

Socialbots in politischen Online-Konversationen. Eine (überschätzte) Gefahr für demokratische Willensbildung im Internet?

Florian Muhle, Robert Ackland und Timothy Graham

Wenige Tage vor der 58. US-Präsidentenwahl am 9. November 2016 tauchten in den internationalen Medien Schlagzeilen auf, die darauf hinwiesen, dass so genannte Socialbots den Ausgang der Wahl zugunsten des Kandidaten der Republikaner *Donald Trump* beeinflussen könnten. Zu diesem Zeitpunkt war zwar ein knapper Wahlausgang zu erwarten, dennoch waren sich die meisten Beobachter einig, dass *Hillary Clinton* am Ende die Wahl für sich entscheiden würde.¹ Vor diesem Hintergrund ist es nicht verwunderlich, dass nach der Wahl im Zuge der Ursachenforschung für *Trumps* überraschenden Sieg Socialbots wieder in den Blick gerieten und von Politik und Öffentlichkeit für den Wahlausgang (mit-)verantwortlich gemacht wurden. So wies Bundeskanzlerin *Angela Merkel* nur wenige Tage nach der US-Wahl in einer Rede vor dem Deutschen Bundestag auf Gefahren hin, die Bots für Meinungsbildungsprozesse bedeuten können², und die Grünen forderten bereits mit Blick auf die Bundestagswahl 2017 in einem Beschluss der Bundestagsfraktion die Kennzeichnungspflicht von Socialbots³, um Mediennutzer vor Täuschung und Manipulation zu schützen. Dazu passend stellte der Deutschlandfunk fest, dass „der Weg an die Macht [...] neuerdings über Socialbots [führt]⁴“, während auf Arte angesichts der kostengünstigen Herstellung von Socialbots die Frage gestellt wurde, ob „unsere Wahlen gekauft“⁵ seien.

Deutlich wird durch die Reaktionen auf den Einsatz von Socialbots, dass diese seit der Wahl *Donald Trumps* auch in Deutschland als Problem, wenn nicht gar Gefahr für die demokratische Willensbildung betrachtet werden, der begegnet werden muss, um Manipulation und Beeinflussung von Wahlentscheidungen vorzubeugen. Zugleich ist aber auch feststellbar, dass die Debatte und Aufregung um Socialbots und deren Einfluss auf demokratische Meinungsbildungsprozesse im Wesentlichen auf der (oberflächlichen) Rezeption einiger weniger Studien beruhen, die mit automatisierten Methoden der Big Data-Analyse arbeiten, um große Datenmengen auswerten zu können.

Wie im vorliegenden Beitrag anhand der Forschung zur Präsidentschaftswahl in den USA dargelegt werden soll, geht diese Orientierung an großen Daten jedoch (notwendig?) mit einer im Vergleich zu traditionellen Instrumenten der sozialwissenschaftlichen For-

1 Vgl. *Bernhard Kornelius*, Die US-Präsidentenwahl vom 8. November 2016. *Trumps* Triumph, in: ZParl, 48. Jg. (2017), H. 2, S. 287 – 310, S. 288 f.

2 Vgl. Rede von Bundeskanzlerin Dr. Angela Merkel zum Haushaltsgesetz 2017 vor dem Deutschen Bundestag am 23. November 2016 in Berlin, Bulletin 138-1, <https://www.bundesregierung.de/Content/DE/Bulletin/2016/11/138-1-bkin-bt.html> (Abruf am 10. September 2018).

3 Vgl. Bündnis 90/Die Grünen Bundestagsfraktion, „Verantwortung, Freiheit und Recht im Netz“, Fraktionsbeschluss vom 13. Januar 2017, https://www.gruene-bundestag.de/files/beschluesse/Fraktionsbeschluss_Verantwortung_im_Netz_Weimar17.pdf (Abruf am 10. September 2018).

4 Peter Welcherling, Social Bots. Wahlkampf der Algorithmen, in: Deutschlandfunk online vom 22. Januar 2017, http://www.deutschlandfunk.de/social-bots-wahlkampf-der-algorithmen.740.de.html?dram:article_id=376345 (Abruf am 8. August 2018).

5 Hanna Peters, Sind unsere Wahlen gekauft?, in: Arte online vom 6. Februar 2017, <https://info.arte.tv/de/social-bot-bundestagswahl-manipulation> (Abruf am 8. August 2018).

schung geringeren Sensibilität für die Besonderheiten der Online-Kommunikation sowie einer geringeren Validität der Messinstrumente einher. In der Konsequenz gelangen wir auf Grundlage eines eigenen netzwerkanalytischen „Small Data-Ansatzes“⁶ zu anderen Ergebnissen bezüglich der Bedeutung von Socialbots in der politischen Online-Kommunikation, die möglicherweise zu einer etwas nüchternen geführten Debatte beitragen können.

1. Socialbots in der politischen Online-Kommunikation – zum Stand der Forschung

Socialbots sind Computerprogramme, die in sozialen Netzwerken im Internet User-Accounts (Nutzerkonten) steuern. Dabei geben sie fälschlicherweise vor, „echte“ Nutzer zu sein, da sie (in der Regel) manipulativen Zwecken dienen, die unentdeckt bleiben sollen. Häufig werden Socialbots nicht einzeln, sondern im Verbund mit anderen Bots eingesetzt, um gezielt und massenhaft bestimmte Nachrichten oder Hashtags weiterzuverbreiten. Auf diese Weise soll die Aufmerksamkeit der Online-Community sowie interessierter Dritter (zum Beispiel Politik und Massenmedien) auf spezifische Themen und Inhalte gelenkt und deren Popularität beeinflusst werden.⁷ Bevorzugtes Einsatzgebiet von Socialbots ist der Kurznachrichtendienst Twitter, der einerseits ein zunehmend bedeutsames Medium für politische Kommunikation darstellt⁸ und es andererseits über seine offene API (Application Programming Interface, Programmschnittstelle) verhältnismäßig leicht macht, Accounts zu automatisieren.

Die Verbreitung von Socialbots in sozialen Netzwerken und deren gesellschaftlichen und politischen Implikationen sind bisher kaum (sozial-)wissenschaftlich erforscht. So hat das Büro für Technikfolgen-Abschätzung beim Deutschen Bundestag im Rahmen einer systematischen Literatursuche im Frühjahr 2017 weltweit nur 59 Publikationen zum Thema finden können und festgestellt, dass die wissenschaftliche Aufmerksamkeit erst seit 2015 langsam zunimmt.⁹ Viele der bisher existierenden Arbeiten befassen sich aus informationswissenschaftlicher Perspektive mit grundlegenden Fragen der Bot-Erkennung und verwenden automatische Verfahren zur Klassifizierung automatisierter Accounts in sozialen Netz-

6 Diesen entwickeln wir derzeit im Forschungsprojekt „Socialbots as Political Actors?“ unter der Leitung von *Robert Ackland* (Australian National University) und *Florian Muhle* (Universität Bielefeld). Das Projekt wird finanziell vom BMBF im Rahmen des DAAD-Programmes „Projektbezogener Personenaustausch Australien 2017“ gefördert.

7 Vgl. *Norah Abokhodair / Daisy Yoo / David W. McDonald*, Dissecting a Social Botnet, in: *Dan Cosley / Andrea Forte / Luigina Ciolfi / David McDonald* (Hrsg.), CSCW, 15. Proceedings of the 2015 ACM International Conference on Computer-Supported Cooperative Work and Social Computing: March 14-18, 2015, Vancouver, BC, Canada / New York 2015, S. 839 – 851; *Simon Hegelich / Dietmar Janetzko*, Are Social Bots on Twitter Political Actors? Empirical Evidence from a Ukrainian Social Botnet (2016), <https://www.aaai.org/ocs/index.php/ICWSM/ICWSM16/paper/view/13015> (Abruf am 8. August 2018); *Andree Thielges / Simon Hegelich*, Manipulation in sozialen Netzwerken, in: ZfP, 64. Jg. (2017), H. 4, S. 493 – 512, S. 496.

8 Vgl. *Marco Dohle / Uli Bernhard*, Mediennutzung und -wahrnehmung von Bundestagsabgeordneten. Ersetzen oder ergänzen Online-Medien klassische Informations- und Kommunikationskanäle?, in: ZParl, 45. Jg. (2014), H. 4, S. 763 – 774, S. 764.

9 Vgl. *Sonja Kind / Tobias Jetzke / Sebastian Weide / Simone Ehrenberg-Silkes / Marc Bovenschulte*, Social Bots. TA-Vorstudie 2017, TAB-Horizon-Scanning Nr. 3 (2017), S. 22, <https://www.tab-beim-bundestag.de/de/pdf/publikationen/berichte/TAB-Horizon-Scanning-hs003.pdf> (Abruf am 8. August 2018).

werken.¹⁰ Genuin sozialwissenschaftliche Arbeiten, die sich Socialbots widmen, existieren kaum und setzen sich auch nicht im engeren Sinne empirisch mit dem (politischen) Einsatz von Socialbots auseinander.¹¹ Dies bleibt bisher Forschenden vorbehalten, die in der Lage sind, große Datenmengen aus sozialen Netzwerken automatisiert auszuwerten. Indem diese sich der politischen Kommunikation zuwenden, befassen sie sich zwar mit einem traditionell sozialwissenschaftlichen Thema, methodisch arbeiten sie jedoch primär mit den Mitteln der Informatik und der „Big Data-Analyse“, das heißt der „informationstechnologischen Auswertung großer Datenmengen“¹², deren großes Versprechen darin liegt, allein aufgrund der Quantität der untersuchten Daten zu objektiven Aussagen über die Wirklichkeit zu gelangen: „With enough data, the numbers speak for themselves.“¹³

Entsprechende Analysen zur politischen Kommunikation liegen bisher vor allem mit Blick auf die Brexit-Kampagne¹⁴, den Ukraine-Konflikt¹⁵ und die auch in diesem Beitrag im Vordergrund stehende Präsidentschaftswahl 2016 in den USA¹⁶ vor, wobei erst die überraschende Wahl Trumps dafür gesorgt hat, dass auch die breite Öffentlichkeit auf das Thema aufmerksam wurde. Das Vorgehen bisheriger Arbeiten zum politischen Einsatz von Socialbots, obwohl im Detail mit unterschiedlichen Methoden arbeitend, ist im Prinzip stets sehr ähnlich. Im ersten Schritt wird mithilfe von Hashtags ein Datenkorpus erstellt,

