads on Coronavirus, Unifying the Country. In: https://www.pewresearch.org/politics/2020/10/09/amid-campaign-turmoil-biden-holds-wide-leads-on-coronavirus-unifying-the-country/ Zugegriffen am 13.01.2021

Politico 2016: 2016 Presidential Election Results. In: https://www.politico.com/2016-election/results/map/president/ Zugegriffen am 27.01.2021

Politico 2021: Presidential Election Results – Joe Biden has been declared the winner, toppling Donald Trump after four years of upheaval in the White House. In: https://www.politico.com/2020-election/results/president/ Zugegriffen am 27.01.2021

Postfity 2020: Twitter Follow Bots – Good or Evil? The Pros and Cons of Using Twitter Bots in Marketing. In: https://postfity.com/blog/twitter-follow-bots/ Zugegriffen am 29.01.2021

Reuters 2015: Suspect charged in shooting of police officers in Ferguson, Missouri. Archiviert in: https://archive.is/3CbFQ Zugegriffen am 15.03.2021

Roberts, Margaret E./ Stewart, Brandon M./ Tingley, Dustin/ Lucas, Christopher/ Leder-Luis, Jetson/ Kushner Gadarian, Shana/ Albertson, Bethany/ Rand, David G. 2014: Structural Topic Models for Open-Ended Survey Responses. In: American Journal of Political Science Vol. 58, No. 4. S. 1064-1082

Roberts, Margaret E./ Stewart, Brandon M./ Tingley, Dustin 2019: stm. An R Package for Structural Topic Models. In: Journal of Statistical Software Vol. 91, Iss. 2. S. 1-40.

Select Commitee on Intelligence o.J.: Report of the Select Committee on Intelligence, United States Senate, on Russian Active Measures, Campaigns and Interference in the 2016 U.S. Election – Volume 2: Russia's use of Social Media with additional Views. In: https://www.intelligence.senate.gov/sites/default/files/documents/Report_Volume2.pdf Zugegriffen am 13.01.2021

Time 2019: Here Are All of the Indictments, Guilty Pleas and Convictions From Robert Mueller's Investigation. In: https://time.com/5556331/mueller-investigation-indictments-guilty-pleas/ Zugegriffen am 23.03.2021

Twitter ohne Jahr: Information Operations. In: https://transparency.twitter.com/en/reports/information-operations.html Zugegriffen am 25.01.2021

Twitter 2018: Enabling further research of information operations on Twitter. In: https://blog.twitter.com/en_us/topics/company/2018/enabling-further-research-of-information-operations-on-twitter.html Zugegriffen am 25.01.2021

Twitter Help ohne Jahr: How to change your language settings. In: https://help.twitter.com/en/managing-your-account/how-to-change-language-settings Zugegriffen am 13.01.2021

World Population Review 2021: Us States - Ranked by Population 2021. In: https://worldpopulationreview.com/states Zugegriffen am 27.01.2021

Yonder 2018: SSCI Research Summary – 2016 Disinformation Report. In: https://go.yonder-ai.com/2016-disinformation-report-pdf Zugegriffen am 22.01.2021

Appendix A – Topics und Topic-Inhalte

Topic	Themen-Label	Top 5 wahrscheinlichste Worte Top 5 exklusivste Worte	Anteil (ø, gerundet)
News			
4	Märsche und Demonstrationen	million march thousand news protest march parad thousand hundr renamemillionwomenmarch	0.0079
5	Celebrities, Seth Rich Verschwörungstheorien	william philli news da rich rich seth william sacramento clayton	0.0041
8	Kunst	new orlean news citi street art street orlean york jersey	0.0178
18	Transit und Technologie	busi news project road develop project listen podcast rail construct	0.0086
30	Urlaub, Feste	famili celebr christma light holiday christma holiday gift tree eve	0.0081
34	Celebrities, Musik	entertain showbiz top news star entertain jenner album movi the-top-ten	0.0220
35	Lokalnachrichten Chicago	news chicago area local park emanuel rauner levin chicago illinoi	0.0189
37	US-Politik, Obama	obama presid polit visit hous michell barack obama polici administr	0.0165
38	Sport, Super Bowl 49/50	local bowl super ad 50 bowl super 50 falcon pro	0.0082
39	Sport, Cleveland	sport news coach nfl say fifa suspens bree soccer payton	0.0179
44	Syrien, Russlands Wahleinmischung 2016	immigr russian russia hack putin putin ambassador unfair gruber vladimir	0.0046
48	US-Politik, Biden, Bush	polit washington bush campaign run georg bush jeb w carter	0.0101
49	Hunde, Tierheime	dog owner news hot rescu dog pet cat puppi anim	0.0083
52	Celebrities, Foke-Medien (Name)	foke reveal uk british london foke reveal bbc labour britain	0.0094