- 10 Vgl. *Samuel C. Woolley / Douglas Guilbeault*, Computational Propaganda in the United States of America: Manufacturing Consensus Online. Working Paper 2017.5, Oxford 2017; *Onur Varol / Emilio Ferrara / Clayton A. Davis / Filippo Menczer / Alessandro Flammini*, Online Human-Bot Interactions. Detection, Estimation, and Characterization (2017/03/27), <https://arxiv.org/pdf/1703.03107.pdf> (Abruf am 8. August 2018); *Emilio Ferrara / Onur Varol / Clayton A. Davis / Filippo Menczer / Alessandro Flammini*, The Rise of Social Bots, in: Communications of the ACM, 59. Jg. (2016), H. 7, S. 96 – 104; *Bence Kollanyi / Philip N. Howard / Samuel C. Woolley*, Bots and Automation over Twitter during the U.S. Election, Oxford 2016; *V. S. Subrahmanian / Amos Azaria / Skylar Durst / Vadim Kagan / Aram Galstyan / Kristina Lerman / Linhong Zhu / Emilio Ferrara / Alessandro Flammini / Filippo Menczer*, The DARPA Twitter Bot Challenge, in: Computer, 49. Jg. (2016), H. 6, S. 38 – 46; *Simon Hegelich / Dietmar Janetzko*, a.a.O. (Fn. 7).
- 11 Vgl. *Robert W. Gehl / Maria Bakardjieva* (Hrsg.), Socialbots and their Friends. Digital Media and the Automation of Sociality, London / New York 2017.
- 12 *Michael Klemm / Sascha Michel*, Big Data-Big Problems? Zur Kombination qualitativer und quantitativer Methoden bei der Erforschung politischer Social-Media-Kommunikation, in: Heike Ortner / Daniel Pfurtscheller / Michaela Rizzolli / Andreas Wiesinger (Hrsg.), Datenflut und Informationskanäle, Innsbruck 2014, S. 83 – 98, S. 84.
- 13 *Chris Anderson*, The End of Theory. The Data Deluge Makes the Scientific Method Obsolete, in: Wired online vom 23. Juni 2008, <https://www.wired.com/2008/06/pb-theory/> (Abruf am 8. August 2018).
- 14 Vgl. *Philip N. Howard / Bence Kollanyi*, Bots, #StrongerIn, and #Brexit. Computational Propaganda during the UK-EU Referendum, Oxford 2016.
- 15 Vgl. *Simon Hegelich / Dietmar Janetzko*, a.a.O. (Fn. 7).
- 16 Vgl. *Bence Kollanyi / Philip N. Howard / Samuel C. Woolley*, a.a.O. (Fn. 10); *Alessandro Bessi / Emilio Ferrara*, Social Bots Distort the 2016 U.S. Presidential Election Online Discussion, in: First Monday, 21. Jg. (2016), H. 11, <http://firstmonday.org/ojs/index.php/fm/article/view/7090/5653> (Abruf am 14. September 2018); *Samuel C. Woolley / Douglas Guilbeault*, Computational Propaganda in the United States of America: Manufacturing Consensus Online. Working Paper 2017.5, Oxford 2017; *Philip N. Howard / Samuel Woolley / Ryan Calo*, Algorithms, Bots, and Political Communication in the US 2016 Election. The Challenge of Automated Political Communication for Election Law and Administration, in: Journal of Information Technology & Politics, 15. Jg. (2018), H. 2, S. 81 – 93; *Bence Kollanyi / Philip N. Howard / Samuel C. Woolley*, a.a.O. (Fn. 10).

der Accounts enthält, die in einem bestimmten Zeitraum Tweets zum interessierenden Thema versendet haben. Dabei wird davon ausgegangen, dass Hashtags bei Twitter die kommunikative Funktion erfüllen, Tweets als zu bestimmten Themen beitragend zu markieren und damit auch öffentlich sichtbar zu machen.¹⁷ Im zweiten Schritt wird darauf aufbauend versucht, im Datenkorpus solche Accounts zu identifizieren, die botgesteuert sind. Dies ermöglicht es im dritten Schritt, deren Tweets daraufhin zu untersuchen, wie sich diese zum interessierenden Thema politisch positionieren, um hieraus viertens Schlussfolgerungen über deren Wirkung in politischen Online-Diskussionen zu ziehen, die gegebenenfalls durch weitergehende Analysen gestützt werden.

Prinzipiell handelt es sich hierbei um ein plausibles und nachvollziehbares Vorgehen. Doch der Teufel steckt im Detail, und bei einer genaueren Betrachtung der so arbeitenden Studien fällt auf, dass die Möglichkeit, automatisiert große Datenmengen auszuwerten, zu Lasten der Validität der gewählten Forschungsinstrumente geht, wodurch auch die Aussagekraft der Ergebnisse eingeschränkt wird. Dies wird im Folgenden anhand der Forschungsarbeiten illustriert, auf die sich im November 2016 die weltweite Berichterstattung über den Einsatz von Socialbots im US-Wahlkampf bezog¹⁸ und die aus dem Kontext der beiden Forschungszusammenhänge stammen, die in der Untersuchung von Socialbots weltweit bisher federführend sind. Hierbei handelt es sich um das amerikanische „Observatory on Social Media“ an der Indiana University und das „Computational Propaganda Project“ an der University of Oxford, das in jüngster Zeit auch Fallstudien zu computergesteuerter Propaganda in verschiedenen Ländern¹⁹, darunter Deutschland veröffentlicht hat.²⁰

2. Konzeptionelle Probleme bisheriger Forschung

Mit Bezug zur U.S.-Wahl haben beide Forschungsgruppen automatisiert mehrere Millionen Accounts und deren Tweets ausgewertet²¹, die im Vorfeld der Wahl gesendet wurden und sich mittels bestimmter Hashtags wie etwa #elections2016, oder #trump beziehungsweise #hillary auf den Wahlkampf und die konkurrierenden Kandidaten bezogen. Ziel war es dabei, die „activity of Socialbots involved in the 2016 U.S. Presidential election conversation ongoing on social media“²² beziehungsweise „the character and process of

17 Vgl. hierzu auch Axel Bruns / Hallvard Moe, Structural Layers of Communication on Twitter, in: Jean Burgess / Merja Mahrt / Axel Bruns / Katrin Weller (Hrsg.), Twitter and Society, New York 2014, S. 15 – 28, S. 17.

18 Dies sind Bence Kollanyi / Philip N. Howard / Samuel C. Woolley, Bots and Automation over Twitter during the First U.S. Presidential Debate, Oxford 2016; dies., a.a.O. (Fn. 10); Alessandro Bessi / Emilio Ferrara, a.a.O. (Fn. 16).

19 Vgl. Samuel C. Woolley / Philip N. Howard, Computational Propaganda Worldwide: Executive Summary, Oxford 2017.

20 Vgl. Lisa-Maria N. Neudert, Computational Propaganda in Germany: A Cautionary Tale, Oxford 2017; dies. / Bence Kollanyi / Philip N. Howard, Junk News and Bots during the German Parliamentary Election: What are German Voters Sharing over Twitter?, Oxford 2017.

21 Vgl. Alessandro Bessi und Emilio Ferrara, a.a.O. (Fn. 10), haben rund 2,8 Millionen Accounts in ihrem Datenkorpus, von denen zwischen dem 16. September 2016 und dem 21. Oktober 2016 rund 20 Millionen Tweets versendet wurden. Bei Bence Kollanyi, Philip N. Howard und Samuel C. Woolley, a.a.O. (Fn. 10), sind es 3,7 Millionen Accounts mit 19,4 Millionen Tweets im Zeitraum vom 1. bis 9. November 2016.

22 Alessandro Bessi / Emilio Ferrara, a.a.O. (Fn. 16).

political conversation over Twitter”²³ zu untersuchen. Bei einem genauen Blick auf die Arbeiten zeigt sich allerdings, dass sie, zumindest im engeren Sinne, gar keine Konversationen untersuchen, sondern lediglich isolierte Aktivitäten und – im Falle von *Alessandro Bessi* und *Emilio Ferrara* – Prozesse der Informationsdiffusion. Damit ist bereits das erste Problem angesprochen.

2.1. Problematisches Verständnis der Kommunikationsstrukturen auf Twitter

Sowohl *Alessandro Bessi* und *Emilio Ferrara* als auch *Bence Kollanyi* geben an, Aussagen über Aktivitäten von Bots in den Online-Konversationen zur US-Wahl treffen zu wollen. Grundlage hierfür ist ein sehr weiter und analytisch unscharfer Begriff von Konversation. So fassen beide Autorengruppen diesen als Sammelbegriff für sämtliche Aktivitäten bei Twitter und zählen auch isolierte Aktivitäten, die keinerlei Reaktion erfahren oder Prozesse der Informationsdiffusion, wie sie eintreten, wenn Nachrichten per Retweet weiterverbreitet werden, als Bestandteil von Online-Konversationen. Dies widerspricht aber gängigen Assoziationen und Anforderungen, die typischerweise mit dem Begriff der Konversation verbunden werden. Denn sowohl im alltäglichen als auch im sozialwissenschaftlichen Verständnis werden Konversationen als ein sequenzielles und reziprokes Geschehen begriffen, in dem mehrere Beteiligte in einen Austausch miteinander treten und wechselseitig aufeinander Bezug nehmen.²⁴

Der Begriff der Online-Konversation ruft damit Assoziationen eines dialogischen Geschehens hervor, in dem die Beteiligen ihre (politischen) Standpunkte austauschen und abwechselnd die Rollen der Sprechenden beziehungswise Zuhörenden übernehmen. Ein solches sequenzielles und reziprokes kommunikatives Geschehen wird aber von den genannten Autoren gar nicht untersucht. *Kollanyi* u.a. betrachten lediglich die Summe isolierter Aktivitäten, die sich unter den sie interessierenden Hashtags versammeln, um auf diese Weise Häufigkeit und politische Ausrichtung von Äußerungen zu bestimmen, die durch automatisierte Accounts erzeugt werden. Auf diese Weise können sie zwar feststellen, wie oft bestimmte Hashtags verwendet werden, es gerät aber keineswegs in den Blick, wie verschiedene User aneinander anschließen. Kommunikationsprozesse werden von *Kollanyi* u.a. also nicht untersucht.

Bessi und *Ferrara* gehen hier zumindest einen Schritt weiter. Ergänzend zur Analyse von Hashtags widmen sie sich insoweit auch sequenziellen Kommunikationsprozessen, als sie Retweet-Netzwerke einbeziehen, um die Rolle von Bots darin zu eruieren.²⁵ Über Konversationen im eigentlichen Sinne, die nicht nur sequenziell, sondern eben auch reziprok ablaufen, sagt dies aber nichts aus, denn Retweet-Netzwerke etablieren keine bidirektionale

23 *Bence Kollanyi / Philip N. Howard / Samuel C. Woolley*, a.a.O. (Fn. 10).

24 In der Soziologie findet sich diese Vorstellung ebenso bei *Jürgen Habermas* wie in der Konversations- und der Netzwerkanalyse. Vgl. *Jürgen Habermas*, Theorie des kommunikativen Handelns, Frankfurt am Main 1995; *Jack Sidnell*, Conversation Analysis. An Introduction, Chichester 2010; *Robert Ackland*, Web Social Science. Concepts, Data and Tools for Social Scientists in the Digital Age, Los Angeles / London 2013; *David R. Gibson*, Participation Shifts. Order and Differentiation in Group Conversation, in: Social Forces, 81. Jg. (2003), H. 4, S. 1335 – 1380.

25 Hierbei stellen sie fest, dass es zumindest einigen Bots gelingt, nicht nur von anderen Bots, sondern auch von Menschen Retweets zu erhalten, vgl. *Alessandro Bessi / Emilio Ferrara*, a.a.O. (Fn. 16).

Kommunikation; sie dienen vielmehr unidirektional der Verbreitung von Nachrichten. Die Analyse der Position von Bots in Retweet-Netzwerken bringt folglich auch keine Erkenntnisse über deren Rolle in Konversationen, sondern über deren erfolgreichen Einsatz als Informationsdistributoren beziehungsweise Verbreitungsmedien.²⁶

Folgt man einer Differenzierung verschiedener Ebenen der Kommunikation auf Twitter, wie sie *Axel Bruns* und *Hallvard Moe* anbieten²⁷, wird damit in der bisherigen Forschung alles untersucht, bloß keine Konversationen. Die Analyse von *Bence Kollanyi* u.a. verbleibt komplett auf der Makro-Ebene der Twitterkommunikation bei der Untersuchung so genannter „hashtagged exchanges“²⁸, womit schlichtweg die Summe aller Nachrichten gemeint ist, die zu bestimmten Hashtags versendet werden. Konversationen – *Axel Bruns* und *Hallvard Moe* sprechen mit Blick auf Twitter von „@Reply Conversations“²⁹ – finden dagegen auf der Mikro-Ebene der Twitterkommunikation statt, indem hier User aufeinander eingehen und einander antworten. Retweets, wie sie von *Alessandro Bessi* und *Emilio Ferrara* untersucht werden, begreifen *Axel Bruns* und *Hallvard Moe* als Mittel, das es ermöglicht, „to move tweets across layer boundaries“³⁰, indem beispielsweise durch Retweets einzelne Nachrichten aus der persönlichen Öffentlichkeit von Nutzern weiterverbreitet werden oder umgekehrt bestimmte Nachrichten in ihre persönliche Öffentlichkeit „importiert“ werden. Retweets können damit Konversationen hervorrufen oder aus diesen hervorgehen, sie werden aber selbst erst dann zum Bestandteil von Konversationen, wenn sie von anderen beantwortet werden.

Im Vergleich mit anderen Kommunikationsformen spielen Konversationen, wie sie sich auf der Mikro-Ebene etablieren, nur eine untergeordnete Rolle auf Twitter, das insgesamt eher als Informations-, denn als Kommunikationsnetzwerk (im Sinne wechselseitigen Austauschs) charakterisiert werden kann.³¹ Dennoch erscheint ein analytischer Fokus auf entsprechende Konversationen im engeren Sinne vielversprechend, denn einerseits wurden diese insgesamt auf Twitter bisher kaum untersucht³², andererseits erlaubt der Blick auf bidirektionale Kommunikationsprozesse auch – im Unterschied zur Untersuchung von Retweet-Kaskaden und „hashtagged exchanges“ –, in den Blick zu nehmen, wie (neutral, ablehnend, zustimmend) sich Nutzer aufeinander beziehen und so tatsächlich Relevanz, Charakter und Prozess von Online-Konversationen und deliberativen Prozessen auf Twitter sowie die Rolle von Socialbots in diesen zu erfassen. Damit Letzteres gelingen kann, ist jedoch ein valides Verfahren zur Bot-Erkennung notwendig, das es zuverlässig ermöglicht, automatisierte Accounts und ihre Aktivitäten in Online-Konversationen zu identifizieren. Die in der bisherigen Forschung zur Anwendung gebrachten Verfahren scheinen hierfür – zumindest mit Blick auf die Twitterkommunikation zur Präsidentschaftswahl in den USA

26 Mit *Niklas Luhmann*, Die Gesellschaft der Gesellschaft, Band 2, Frankfurt 1998, S. 202, verstehen wir Verbreitungsmedien als solche Medien, die den Empfängerkreis einer Kommunikation erweitern und bestimmen, indem sie dieselbe Information an zahlreiche Empfänger/innen ausspielen.

27 Vgl. *Axel Bruns* / *Hallvard Moe*, a.a.O. (Fn. 17).

28 Ebenda, S. 17.

29 Ebenda, S. 19.

30 Ebenda, S. 22.

31 Vgl. *Robert Ackland*, a.a.O. (Fn. 24), S. 74.

32 Vgl. *Ryosuke Nishi* / *Taro Takaguchi* / *Keigo Oka* / *Takanori Maehara* / *Masashi Toyoda* / *Ken-ichi Kawarabayashi* / *Naoki Masuda*, Reply Trees in Twitter. Data Analysis and Branching Process Models, in: Social Network Analysis and Mining, 6. Jg. (2016), H. 1, Artikel 26.

– allerdings nur bedingt geeignet. Dies ist der zweite problematische Punkt der existierenden Forschung zur US-Wahl.

2.2. Unzuverlässige Verfahren zur Bot-Erkennung

Bisher existiert kein einheitliches Verfahren zur Bot-Erkennung. Entsprechend kamen in der Forschung zur US-Wahl auch unterschiedliche Verfahren zum Einsatz, die stark voneinander abweichende Ergebnisse produzieren. So nutzten *Bence Kollanyi* u.a. einen verhältnismäßig simplen Ansatz zur Identifikation von Socialbots, indem sie als einzigen Indikator hierfür die Anzahl der gesendeten Tweets pro Tag wählten. Als automatisiert galten ihnen solche Accounts, die durchschnittlich mehr als 50 Tweets täglich versenden.³³ Auf dieser Grundlage identifizierten sie in ihrem Datenkorpus „4.160 highly automated accounts“³⁴, die damit etwas mehr als 0,1 Prozent aller untersuchten Nutzerkonten ausmachen, allerdings als hochgradig aktive Accounts zugleich für bis zu 25 Prozent der untersuchten Tweets verantwortlich sind.³⁵

Demgegenüber wandten *Bessi* und *Ferrara* ein ungleich komplexeres Verfahren zur Bot-Erkennung an, indem sie den „Botometer“³⁶ verwendeten, der an der Indiana University entwickelt wurde. Dabei handelt es sich um eine Software, die nicht nur die Häufigkeit von Tweets zur Identifikation von Socialbots heranzieht, sondern zusätzlich zahlreiche weitere Merkmale analysiert. Zu diesen gehören unter anderem die Inhalte gesendeter Tweets sowie die Art der Vernetzung eines Accounts oder dessen Profilinformationen.³⁷ Aus der Gesamtheit der untersuchten Merkmale errechnet der Botometer dann einen „Botscore“ zwischen 0 und 1, der anzeigt, ob ein Account (wahrscheinlich) menschlich gesteuert oder automatisiert betrieben wird.³⁸

Indem der Botometer nicht nur die Häufigkeit des Versendens von Nachrichten untersucht, sondern auch andere (verdächtige) Merkmale einbezieht, besteht die Möglichkeit, dass er auch solche Konten als botgesteuert identifiziert, die vergleichsweise selten twittern, dafür aber andere Merkmale aufweisen, die stark auf eine Automatisierung hinweisen. Beispielsweise gelten das Fehlen umfassender Profilinformationen oder geographischer Metadaten in den Tweets neben der Aktivität als wichtige Indikatoren, um die Automatisierung eines Accounts festzustellen.³⁹ In der Konsequenz führt dies zu einer ganz anderen Einschätzung der Anzahl an Bots im eigenen Datenkorpus, als dies bei *Bence Kollanyi* u.a. der Fall ist. So identifizierten *Alessandro Bessi* und *Emilio Ferrara* in einer Stichprobe, welche die 50.000 aktivsten Accounts aus ihrem Korpus enthält⁴⁰, 7.183 botgesteuerte Accounts. Dies

33 Vgl. *Bence Kollanyi / Philip N. Howard / Samuel C. Woolley*, a.a.O. (Fn. 10); *dies.*, a.a.O. (Fn. 18).

34 Vgl. *dies.*, a.a.O. (Fn. 10).

35 Vgl. *Bence Kollanyi / Philip N. Howard / Samuel C. Woolley*, a.a.O. (Fn. 10), S. 3.

36 Ursprünglich hieß diese Software BotORNot. Sie wurde aber vor einiger Zeit umbenannt.

37 Vgl. *Clayton A. Davis / Onur Varol / Emilio Ferrara / Alessandro Flammini / Filippo Menczer*, BotOrNot, in: *Jacqueline Bourdeau / Jim A. Hendler / Roger Nkambou Nkambou / Ian Horrocks / Ben Y. Zhao* (Hrsg.), WWW'16 Companion. Proceedings of the 25th International Conference on World Wide Web, May 11-15, 2016, Montreal, Canada, S. 273 – 274.

38 Ab einem Wert von 0,5 werden Accounts als botgesteuert eingestuft.

39 Vgl. *Alessandro Bessi / Emilio Ferrara*, a.a.O. (Fn. 16).

40 Da die Klassifizierung von Accounts mittels Botometer nicht nur komplex, sondern auch auf die Benutzung der Twitter API als Schnittstelle angewiesen ist, die einige Beschränkungen in der

sind rund 15 Prozent der Accounts im Korpus (im Unterschied zu 0,1 Prozent bei *Bence Kollanyi* u.a.).

Diesen prozentualen Anteil behandeln sie als repräsentativ für ihren Gesamtkorpus, so dass sie zum Schluss gelangen, dass sich darin insgesamt circa 400.000 Socialbots befinden. Trotz eines etwas kleineren Korpus liegt die Zahl der identifizierten Bots bei *Bessi* und *Ferrara* damit 100mal höher als bei *Kollanyi* u.a. Hieraus folgt, dass entweder eine der beiden Erkennungsmethoden zu gravierend falschen Ergebnissen führen muss oder gar beide. Bei *Kollanyi* u.a. finden sich keinerlei Hinweise auf eine entsprechende Kontrolle der gewählten Methode zur Bot-Erkennung. Dagegen verweisen *Bessi* und *Ferrara* darauf, dass der Botometer mit 95-prozentiger Wahrscheinlichkeit richtige Ergebnisse zeitige. Grundlage für diese Einschätzung ist jedoch nicht die Validierung anhand der eigenen Daten, sondern eines anderen älteren Datenkorpus, die durch die Entwickler des Botometers vorgenommen wurde. Diese weisen darauf hin, dass die Genauigkeit der Trefferquote bei anderen Daten schlechter sein könnte und bemerken zudem, dass der Botometer oft fälschlicherweise Accounts als automatisiert einstuft, die von Organisationen oder zu Werbezwecken eingerichtet wurden⁴¹, so dass Anlass besteht, auch die Validität des Botometers zu prüfen.

Eine solche Prüfung haben wir anhand eines eigenen Datenkorpus vorgenommen, der knapp 6,5 Millionen Tweets von 1,44 Millionen Usern enthält, die wir am 26. September 2016 während der ersten Fernsehdebatte zwischen *Hillary Clinton* und *Donald Trump* sowie in den 15 Minuten vor und nach der Debatte erhoben haben.

Aufgrund der großen Datenmenge war es nur möglich, eine kleine Teilmenge der Accounts manuell zu prüfen. Da es in erster Linie um die Rolle von Bots in Konversationen geht, wurden in einem ersten Schritt nur diejenigen Accounts überprüft, die tatsächlich an entsprechenden Konversationen beteiligt waren. Auf diese Weise konnte die Menge der zu überprüfenden Konten bereits deutlich auf 26.564 Nutzer reduziert werden⁴². Diese wurden dann mithilfe des Botometer als mutmaßlich menschlich oder automatisch gesteuert kategorisiert. Im Ergebnis wurden auf diese Weise 1.537 Accounts (5,7 Prozent) als Socialbots eingestuft und 25.027 Konten als menschlich gesteuert, indem – wie in der bisherigen Literatur üblich – diejenigen als automatisiert eingestuft wurden, die einen Botscore von 0,5 oder höher erreichten. Abbildung 1 zeigt die Verteilung der Botscores auf die Zahl der User, von denen ein Großteil einen Wert zwischen 0,2 und 0,4 erhielt. Die als Bot kategorisierten Accounts sind grau hinterlegt.