Topic	Themen-Label	Top 5 wahrscheinlichste Worte Top 5 exklusivste Worte	Anteil (ø, gerundet)
55	Clinton, E-Mails	clinton hillari polit email privat clinton email hillari iloveobama benghazi	0.0085
56	Syrien, Iran, Sanktionen	deal talk news world iran iran deal kerri middl talk	0.0110
57	Gesundheit, FDA, Zika-Virus	health tech drug news busi health zika fda outbreak studi	0.0290
60	Probleme mit Technologien	place caus gas unit leak place gas taken incid leak	0.0068
62	Sport, Fischen	open lake river spring local fish lake outdoor trout columbia	0.0082
63	Konflikte: Südchinesisches Meer, ISIS	news world u. environ china earthquak quak environ hurrican tropic	0.0215
66	US Supreme Court, gleichgeschl. Ehe	polit court news bill senat court suprem rule lawmak marriag	0.0321
68	Lokalnachrichten, Verbrechen	man polic news shoot kill fatal robberi identifi shoot stab	0.0794
72	Sport	sport win beat lead basebal titl 2-1 djokov yanke ot	0.0449
73	Gerichte und Prozesse	offic case polic news trial slain sheriff box attorney funer	0.0161
74	Sport	video game play watch fan video play hashtag maketvsexi game	0.0120
76	Sport, Golf	year break tiger sinc second tiger wood golf master sexysport	0.0092
77	Lokalnachrichten, Feuer, Verkehr, Unfälle	fire news home local san fire firefight crew burn valley	0.0269
79	Lokalnachrichten, Unfälle	found 1 bodi dead leav wash found pretend bodi 1	0.0093
80	Technologie, Tech-Unternehmen	secret agent news trust servic secret agent trust booz serviceofnosecret	0.0044

Topic	Themen-Label	Top 5 wahrscheinlichste Worte Top 5 exklusivste Worte	Anteil (ø, gerundet)
81	Lokalnachrichten, Verbrechen, Polizeigewalt	stop baltimor polic search traffic baltimor riot baltimorevsrac stop cop swillbecop	0.0079
85	Sport	news sign bear worth fort bear lion cutler polar matt	0.0076
86	Gerichtsprozesse, Boston Marathon Bomber	topnew bomb threat news terror boston plot marathon jewish suicid	0.0074
87	Lokalnachrichten, Verbrechen, Russland-Nachforschungen	investig bust caught protest call edit stole bust thief investig	0.0065
88	Sexuelle Verbrechen von Sportlern	former sport news player smith smith vega raider las 49er	0.0099
89	Naher Osten, Syrien, ISIS	forc isi syria kill news saa aleppo raqqa deirezzor sdf	0.0144
Personen			
25	Pro-Amerika, Pro-Trump, Reagan	goe emojflagunitedst viral news captain viral emojflagunitedst superoldhero ronald emojhighvoltag	0.0035
28	Pro-Amerika, Tea Party	miss usa welcom born trend usa materialevid emot trend songshannibalwoulds	0.0055
29	Wahlen 2016, pro-Trump	make money vote america elect money trumpforpresid vote makeamericagreatagain hillaryforprison2016	0.0114
36	Pro-Black Lives Matter	live black matter activist voic live matter voic size activist	0.0058
41	Zitate	iamonfir word book other read write iamonfir letter paper unknown	0.0069
42	Pro-Trump, generelle Tweets	comment news figur compet broken purpos emojfacescreaminginfear distanc heel ifihadabodydoubl	0.0031
43	Anti-Islam, Anti-Demokraten	media social news orlando radic radic emojbackhandindexpointingdown typic media hypocrit	0.0037