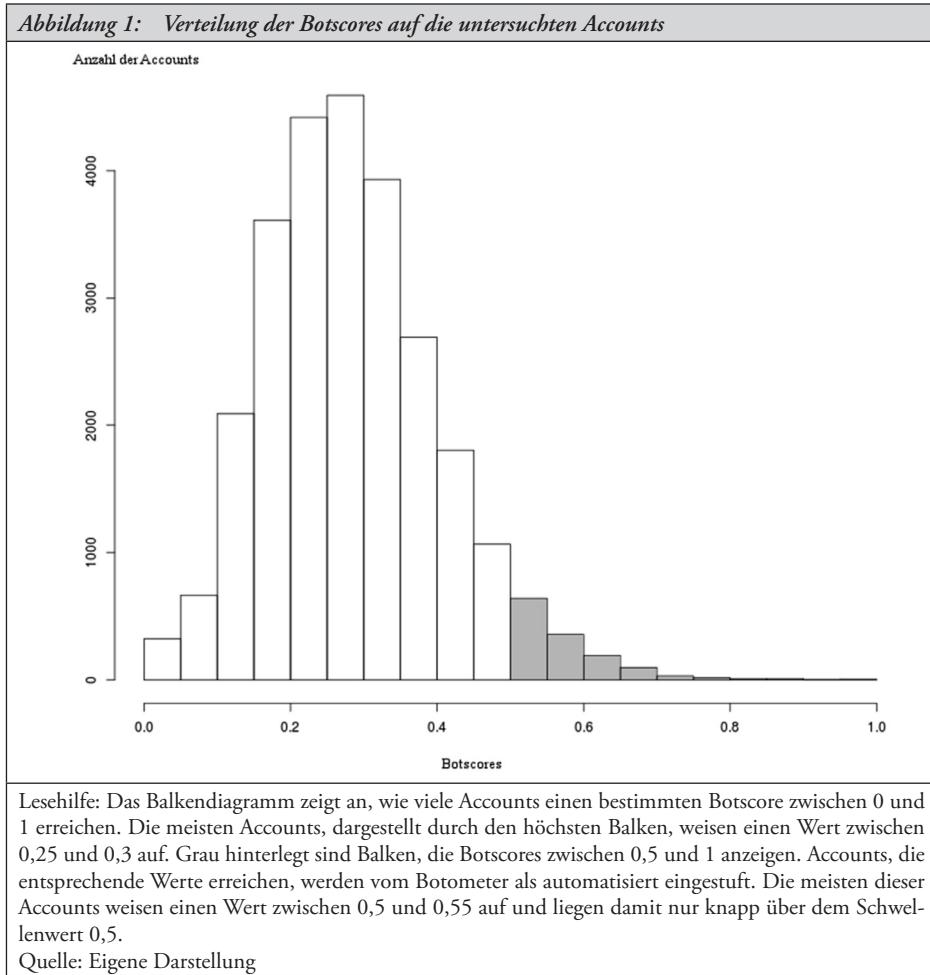
Im Anschluss an die Kategorisierung der Accounts als menschlich oder automatisch gesteuert wurde aus beiden Gruppen eine Zufallsstichprobe gezogen, die aus 550 als menschlich und 100 als automatisiert eingestuften Konten bestand.⁴³ Sämtliche Accounts wurden

Nutzung aufweist, werten *Alessandro Bessi* und *Emilio Ferrara* anders als *Bence Kollanyi* u.a. nicht alle Accounts in ihrem Korpus aus, sondern nur die aktivsten 50.000, von denen sie 7.183 als botgesteuert identifizieren.

41 Vgl. *Clayton A. Davis / Onur Varol / Emilio Ferrara / Alessandro Flammini / Filippo Menczer*, a.a.O. (Fn. 37), S. 2; *Onur Varol / Emilio Ferrara / Clayton A. Davis / Filippo Menczer / Alessandro Flammini*, a.a.O. (Fn. 10). Dies nehmen *Alessandro Bessi* und *Emilio Ferrara* aber nicht zum Anlass, um die Akkuratheit des Botometer für den ihnen vorliegenden Datenkorpus zu testen, so dass weder *Bence Kollanyi* u.a. noch *Alessandro Bessi* und *Emilio Ferrara* eine angemessene Validierung ihrer Bot-Erkennungsmethode anbieten.

42 Das heißt, nur knapp zwei Prozent der User partizipieren an Konversationen.

43 Es sind mehr menschlich eingestufte Accounts im Sample, da deren Anzahl auch insgesamt deutlich größer ist.

Abbildung 1: Verteilung der Botscores auf die untersuchten Accounts

gesichtet, um ihr Profil und ihre Aktivitäten in den Blick zu nehmen und das Ergebnis der automatisierten Kategorisierung zu überprüfen.⁴⁴ Zusätzlich wurde hierbei das kostenlose Web Analytics-Tool „Socialbearing“ hinzugezogen, das unter anderem die durchschnittliche

⁴⁴ Die Überprüfung erfolgte mittels qualitativer Inhaltsanalyse, vgl. hierzu *Udo Kuckartz, Qualitative Inhaltsanalyse. Methoden, Praxis, Computerunterstützung*, Weinheim 2016. In deren Rahmen wurden zunächst die Accounts entlang von fünf Oberkategorien (Accounttyp, Aktivitätsgrad, Aktivitätstyp, Vernetzungsgrad sowie Vernetzungstyp) kategorisiert, um darauf aufbauend zu entscheiden, ob es sich um menschlich oder maschinell gesteuerte Accounts handelte. Die Intercoder-Übereinstimmung lag bei 80 Prozent und es wurde beschlossen, nur diejenigen Accounts als Bots zu kategorisieren, die übereinstimmend von zwei Kodierenden als solche identifiziert wurden. Konten, bei denen Uneinigkeit herrschte, wurden als „nicht entscheidbar“ kategorisiert. Die manuelle Kategorisierung erfolgte durch *Florian Muhle* sowie Teilnehmende des Seminars „Socialbots in der Online-Kommunikation“, das im Sommersemester 2018 an der Universität Bielefeld durchgeführt wurde.

che Tweethäufigkeit eines Accounts pro Tag anzeigt und auf diese Weise ermöglicht, auch die Validität des Bot-Erkennungsverfahren von *Bence Kollanyi* u.a. zu testen.

Das Ergebnis zeigt, dass beide Verfahren nicht zuverlässig in der Lage sind, botgesteuerte von menschlich gesteuerten Accounts zu unterscheiden. So konnten von den einhundert geprüften Accounts, die vom Botometer als botgesteuert kategorisiert wurden, lediglich 16 als tatsächlich botgesteuert identifiziert werden. Bei 69 Konten wurde dagegen mit Sicherheit festgestellt, dass diese nicht als Socialbots einzustufen waren.⁴⁵ Beim überwiegenden Teil dieser durch den Botometer falsch kategorisierten Accounts handelte es sich um Konten von Organisationen (insgesamt 27) oder Personen öffentlichen Interesses (insgesamt 30). Dies weist darauf hin, dass in der Debatte zur Präsidentschaftswahl – im Vergleich zur durchschnittlichen Twitterkommunikation – offensichtlich solche Account-Typen überrepräsentiert sind, bei denen die Entwickler des Botometers eingestehen, dass sie von der Software häufig nicht richtig zugeordnet werden.

Dies ist aber nicht das einzige Problem der Bot-Erkennung mittels Botometer. Interesanterweise wurde in der vorliegenden Analyse auch festgestellt, dass sich im Sample der von der Software als menschlich eingestuften Accounts insgesamt 53 befanden, die richtigerweise als botgesteuert zu kategorisieren gewesen wären. Während es sich bei fast 70 Prozent der als Socialbot eingestuften Accounts um „false positives“⁴⁶ handelte, erzeugt der Botometer also auch einige „false negatives“ (knapp zehn Prozent), bei denen er nicht erkennt, dass es sich um automatisierte Konten handelt. Deutlich wird somit, dass der Botometer mit Blick auf die Twitterkommunikation zur US-Wahl keine zuverlässige Identifikation von Bots gewährleisten kann.

Gleiches gilt für die einfachere Erkennungsmethode von *Bence Kollanyi* u.a. So finden sich im gesamten Sample (alle 650 Accounts) 26 Konten, die durchschnittlich häufiger als 50mal täglich Nachrichten senden. Von diesen konnten jedoch nur 17 als tatsächlich botgesteuert eingestuft werden (also knapp zwei Drittel). Bei den verbliebenen neun handelt es sich – ganz ähnlich wie bei den durch den Botometer ermittelten „false positives“ – um Accounts von Medienorganisationen (zum Beispiel der Washington Post) oder Personen öffentlichen Interesses (zum Beispiel *Rick Wilson*, ein konservativer politischer Berater und Medienexperte). Daneben brachte die manuelle Analyse zahlreiche „false negatives“ hervor, denn die große Mehrzahl der von uns manuell verifizierten Socialbots twitterte seltener als 50mal täglich und wird entsprechend von *Kollanyi* u.a. nicht erkannt. Das heißt, einerseits sind – anders als von den Autoren nahegelegt – nicht alle hochgradig aktiven Konten automatisch Bots, andererseits werden die zahlreichen weniger aktiven Bot-Accounts durch das gewählte Verfahren nicht erkannt, so dass auch dieses keine zuverlässigen Ergebnisse zeitigt.

Die zur Verfügung stehenden Mittel der automatisierten Bot-Erkennung scheinen also nicht tauglich, um zu validen Ergebnissen zu gelangen. Gerade mit Blick auf die „false positives“ ist dies besonders problematisch, da es sich bei diesen überwiegend um professionell betriebene beziehungsweise organisatorische Accounts handelt die oftmals zentrale Akteure

45 Daneben haben wir vier Accounts als semi-automatisiert eingestuft, da sie teilweise automatisierte Skripte verwenden. Bei elf Accounts waren wir nicht in der Lage, diese klar zuzuordnen, da diese gelöscht (2), gesperrt (2) oder schwer einzuordnen waren (7).

46 Damit werden solche Accounts bezeichnet, die fälschlicherweise als Bots identifiziert werden, vgl. *Onur Varol / Emilio Ferrara / Clayton A. Davis / Filippo Menczer / Alessandro Flammini*, a.a.O. (Fn. 10).

innerhalb der Twitterkommunikation im Allgemeinen und dort ablaufender Konversationen im Besonderen sind (vgl. Abschnitt 3)⁴⁷, so dass eine Einstufung dieser Konten als Socialbots auch zu falschen Annahmen hinsichtlich ihrer Bedeutung innerhalb der Online-Kommunikation führt. Folglich ist es dringend nötig, existierende Verfahren der automatisierten Bot-Erkennung zu hinterfragen, weiter zu entwickeln und stets anhand konkret vorliegender Daten manuell zu validieren. Dasselbe gilt für die in der bisherigen Forschungspraxis verwendeten Methoden zur Feststellung der politischen Agenda der als Socialbot identifizierten Accounts.

2.3. Unzuverlässige Verfahren zur Erkennung der politischen Agenda von Bots

Bence Kollanyi u.a. sowie *Alessandro Bessi* und *Emilio Ferrara* wählten nahezu identische Vorgehensweisen zur Erkennung der politischen Agenda von Socialbots, indem sie auf die automatisierte Auswertung der Verwendung von Hashtags setzten.⁴⁸ Die Forscher schauten sich also nicht ganze Nachrichten an, sondern nur ausgewählte Aspekte, denen eine Aussagekraft hinsichtlich der Fragestellung zugesprochen wurde. Damit folgten sie einer für Big Data-Forschung typischen Strategie der Reduktion von Datenkomplexität, um überhaupt eine automatisierte Analysierbarkeit herzustellen.⁴⁹ Konkret legten sie hierzu *a priori* fest, welche Hashtags in ihrem Korpus neutral und welche positiv mit *Clinton* oder mit *Trump* assoziiert waren und werteten darauf aufbauend aus, wie viele dieser Hashtags von als Bot klassifizierten Accounts verwendet wurden. Dies sollte es erlauben, die politische Agenda der Bot-Accounts herauszuarbeiten.

Prinzipiell erscheint ein solches Vorgehen plausibel, und es ist auch nachvollziehbar, etwa die Verwendung von Hashtags wie #MAGA oder #votetromp als eindeutig positiv mit *Trump* assoziiert zu deuten, während Hashtags wie #imwithher oder #hillarysupporter positiv mit *Clinton* in Verbindung gebracht werden.⁵⁰ Dies gilt jedoch nicht für alle Hashtags, die von den Forschungsgruppen un hinterfragt als positiv mit *Clinton* oder *Trump* assoziiert betrachtet wurden. So behandelten beide Teams die Hashtags #trump und #hillary, die die am häufigsten verwendeten einschlägigen Hashtags waren⁵¹, jeweils als positiv mit der referierten Person assoziiert, ebenso die Hashtags #hillaryclinton und #donaldtrump⁵². Das Problem hieran ist aber, dass Zweifel an der Eindeutigkeit dieser Hashtags angebracht sind, da es sich nur um die beiden Namen handelt, ohne dass diese – anders als etwa #votetromp oder #imwithher – in irgendeiner Form eine Wertung erfahren. Daher wurden auch diese Hashtags einer manuellen Validierung unterzogen. Konkret wurden je 100 Tweets, die die

47 Vgl. *Luca Rossi / Matteo Magnani*, Conversation Practices and Network Structure in Twitter, in: Proceedings of the Sixth International AAAI Conference on Weblogs and Social Media 2012, S. 563 – 566, S. 565.

48 Unterschiede bestehen lediglich hinsichtlich der Auswahl der Hashtags, die ausgewertet wurden.

49 Vgl. *Danah Boyd / Kate Crawford*, Critical Questions for Big Data. Provocations for a Cultural, Technological, and Scholarly Phenomenon, in: Information, Communication & Society, 15. Jg. (2012), H. 5, S. 662 – 679, S. 667.

50 Vgl. *Ryosuke Nishi / Taro Takaguchi / Keigo Oka / Takanori Maehara / Masashi Toyoda / Ken-ichi Kawarabayashi / Naoki Masuda*, a.a.O. (Fn. 32), S. 301.

51 Vgl. die Übersicht bei *Alessandro Bessi / Emilio Ferrara*, a.a.O. (Fn. 16).

52 Vgl. ebenda; *Bence Kollanyi / Philip N. Howard / Samuel C. Woolley*, a.a.O. (Fn. 10); *dies.*, a.a.O. (Fn. 18).

Hashtags #trump, #donald trump sowie #hillary und #hillaryclinton enthielten⁵³, daraufhin untersucht, ob diese tatsächlich positiv oder aber negativ oder neutral auf die jeweilige Person verwiesen.⁵⁴

Die Ergebnisse sprechen eine deutliche Sprache: Von den Tweets, die im Sample das Hashtag #trump verwendeten, bezogen sich lediglich neun positiv auf *Donald Trump*. Demgegenüber positionierten sich 72 Tweets eindeutig negativ (zum Beispiel „#Trump is a MORON. Did he go to ANY DEBATE coaching sessions whatsoever??? #debatenight“), während 19 Tweets neutrale Aussagen trafen. Ähnlich sieht dies mit dem Hashtag #donald-trump aus. Von 100 Tweets, die dieses Hashtag enthielten, bezogen sich 77 negativ, 17 neutral und nur sechs positiv auf *Donald Trump*. Das heißt, beide Hashtags sind überwiegend negativ mit *Donald Trump* assoziiert, wurden aber fälschlicherweise als pro-Trump-Hashtags behandelt. Berücksichtigt man, dass das Hashtag #trump das mit Abstand am häufigsten genutzte im gesamten „hashtagged exchange“ zur Präsidentschaftswahl zu sein scheint⁵⁵, folgt hieraus, dass die automatisierte Analyse wie sie von *Bence Kollanyi* u.a. sowie *Alessandro Bessi* und *Emilio Ferrara* umgesetzt wurde, eindeutig falsche Ergebnisse generiert.

Bei den Hashtags #hillary und #hillaryclinton sieht es etwas differenzierter aus. Aber auch diese sind tendenziell negativ mit der Kandidatin assoziiert, nicht positiv. So wurde das Hashtag #hillary in 56 von 100 Fällen im Rahmen negativer Erwähnungen der Kandidatin verwendet, das Hashtag #hillaryclinton in 42 Fällen. In positivem Zusammenhang standen die Hashtags dagegen nur in 37 beziehungsweise 30 Fällen, während sie 21- beziehungsweise 14mal in neutralem Zusammenhang verwendet wurden. Auch hier zeigt die Betrachtung der Hashtags im Kontext ihres tatsächlichen Gebrauchs also, dass sie nicht als Indikatoren geeignet sind, um pro-Clinton-Aussagen zu identifizieren.

Doch selbst, wenn die Autoren auf Grundlage einer Validierung beschlossen hätten, die Hashtags #trump, #donaldtrump, #hillary und #hillaryclinton nicht als Indikatoren zu verwenden⁵⁶, würde ein weiteres Problem fortbestehen: Zahlreiche Tweets enthielten keinerlei eindeutig wertende Hashtags, sind aber dennoch klar als parteiisch zu identifizieren, sobald man sich ihren Inhalt ansieht. Wie eine weitere manuelle Analyse von 100 Tweets mit den von *Kollanyi* u.a. beziehungsweise *Bessi* und *Ferrara* als neutral erachteten Hashtags #debatenight2016, #debates2016, #election2016 und #debates zeigt, gilt nur in 22 Prozent der Fälle, dass die Nachrichten auch tatsächlich neutral waren. Die Mehrheit der Nachrichten (40 von 100), die neutrale Hashtags verwendeten, nahm dagegen negativ auf *Donald Trump* Bezug⁵⁷, wie exemplarisch an dem Tweet „Donald Trump only wants to make America great again for rich white Christians #debatenight“ sichtbar wird, der den Republikanischen Kandidaten klar kritisiert.

53 Alessandro Bessi und Emilio Ferrara verwenden insgesamt neun Hashtags, darunter die vier von uns überprüften. Bence Kollanyi u.a. verwenden dagegen mehr als 40 Hashtags, kommen aber zu einem ganz ähnlichen Ergebnis wie Alessandro Bessi und Emilio Ferrara.

54 Auch hier kam die qualitative Inhaltsanalyse zum Einsatz.

55 Vgl. Alessandro Bessi / Emilio Ferrara, a.a.O. (Fn. 16).

56 Diesen Weg wählen Marian-Andrei Rizoiu / Timothy Graham / Rui Zhang / Yifei Zhang / Robert Ackland / Lexing Xie, #DebateNight. The Role and Influence of Socialbots on Twitter During the 1st 2016 U.S. Presidential Debate, 12. International Conference on Web and Social Media (ICWSM), Stanford, USA, 25. bis 28. Juni 2018.

57 Hillary Clinton wird 13mal negativ erwähnt.

Damit erhärtet sich der Eindruck, dass die Methode der Hashtag-Analyse, wie sie in den beiden genannten Arbeiten eingesetzt wurde, insgesamt zu einer deutlichen Fehleinschätzung der politischen Positionierungen der untersuchten Accounts führt. Dies verweist deutlich auf die Grenzen automatisierter Inhaltsanalysen und legt nahe, diese entweder besonders gut zu validieren⁵⁸ oder alternativ ganz auf manuelle Inhaltsanalysen zu setzen, die auch kontextspezifische Interpretationen des Untersuchungsmaterials erlauben und damit im Vergleich zu automatisierten Analysen als deutlich valide Instrumente erscheinen – natürlich zu Lasten der Menge analysierbarer Daten.⁵⁹

2.4. Zusammenfassung

Nimmt man die vorgetragenen Einwände gegen die bisherige Forschung zu Socialbots in der politischen Online-Kommunikation ernst, wird deutlich, dass ein verändertes Forschungsdesign, das weniger Daten untersucht, diese aber manuell und umfassender analysiert, höhere Validität verspricht. Unseres Erachtens folgt hieraus die Einsicht, nicht allein auf automatisierte Big Data-Analysen zur politischen Online-Kommunikation zu setzen, die möglicherweise die Validitätsstandards sozialwissenschaftlicher Forschung nicht erfüllen.⁶⁰ Diese müssen zwingend – notwendigerweise auf Grundlage kleinerer Samples – validiert und durch „Small Data-Ansätze“ ergänzt werden.⁶¹

3. Socialbots in politischen Online-Konversationen: Erste Erkenntnisse aus dem Projekt „Socialbots as Political Actors?“

Im Rahmen des Forschungsprojektes „Socialbots as Political Actors?“⁶² versuchen wir gegenwärtig ein Verfahren zu entwickeln, das die bis hierhin skizzierten Probleme existierender Arbeiten zum Einsatz von Socialbots in der politischen Kommunikation zu vermeiden sucht und sich im Rahmen eines Small Data-Ansatzes auf die Analyse von Online-Konversationen in einem engeren Sinne konzentriert. Konkret besteht das Konzept aus vier Bausteinen, die in ihrem Zusammenspiel zu validen und tiefer gehenden Einsichten in Bedeutung und Charakter politischer Online-Konversationen zur Präsidentschaftswahl in den USA sowie in die Rolle von Bots darin führen sollen. Diese Bausteine sind:

- (1) Methoden der Netzwerkanalyse, um formale Charakteristika von Online-Konversationen herauszuarbeiten,
- (2) manuelle Kategorisierung automatisierter Accounts (wie in Abschnitt 2.2. dargestellt),

58 Vgl. *Justin Grimmer / Brandon M. Stewart*, Text as Data. The Promise and Pitfalls of Automatic Content Analysis Methods for Political Texts, in: *Political Analysis*, 21. Jg. (2013), H. 3, S. 267 – 297, S. 271.