Topic	Themen-Label	Top 5 wahrscheinlichste Worte Top 5 exklusivste Worte	Anteil (ø, gerundet)
50	GOP-Präsident-schaftsdebatten	parti paul candid debat problem gopdeb carson vegasgopdeb stopt-hegop ben	0.0071
53	Generelle Tweets, F-Wort	king f stephen news k f k kennedi ck pave	0.0037
58	Pro-Black Lives Matter	never student high school colleg emojraisedfist graduat x malcolm hat	0.0075
59	Pro-Trump, gene-relle Tweets	love best ever friend seen seen rap emojredheart best love	0.0079
69	Immigration, Bor-der Wall	secur illeg border legal news illeg secur homeland alien immigrati-onact	0.0045
71	Generelle Tweets	see stori financ cri girlfriend emojloudlycryingfac istartcryingwhen doom stori nose	0.0056
83	Zitate, Black Em-powerment	strong news fight embrac chang emojflexedbicep embrac emojparty-popp 2017survivaltip signsyouarea-merican	0.0029
90	Antworten „white guilt“	news say local get call face guilt news hous get	0.0024
Spam			
1	Workout	good exercis much walk lot count daili twice 2014 fun	0.0176
2	Reaktionen auf Terrorangriffe	rt stand left agre pleas enlist agre prayforbrussel barbmuen-chen usfa	0.0108
6	Texit	like look thing question answer forward like answer look everybodi	0.0097
7	Anti-Obama	readi point interest complet wonder wow obamalameduck interest won-der fair	0.0056
9	Anti-Demokraten, Confederacy-Sym-bole	rt group flag remov call network confeder remov flag statu	0.0066
10	Followerstatistiken, generelle Tweets	peopl hate everyth stupid mani ihat ilov__butih__ tag lone ex	0.0085

Topic	Themen-Label	Top 5 wahrscheinlichste Worte Top 5 exklusivste Worte	Anteil (ø, gerundet)
12	Wetterberichte, Pro-Bannon/Anti-Kushner	weather forecast heat news rain cool emojgrinningfacewithbigey forecast emojpolicecarlight heat	0.0070
16	Veterans Day	day thank happi free honor veteran rememberyourhero serv remembhero	0.0114
20	Workout, Bernie Sanders	workout today better sander safe leg differ yesterday feelthebern preworkout	0.0118
22	Intellectual Dark Web-Mentalitäten	emojfacewithtearsofjoy guy liber sure idea emojfacewithtearsofjoy loser hilari doubt laugh	0.0107
24	Falschnachricht Columbian Cemicals	world attack louisiana fake terrorist louisiana recogn jarvi disturb opinion	0.0083
26	Falschnachricht Columbian Cemicals, News Brief, 9/11	9 11 news mark blast 9 11 a.m demand bankruptci	0.0053
31	Falschnachricht ukrainisches AKW, Workout	time may go everi long time may explod squar everi	0.0100
33	Falschnachrichten Columbian Chemicals, Louisiana Explosion, Ebola	realli well god wait oh wait yeah sorri well louisianaexplos	0.0132
40	Falschnachricht ukrainisches AKW	fukushima2015 ukrain nuclear power fukushimaagain ukrain disast danger chernobyl experi	0.0211
45	Pro-/Anti-Islam/Muslime	muslim islam human peac oscarformuhammad oscarformuhammad muhammad oscarssowhit whoismuhammad allah	0.0062
46	Falschnachricht ukrainisches AKW	offici head care govern t offici don anyth fact fool	0.0079

Topic	Themen-Label	Top 5 wahrscheinlichste Worte Top 5 exklusivste Worte	Anteil (ø, gerundet)
51	Falschnachricht Ebola in Atlanta, demok. Präsidents-debatten	anoth yet demndeb demdeb economi yet demdeb demndeb blah economi	0.0042
54	Genereller Spam	amp m r e news q steven jess nicol h	0.0080
61	Falschnachricht vergiftete Truthäne	newyork nyc turkey fail ny kochfarm ny turkey walmart	0.0072
64	Falschnachricht verunreinigtes Trinkwasser	american fall water phosphorusdisast michigan phosphorusdisast water flint phosphorus idaho	0.0095
65	NRA, 2nd Amendment	gun tcot pjnet conserv patriot tcot 2a pjnet ccot wakeupamerica	0.0061
67	Genereller Spam, Justin Bieber	justin mr andrew jack nick justin amber jennif emojheartsuit bieber	0.0075
70	Mobile-Game-Werbung, Bürgerkrieg	us right war refuge stock civil refuge war right us	0.0126
78	Pro-Black Lives Matter, Polizeigewalt	cop blacklivesmatt racist black racism policebrut acab blacktwitt fergusonrememb btp	0.0110
82	Workout	weight know someon someth job weight someon cloth ass buddi	0.0076
Kombination News/Person			
19	Celebrities, trendende Hashtags, Harry Potter	counti roll news harri met harri stone potter style reid	0.0065
21	Finanzkrise, persönliche Bilder	take photo via post stock selfi via stick instagram roy	0.0087
23	Right-Wing, Musiknachrichten	news husband emojclappinghand clark steve emojclappinghand folk shove finest giuliani	0.0030
27	Essen und Trinken	start exercis ice beer news cheat balanc coke cream soda	0.0078