59 Vgl. *Jessica Eispänner / Mark Dang-Anh / Caja Thimm*, Computer-Assisted Content Analysis of Twitter Data, in: *Katrin Weller / Axel Bruns / Jean Burgess / Merja Mahrt / Cornelius Puschmann* (Hrsg.), *Twitter and Society*, New York 2014, S. 97 – 108, S. 102.

60 Vgl. hierzu auch *Danah Boyd / Kate Crawford*, a.a.O. (Fn. 49), S. 666 ff.

61 Vgl. hierzu auch *Joyojeet Pal / A'Ndre Gonawela*, Studying Political communication on Twitter. The Case for Small Data, in: *Current Opinion in Behavioral Sciences*, 18. Jg. (2017), S. 97 – 102, S. 100.

62 Vgl. Fn. 6 für weitere Informationen zum Forschungsprojekt.

- (3) qualitative Inhaltsanalysen von Tweets zur Bestimmung der politischen Agenda von Socialbots (wie in Abschnitt 2.3. dargestellt) und
- (4) Analysen exemplarischer Konversationen, um die ihnen zugrunde liegenden Deutungsmuster und Aushandlungsprozesse zu rekonstruieren.⁶³

Dies stellt eine Mischung quantitativer Forschung dar, wie sie sich in der „Web Social Science“⁶⁴ durchgesetzt hat, mit qualitativen Verfahren der Datenauswertung, die auf die kontextsensitive Interpretation kleiner Datenmenge zielen. Dabei gehen wir mit *Joyojeet Pal* und *A'Ndre Gonawela* davon aus, dass die unterschiedlichen Perspektiven nicht konkurrieren, sondern dazu beitragen, ein vollständig(er)es Bild der politischen Online-Kommunikation zu zeichnen.⁶⁵ Im Moment befindet sich das Projekt noch in der Phase der Datenauswertung, so dass insbesondere mit Blick auf den Baustein vier noch keine Ergebnisse präsentiert werden können und alle anderen Resultate noch vorläufigen Charakter besitzen. Dennoch verfügen wir über erste Erkenntnisse über Bedeutung und Charakter der Online-Konversationen zur US-Wahl, die ein deutlich anderes Bild zeichnen als die bisherige Forschung.

Konversationen werden hier aus einer netzwerkanalytischen Perspektive als spezifischer Netzwerk-Typus begriffen, der sich im Unterschied etwa zu Informationsnetzwerken, die der unidirektionalen Verbreitung von Nachrichten dienen, durch bidirektionale Beziehungen auszeichnet, so dass sich in Konversationen (zumindest schwache) Formen der Reziprozität herausbilden. Bei Twitter realisiert sich dieser Typus in Form der von *Axel Bruns* und *Hallvard Moe* so bezeichneten „@Reply-Conversations“⁶⁶, die entstehen, wenn User sich über die @Reply-Funktion des Dienstes aufeinander beziehen.

Netzwerkanalytisch lässt sich die Gesamtheit aller Konversationen als Graph darstellen, in dem die Accounts, die @Replies senden und empfangen, die Knoten bilden und die @Reply-Aktivitäten selbst als Kanten repräsentiert werden, die Relationen zwischen den Knoten stiften.⁶⁷ Die in dem Graphen zusammengefassten Einzel-Konversationen bilden die Komponenten, aus denen der Graph besteht. Im vorliegenden Fall enthält dieser 26.564 User⁶⁸ und besteht aus 6.073 Komponenten. Konversationen machen damit nur einen geringen Teil der Gesamtkommunikation zur US-Wahl aus, da lediglich 1,8 Prozent der Nutzer im Gesamtkorpus an entsprechenden Konversationen beteiligt waren.⁶⁹

Unterzieht man den Graphen einer näheren Betrachtung, fällt auf, dass dieser insgesamt nur eine sehr geringe Dichte sowie eine sehr geringe Reziprozitätsrate aufweist.⁷⁰ Dies bedeutet, dass die einzelnen Komponenten des Graphen (die einzelnen Konversationen) kaum miteinander verbunden waren und Antworten einzelner User nur sehr selten wieder-

⁶³ Zur Konversationsanalyse als qualitativer Methode der Netzwerkforschung, vgl. *Jan Fuhse*, Soziale Netzwerke. Konzepte und Forschungsmethoden, Konstanz 2016, S. 151 ff.

⁶⁴ *Robert Ackland*, a.a.O. (Fn. 24).

⁶⁵ Vgl. *Joyojeet Pal* / *A'Ndre Gonawela*, a.a.O. (Fn. 61), S. 100.

⁶⁶ *Axel Bruns* / *Hallvard Moe*, a.a.O. (Fn. 17), S. 19.

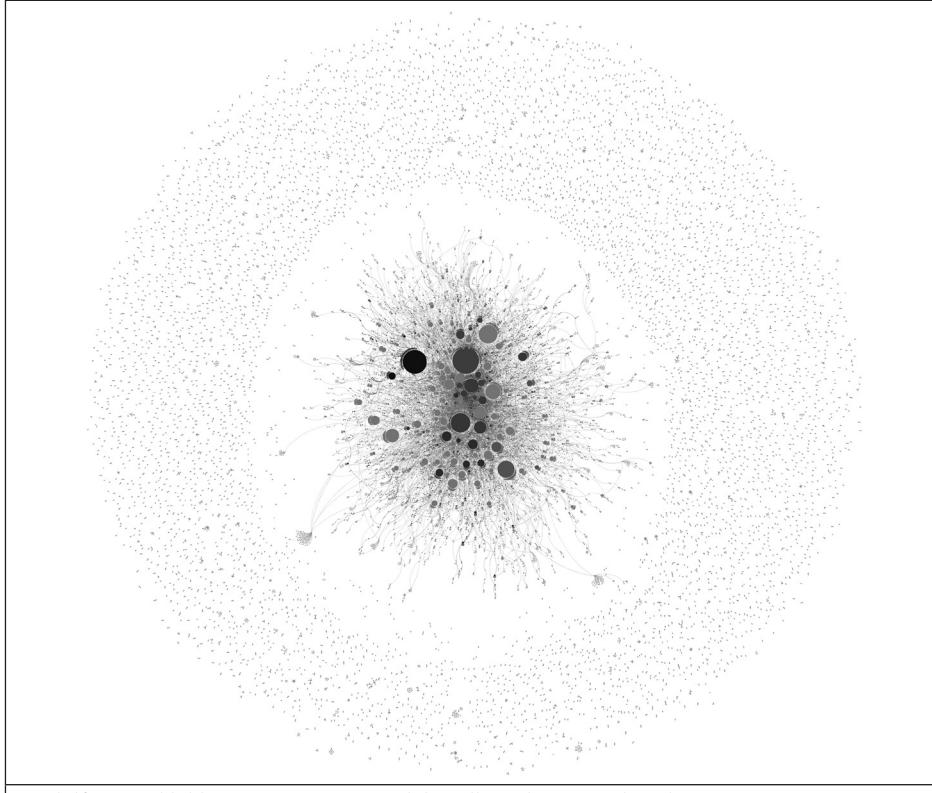
⁶⁷ Vgl. *Jan Fuhse*, a.a.O. (Fn. 63), S. 41.

⁶⁸ Wir zählen hierzu alle User im Korpus, die mindestens eine Antwort erhalten oder gesendet haben. Gelöscht wurden allerdings die Accounts der beiden Präsidentschaftskandidaten. Diese erhielten zwar formal zahlreiche Antworten. Pragmatisch handelte es sich hierbei allerdings um Erwähnungen.

⁶⁹ Die Daten, die *Alessandro Bessi* und *Emilio Ferrara* präsentieren, zeigen dies auch. Replies haben in ihrem Korpus lediglich einen Anteil von 1,65 Prozent des Volumens.

⁷⁰ Der Wert für die Dichte beträgt 3.5726052×10^{-5} , der Wert für die Reziprozität 0.0199929.

Abbildung 2: Visualisierung des Graphen, der die Gesamtheit der Konversationen darstellt



Lesehilfe: Die Abbildung zeigt eine Netzwerkdarstellung der Gesamtheit der Konversationen im untersuchten Graphen. Die einzelnen Punkte repräsentieren die beteiligten Accounts. Im Zentrum der Darstellung finden sich solche Accounts, die stark vernetzt sind. Hierbei sind diejenigen Accounts größer dargestellt, die viele Reaktionen Anderer generieren. Die Punkte, die sich am Rande um das Zentrum herum verteilen, stellen die Accounts dar, die kaum vernetzt sind. Deutlich wird, dass es einerseits einige wenige Accounts gibt, die große Aufmerksamkeit generieren und andererseits eine Vielzahl weitgehend unverbundener (und gleichsam an den Rand gedrängter) Accounts, deren Beiträge, keine größere Öffentlichkeit erreichen.

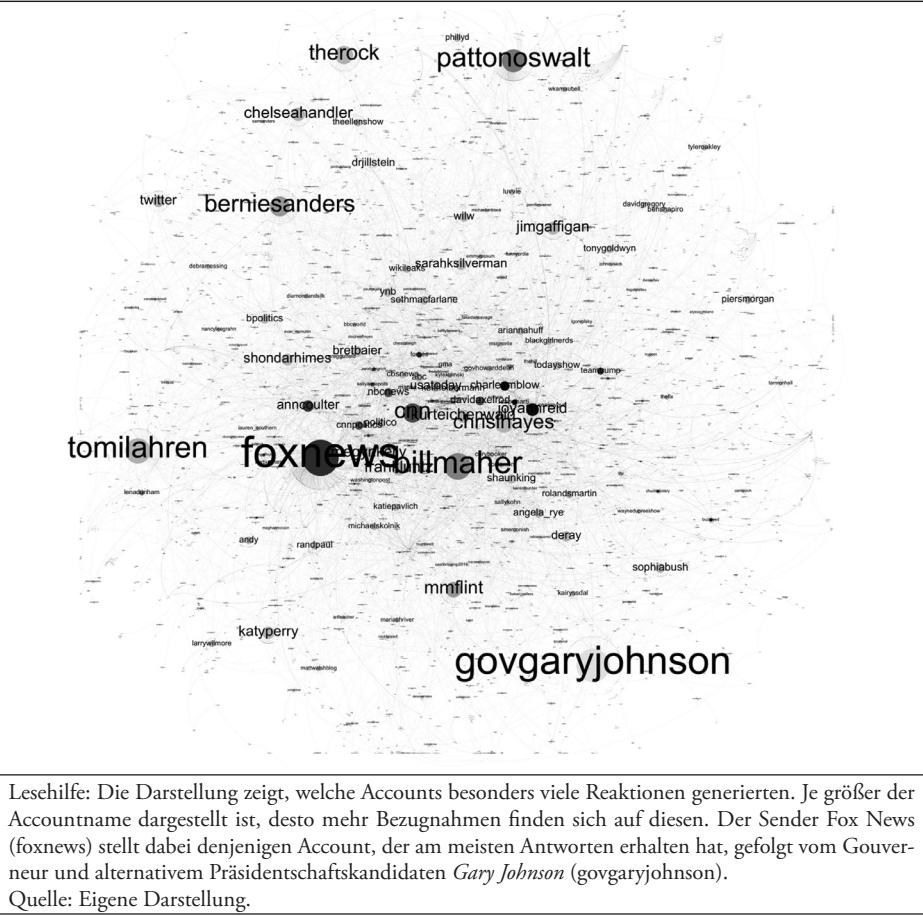
Quelle: Eigene Darstellung.

um eine Reaktion erfuhren. Tatsächlich fand die große Mehrzahl der Konversationen im Datenmaterial (4.949) lediglich zwischen zwei Usern statt, und es gibt insgesamt nur 19 Konversationen mit mehr als zehn Teilnehmenden. Dies bedeutet, dass sich selbst in den „persönlichen Öffentlichkeiten“⁷¹ der Twitter-Nutzer nur wenige Personen an Konversationen beteiligten und diese in der Regel sehr kurz blieben.⁷² Folglich entstanden kaum Prozesse öffentlicher Deliberation.