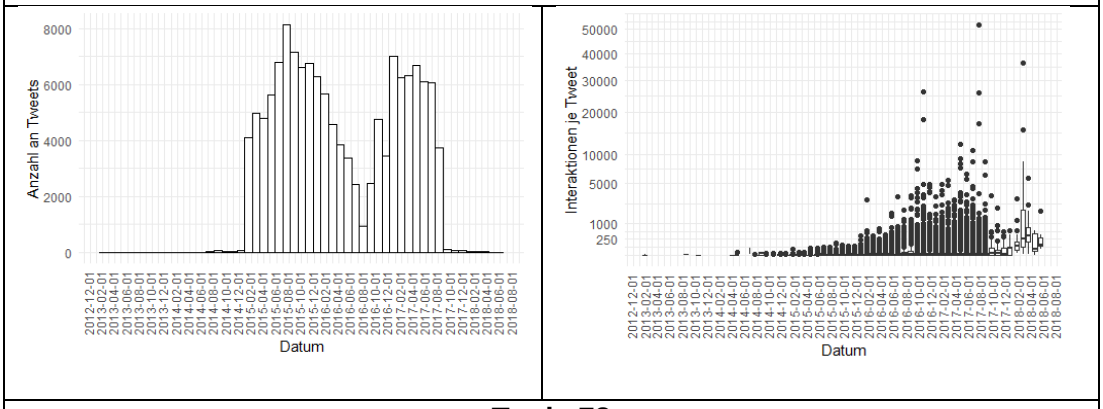
Topic	Themen-Label	Top 5 wahrscheinlichste Worte Top 5 exklusivste Worte	Anteil (Ø, gerundet)
32	Generelle Tweets und Nachrichten	news doubl bee ladi futur emojraisinghand courtroom buse 1990 spice	0.0027
47	Essen und Trinken, trendende Hash-tags	order shop appl news pizza pizza deliveri thingsinventedwhi-lehigh order taco	0.0083
75	Kinder, Schulen, Pro-Impfungen, Pro-Trans-Kinder	kid children school parent news parent children vaccinateus kid fos-ter	0.0065
84	Tech-Nachrichten, trendende Hash-tags	old phone catch internet drive phone ihatepokemongobecaus thingstodoinawaitingroom internet annoy	0.0058
Kombination News/Spam			
13	KSN Threat Tra-cker, Lokalnach-richten Texas	texa 2015 kansa number juli number april june juli 2015	0.0150
14	Vault7/Obamagate, Lokalpolitik, Town Halls	news emojfir doll cheerlead lit emojfir cheerlead doll concept fuckin	0.0024
15	Black History Month, Horoskope für Linda C. Black	women white black girl men 17 c histori crack women	0.0116
17	Trump-Kabinett, „Trump Supporters react to …“	trump donald polit ralli support donald boom penc inaugur trump	0.0139
Kombination Person/Spam			
3	Workout, generelle Tweets	workout just work great feel earli great later feel lol	0.0292
11	Diäten, generelle Tweets	want need help lose tri loss lose tip fast tri	0.0165

Ø Anteil = Durchschnitt der Anteile des jeweiligen Topics an allen Tweets

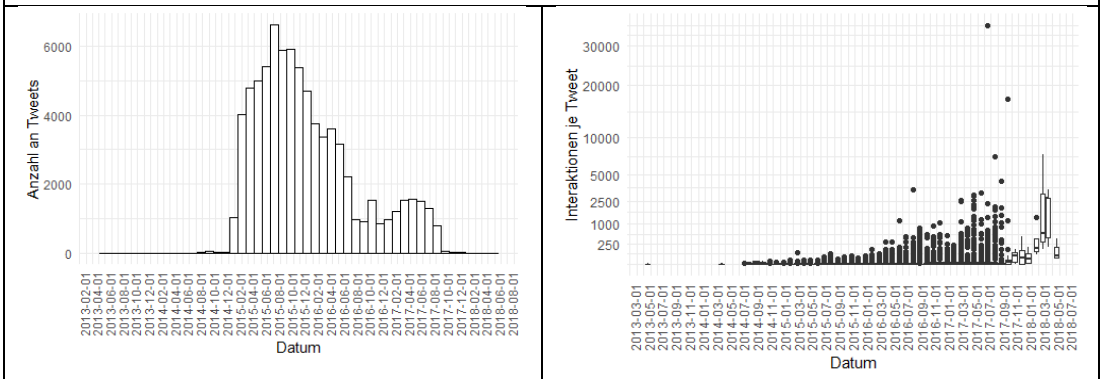
Appendix B – Ausgewählte Topics im zeitlichen Verlauf

1. Reichweitenstärkste Nachrichten-Accounts

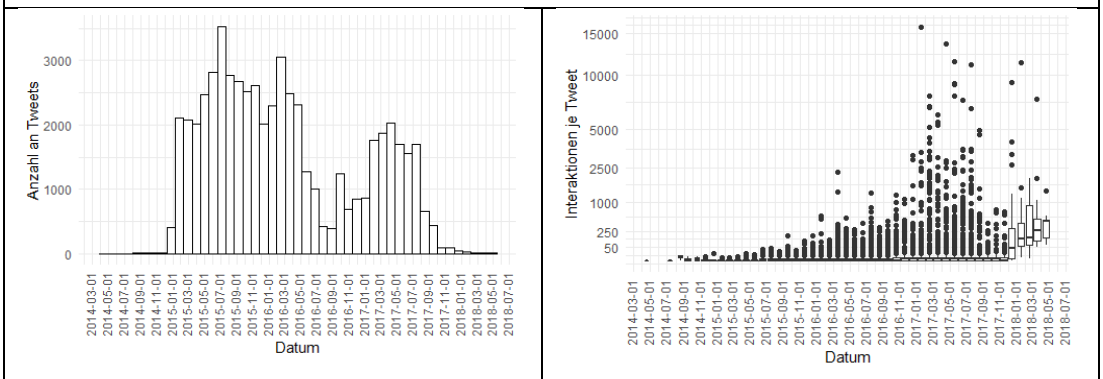
Topic 68



Topic 72



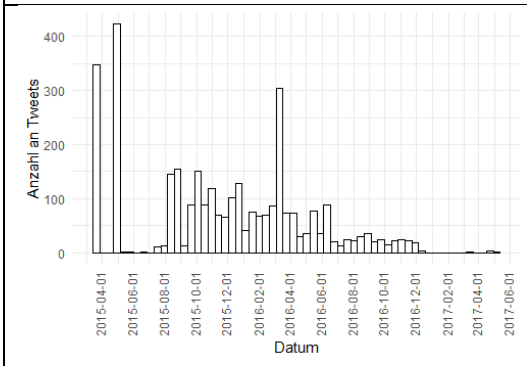
Topic 66

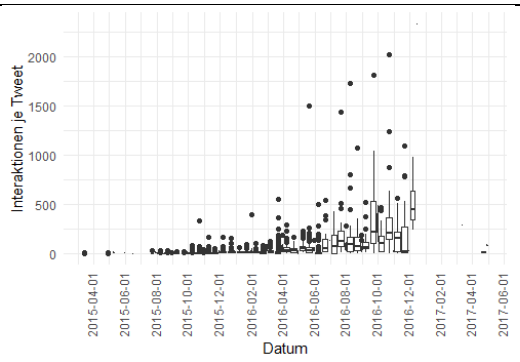


Tweetmengen und erhaltene Interaktionen (Quadratwurzel-transformiert) summiert nach Monatsintervallen für alle drei untersuchen News-Topics (Topics mit größter Reichweite)