71 Jan-Hinrik Schmidt, Social Media, Wiesbaden 2018, S. 27 ff.

72 Dies deckt sich mit Erkenntnissen von Michael Klemm / Sascha Michel, Social TV und Politikaneignung. Wie Zuschauer die Inhalte politischer Diskussionssendungen via Twitter kommentieren, in: Zeitschrift für angewandte Linguistik, 60. Jg. (2014), H. 1, S. 3–35, S. 19, zum fernsehbegleitenden Twittern.

Abbildung 3: Darstellung der Accounts, die besonders viele Antworten erhalten



Die große Ausnahme bildet eine Konversation, an der 12.270 – und damit fast die Hälfte der Accounts im Graphen – beteiligt waren⁷³, so dass sich insgesamt ein Bild zahlreicher kurzer und unverbundener Konversationen auf der einen Seite sowie einer zentralen Konversation auf der anderen Seite ergibt. Letztere zeichnet sich wiederum dadurch aus, dass sich in ihr einige zentrale Akteure herausbildeten, die hochgradig vernetzt waren und insofern bedeutsam für den Zusammenhang der Konversation. Abbildung 2 zeigt diese Struktur des Graphen sehr deutlich. Die große Konversation steht im Zentrum, und diejenigen Accounts, die Reaktionen Anderer generieren, sind größer dargestellt als die weniger zentralen Accounts. Die kleinen Punkte im Umkreis repräsentieren dagegen die kaum vernetzten Accounts, die lediglich kurze Konversationen mit nur wenigen anderen Accounts führen.

73 Damit handelt es sich aufgrund der Anzahl der Teilnehmenden nicht mehr um eine Konversation im klassischen Sinne.

Sieht man sich die zentrale Konversation genauer an, wird schnell klar, dass die Hauptknoten nicht „normale“ User sind, sondern in erster Linie Medienorganisationen sowie berühmte Personen aus den Bereichen Unterhaltung und Politik.⁷⁴ So war es der Fernsehsender Fox News, der die meisten Antworten auf seine Beiträge erhielt, gefolgt vom libertären Präsidentschaftskandidaten *Gary Johnson* und dem Komiker *Bill Maher*. Abbildung 3 zeigt eine Übersicht über die Accounts, die besonders viele Antworten generierten. Deutlich zu erkennen sind hier zusätzlich zu den genannten Top 3-Accounts die Konten der Schauspieler *Patton Oswalt* (pattonoswalt) und *Dwayne Johnson* (therock) sowie der Journalistin *Tomi Lahren* (tomilahren) und des Demokratischen Politikers *Bernie Sanders* (berniesanders).

Typisch für genannte Accounts ist, dass sie viele Antworten generierten, selbst aber nicht auf Antworten reagierten, so dass sich hochgradig spezifische Kommunikationsstrukturen etablierten, die weniger einer üblichen Konversation gleichen, sondern vielmehr die Form massenmedialer Kommunikation vonseiten der zentralen Accounts einerseits und daran anschließender Kommentierung durch das Publikum andererseits annehmen.⁷⁵ Ähnlich wie dies auch schon *Klemm* und *Michel* für das fernsehbegleitende Twittern festgestellt haben, handelt es sich bei diesen Kommentierungen meist um „freistehende Äußerungen“⁷⁶, die nicht von anderen aufgegriffen wurden, so dass also auch mit Blick auf die zentrale Konversation im vorliegenden Graphen nicht davon ausgegangen werden kann, dass hier elaborierte Deliberationsprozesse erfolgten. Deutlich wird dies anhand Abbildung 4. Diese zeigt einerseits das ego-zentrierte Netzwerk⁷⁷ für den Twitter-Account des Komikers *Bill Maher*. Dieser generierte zahlreiche Antworten (dargestellt durch die hellgrauen Pfeile in Richtung des Knotens), vollzog aber selbst keine Aktivitäten in Richtung einzelner anderer Knoten – dies wäre mit dunkelgrauen Pfeilen in deren Richtung dargestellt. Außerdem fanden auch keinerlei Aktivitäten zwischen den übrigen Knoten statt, sondern alle bezogen sich auf *Bill Maher*, so dass diese eher die Form einer Kommentierung der ungerichteten Äußerungen (in Form von nicht spezifisch adressierten Tweets) des zentralen Knotens annahmen.

Will man etwas über die Rolle von Socialbots in den hier identifizierten Online-Konversationen herausfinden, müssen diese zunächst identifiziert werden. Vor dem Hintergrund der in Abschnitt 2.2. beschriebenen manuellen Kategorisierung von Accounts, die sich im Graphen befinden, ist davon auszugehen, dass insgesamt unter den Beteiligten an Konversationen knapp 2.500 (circa 9,4 Prozent) Socialbots sind. Ganz anders als dies *Alessandro Bessi* und *Emilio Ferrara* sowie *Bence Kollanyi* u.a. aufgrund ihrer Hashtaganalyse darstellen, scheinen diese Bots aber keineswegs massiv zugunsten *Donald Trumps* zu agitieren. Zwar finden sich entsprechende automatisierte Konten, aber in der Gesamtheit bestätigt sich dieses Bild nicht. Stattdessen zeigt eine qualitative Inhaltsanalyse von Tweets der hier manuell als Socialbots identifizierten Accounts (vgl. Abschnitt 2.2.), dass diese zwar mehrheitlich *Donald Trump* thematisieren, dies allerdings in hohem Maße negativ. Zumindest mit Blick auf das untersuchte Sample kann also nicht bestätigt werden, dass

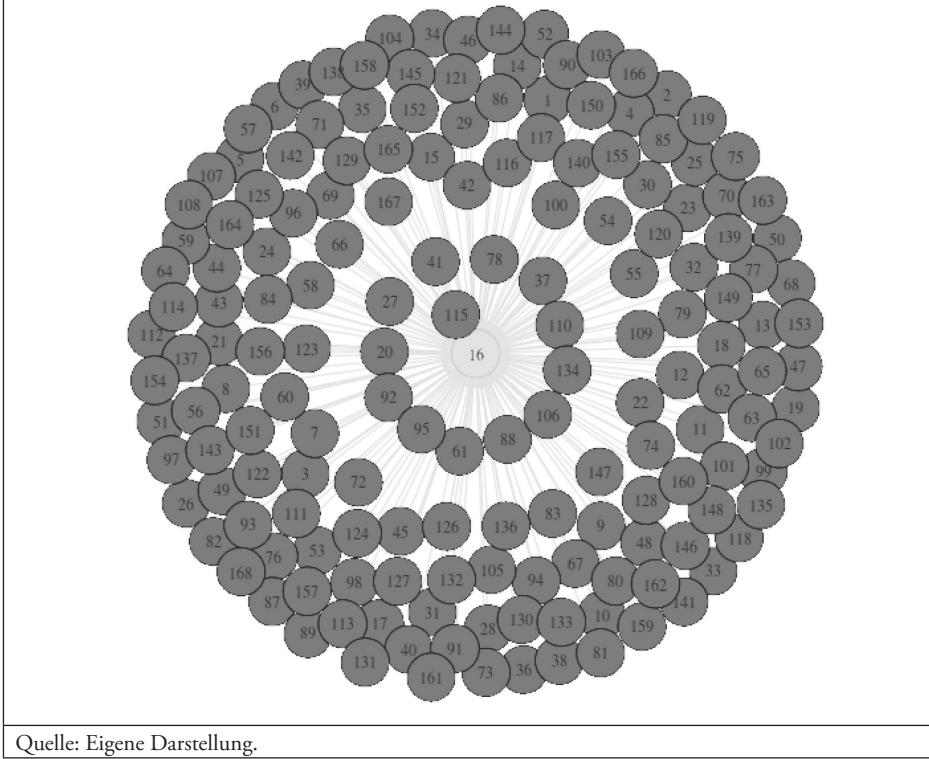
74 Wir haben dies für die Top 20 untersucht. Unter ihnen finden sich nur entsprechende Accounts.

75 Dies beschreiben ganz ähnlich *Luca Rossi / Matteo Magnani*, a.a.O. (Fn. 47), S. 565.

76 *Michael Klemm / Sascha Michel*, a.a.O. (Fn. 72), S. 18.

77 Vgl. hierzu *Jan Fuhse*, a.a.O. (Fn. 63), S. 117 ff.

Abbildung 4: Ego-zentriertes Netzwerk des Twitter-Accounts von Bill Maher



Quelle: Eigene Darstellung.

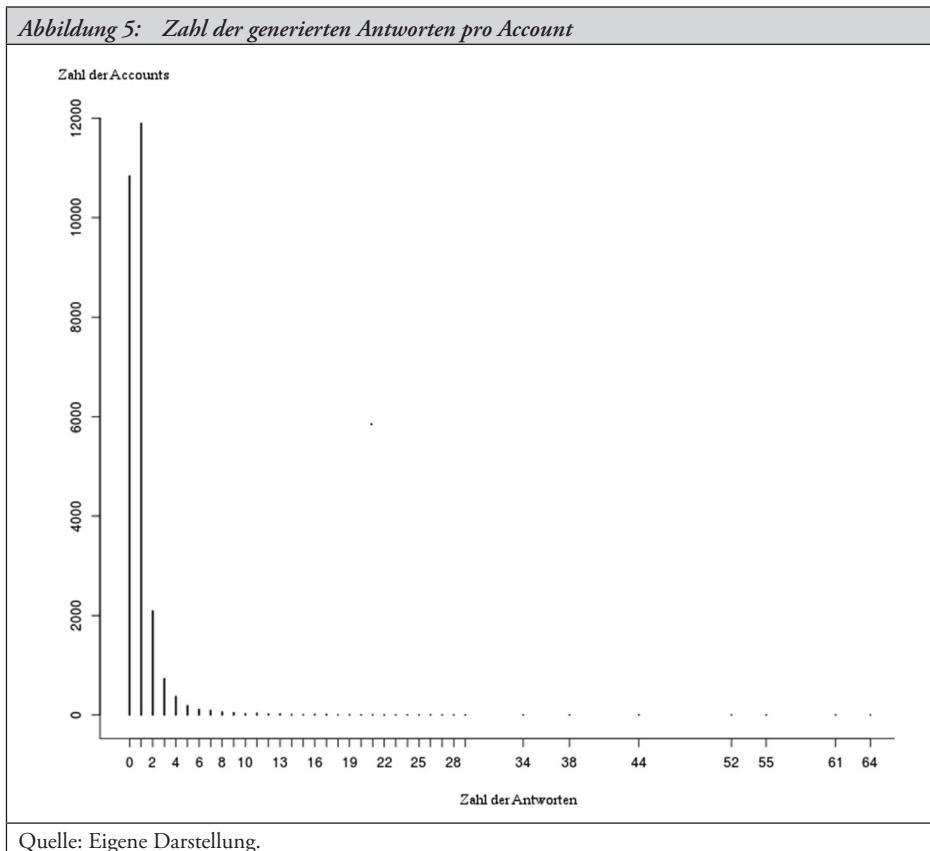
Socialbots im US-Präsidentenwahlkampf primär zur Unterstützung Trumps eingesetzt wurden.⁷⁸

Eine entsprechende Inhaltsanalyse der Tweetaktivitäten allein sagt aber noch nicht viel darüber aus, wie automatisierte Accounts konkret in Konversationen eingebunden sind (vgl. Abschnitt 2.1.). Hierzu bedarf es eingehenderer Analysen, die im Rahmen unseres Projektes in weiten Teilen noch in Arbeit sind.⁷⁹ Herauszufinden konnten wir bisher aber, dass Socialbots zu den aktivsten Accounts innerhalb des untersuchten Graphen zählen. Wie Abbildung 5 zeigt, sendete die Mehrzahl der Konten, die an Konversationen beteiligt waren, gar keine oder nur sehr wenige Antworten. Demgegenüber waren einige wenige Accounts in hohem Maße aktiv, was die Vermutung nahe legt, dass es sich hierbei um Socialbots handeln könnte.

Tatsächlich bestätigt sich dies beim Blick auf die zehn aktivsten Konten. Denn ihre manuelle Überprüfung ergibt, dass es sich bei sechs von ihnen – darunter die beiden aktivsten

78 Wir planen, die Inhaltsanalyse von Tweets auf weitere Accounts auszuweiten, um zu überprüfen, ob dieses Ergebnis bei größerer Datenmenge bestehen bleibt.

79 Hierzu gehören Analysen zum Verhalten von Socialbots und menschlich gesteuerten Accounts mittels „Exponential Random Graph Models“ sowie Analysen zu zeitlichen Dynamiken im Netzwerk. Ebenso gehören hierzu Konversationsanalysen des kommunikativen Geschehens in einzelnen Komponenten (den jeweiligen Konversationen) des Gesamtgraphen.



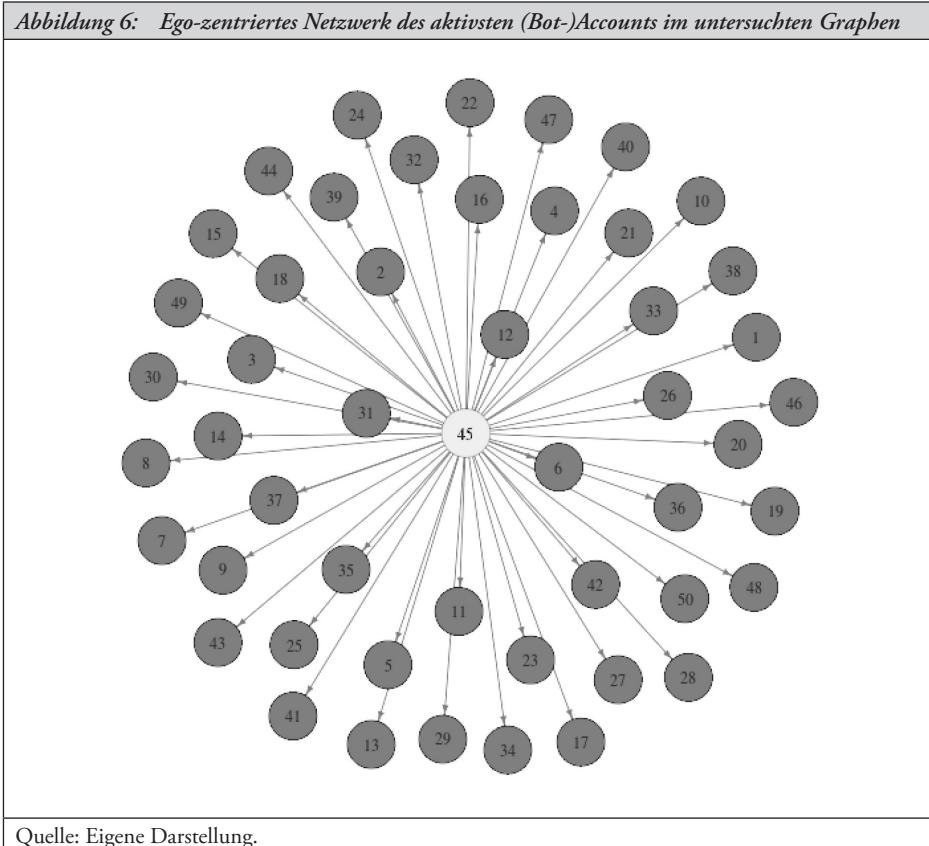
Accounts – um Socialbots handelt. Nur zwei – auf den Plätzen vier und sechs – lassen sich dagegen eindeutig als menschlich gesteuert identifizieren.⁸⁰ Sieht man sich genauer an, ob die Aktivitäten der Socialbots in irgendeiner Form Resonanz (im Sinne sichtbarer Reaktionen) erzeugen, fällt auf, dass dies nicht zutrifft. Vielmehr bleiben die Reply-Tweets ihrerseits unbeantwortet, was in Abbildung 6 deutlich wird.

Diese stellt das ego-zentrierte Netzwerk des zweitaktivsten (Bot-)Accounts im untersuchten Graphen dar und zeigt, dass trotz aller Aktivität seinerseits keinerlei an diesen Account gerichteten Aktivitäten registrierbar sind. Sämtliche Aktivitäten (als dunkelgraue Pfeile dargestellt) richten sich vom Knoten weg hin zu anderen Knoten, die ihrerseits keine Aktivitäten an ihn adressieren. Socialbots scheinen also trotz teilweise massiver Versuche kaum in der Lage, Reaktionen anderer User zu provozieren. Die hohe „Outdegree-Zentralität“⁸¹, die Socialbots aufgrund ihrer Aktivitäten besitzen, gepaart mit einer extrem

80 Die zwei übrigen Accounts wurden gelöscht beziehungsweise gesperrt, so dass eine Kontrolle nicht mehr möglich ist. Löschung und Sperrung legen aber nahe, dass es sich hierbei um automatisierte Accounts gehandelt haben könnte.

81 Zur Berechnung der Outdegree-Zentralität wird die Summe der von einem Knoten ausgehenden Beziehungen errechnet, vgl. Jan Fuhse, a.a.O. (Fn. 63), S. 59.

Abbildung 6: Ego-zentriertes Netzwerk des aktivsten (Bot-)Accounts im untersuchten Graphen



Quelle: Eigene Darstellung.

schwachen „Indegree-Zentralität“⁸² scheint damit in erster Linie ein Indikator für deren Hilflosigkeit⁸³ beziehungsweise Erfolglosigkeit zu sein, kommunikative Beziehungen zu etablieren.

Auch wenn noch weitere Analysen erforderlich sind, um das entstehende Bild der Rolle von Socialbots in Online-Konversationen zu vervollständigen, zeichnet sich somit bereits ab, dass sie – unabhängig von ihrer politischen Ausrichtung – kaum in der Lage zu sein scheinen, sich substantiell an den ohnehin nur rudimentär stattfindenden Online-Konversationen zu beteiligen oder gar manipulativ auf politische Meinungen oder das Wahlverhalten einzuwirken. Stattdessen scheinen selbst die Aktivitäten der aktivsten Socialbots vor allem eines zu bestärken: die Ignoranz der von ihnen adressierten (menschlichen) User.

82 Diese errechnet sich aus der Summe der eingehenden Beziehungen eines Knotens, vgl. ebenda, S. 59.

83 Vgl. hierzu auch ebenda, S. 60.

4. Socialbots als indirekte Gefahr für die Demokratie

Die bisherige Forschung zur Rolle von Socialbots im letzten Präsidentschaftswahlkampf in den USA hat dadurch internationale Aufmerksamkeit erlangt, dass sie einen massiven Einsatz von Socialbots in Online-Konversationen zugunsten Donald Trumps diagnostizierte und dies in einen Zusammenhang mit dessen Wahlsieg stellen konnte. Wie hier gezeigt wurde, sind diesem Narrativ gegenüber jedoch Zweifel anzumelden, da die Instrumente der Big Data-Analyse, mit denen bisherige Studien arbeiteten, nur eingeschränkt valide erscheinen.

Im Unterschied zur bestehenden Forschung liefern die in diesem Beitrag vorgestellten Untersuchungen einige Hinweise darauf, dass Socialbots zwar in Online-Konversationen deutlich präsent und auch hochgradig aktiv sind. Diese Konversationen stellen aber nur einen vergleichsweise kleinen Teil der gesamten Twitterkommunikation zur US-Präsidentenwahl dar und entwickeln kaum elaborierte Strukturen, so dass allein deshalb kaum anzunehmen ist, dass es Socialbots, die in entsprechenden kurzlebigen Konversationen aktiv sind, gelingt, öffentliche Meinungsbildungsprozesse im Internet zu beeinflussen. Dies bestätigt sich dadurch, dass Socialbots offensichtlich nicht in der Lage sind, Reaktionen anderer User hervorzurufen und ihre Aktivitäten stattdessen weitgehend ignoriert werden und unbeantwortet versanden.

Stellen Socialbots deswegen kein Problem für die Demokratie dar? Mitnichten! Allerdings besteht dieses Problem – anders als dies von der bisherigen Forschung sowie von politischer und medialer Öffentlichkeit befürchtet wird – wohl nicht in deren Fähigkeit zur direkten Beeinflussung von Online-Konversationen und Prozessen der Meinungsbildung im Internet. Viel sensibler als die Online-User scheinen Politik und Medien selbst auf das Online-Geschehen zu reagieren: Twitter wird von ihnen zunehmend als Medium behandelt, dessen Beobachtung Auskunft über den Stand politischer Meinungen und Gemütslagen der Bevölkerung geben kann, um hierüber zu berichten beziehungsweise hierauf politisch zu reagieren.⁸⁴ Wenn Politik und Medien sich aber zunehmend daran orientieren, was in den sozialen Medien im Allgemeinen sowie auf Twitter im Besonderen kommuniziert wird, besteht die Gefahr, dass Socialbots, die bestimmte Themen oder Hashtags verbreiten, zu Agenda-Settern werden und auf diese Weise schließlich auf indirektem Wege Einfluss auf politische Debatten und Entscheidungsprozesse nehmen.

84 Exemplarisch zeigt dies folgender Artikel eines ehemaligen Beraters von Barack Obama: Dan Pfeiffer, How Social Media is Revolutionizing Debates, in: CNN online vom 16. September 2015, <https://edition.cnn.com/2015/09/15/opinions/pfeiffer-social-media-debates/index.html> (Abruf am 8. August 2018).