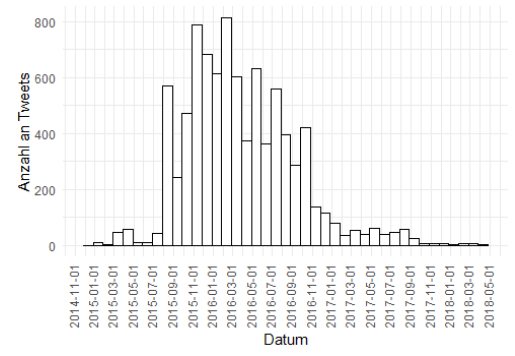
2. Gesellschaftliche Themen um schwarze Amerikaner

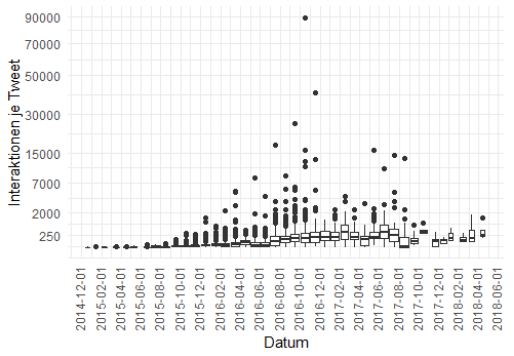
Polizeigewalt



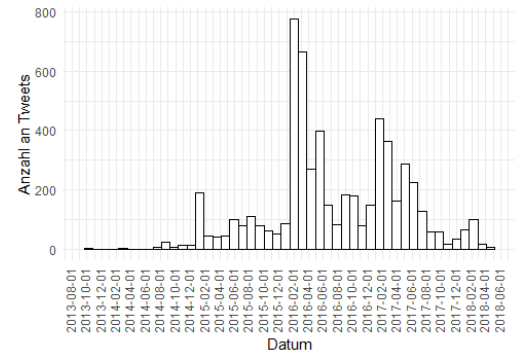


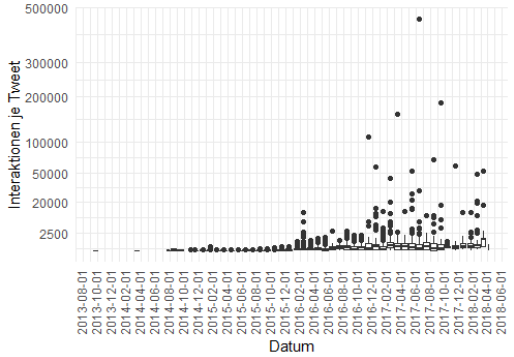
„Black Lives Matter“ / “BlackTwitter“





Black Representation / Black History

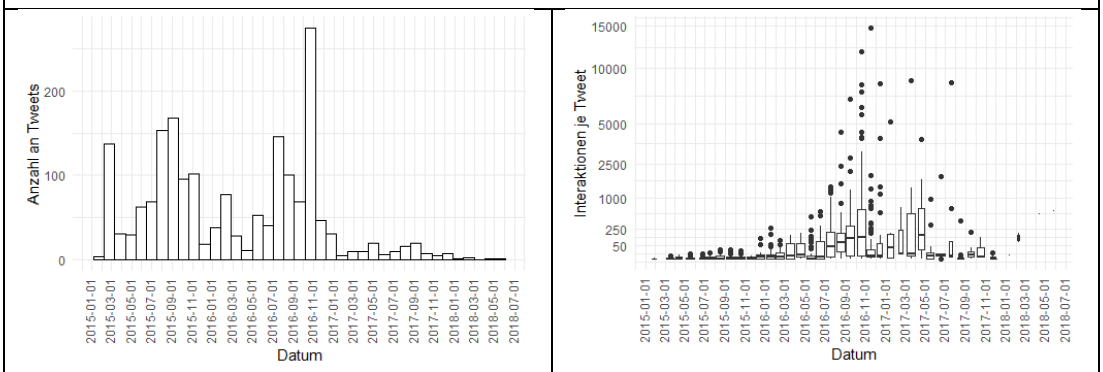




Tweetmengen und erhaltene Interaktionen (Quadratwurzel-transformiert) summiert nach zweiwöchigen (Polizeigewalt) bzw Monatsintervallen für alle drei untersuchen Themenkomplexe zur Lebenswelt schwarzer Amerikaner

2. Hillary Clinton

Clinton-E-Mails und Benghazi



Tweetmengen und erhaltene Interaktionen (Quadratwurzel-transformiert) summiert nach Monatsintervallen für Tweets zu Hillary Clintons E-Mail- und Benghazi-Skandalen.