

Supplement

Dynamische Netzwerkanalyse zum deutschsprachigen #MeToo auf Twitter

Miriam Siemon

Freie Universität Berlin, IfPuK

miriam.siemon@fu-berlin.de

Im Rahmen der #MeToo-Studie wird der deutschsprachige Diskurs auf Twitter langfristig mittels einer Dynamischen Netzwerkanalyse untersucht. Dieses Repositorium enthält die Videoansicht der monatlichen Entwicklung. Damit begegnen wir auch einer methodischen Kritik an der Netzwerkforschung, in der häufig Social-Media-Daten über einen längeren Zeitraum aggregiert werden. So zu vermeintlichen ‚Momentaufnahmen‘ verdichtet (Jackson & Welles, 2016, S. 397), können keine dynamischen Entwicklungen sichtbar gemacht werden, der eigentlich prozessuale Charakter der Netzkommunikation verschwindet in der Analyse. Dies führt dazu, dass Akteur*innen, die zu einzelnen Zeitpunkten zentral sind, in ihrer Bedeutung für den Gesamtzeitraum überschätzt werden (Trier, 2010, S. 215).

Datensatz

Unser Datensatz umfasst deutschsprachige Tweets mit dem Hashtag #MeToo von mehr als vier Jahren: Beginnend mit dem Zeitpunkt des Aufkommens des Hashtags am 17. Oktober 2017 bis zum 31. Dezember 2021. Insgesamt enthält der Datensatz 284.348 Tweets und Retweets (65%) von 87.051 einzelnen Nutzer*innen.

Der Datensatz teilt sich dabei in drei Segmente, die unterschiedliche Stärken und Limitationen aufweisen. Datensatz 1 umfasst den Zeitraum 17.10.17 - 04.12.18 und enthält 185.272 Tweets und Retweets. Diese Daten wurden uns freundlicherweise von Axel Bruns und dessen Team der Queensland University of Technology zur Verfügung gestellt. Da diese Daten zusammen mit englischen und anderssprachigen Tweets gestreamt wurden, wurde vermutlich das 1% Level aller aktuellen Tweets, das die Twitter API als Download-Grenze festlegt, erreicht und das Datensegment enthält folglich nicht alle in diesem Zeitraum geposteten Tweets. Hier ist jedoch von einer zufälligen und nicht von einer systematischen Verzerrung auszugehen, was weniger Einfluss auf die Netzwerkstruktur haben sollte. Datensatz 2 umfasst 47.548 Tweets und Retweets des Zeitraums 04.12.18 - 10.5.20. Für diesen Zeitraum wurde mittels rtweet (Kearney, 2018) und der Twitter-Premium API auf das Twitter-Archiv zugegriffen. Hier lässt sich eine systematische Verzerrung befürchten, da Tweets von gelöschten Accounts nicht rückwirkend heruntergeladen werden können. Datensatz 3 umfasst 51.529 Tweets und Retweets des Zeitraums 10.5.20-31.12.21. In diesem Zeitraum wurden alle deutschsprachigen #MeToo Tweets wöchentlich mittels rtweet heruntergeladen. Auch wenn Twitter mittlerweile den Zugriff für Wissenschaftler*innen auf historische Daten deutlich ausgeweitet hat, halten wir es trotz der angegebenen Limitationen der einzelnen Datensätze für sinnvoller mit diesen Daten zu arbeiten, als zum heutigen Zeitpunkt rückwirkend alle deutschsprachigen #MeToo Tweets herunterzuladen. Eine Stichprobe mit $n = 1.000$ Tweets ergab, dass nur noch ca. die Hälfte der Tweets weiterhin verfügbar sind, d.h. dass entweder die Tweets oder die Accounts, von denen die Tweets stammen, gelöscht wurden. Dies würde vermutlich eine sehr viel größere systematische Verzerrung bedeuten, da Twitter rechtsextreme und -populistische Accounts, die gegen ihre Policy Richtlinien verstoßen, zunehmend sperrt (Van Dijck et al., 2021, S. 2–3).

Aus dem Gesamtdatensatz wurden Node- und Edgelisten für die Netzwerkanalyse erstellt, mit Twitter-Accounts als Knoten und deren kommunikativen Interaktionen in Form von Retweets, Quotes, Replies und Mentions als Kanten. Das Gesamtnetzwerk umfasst 91.330 Knoten und 343.013 Kanten wobei v.a. Retweets (54%) und Mentions (32%) häufig genutzte Interaktionsformen darstellen. Replies (8%) und Quotes (6%) kommen weniger häufig vor. Für Quotes gilt die Einschränkung das, zumindest für den ersten Teil des Datensatzes (17.10.17-4.12.18), einige Quote-Verbindungen (22%) nicht in die Edgelist einfließen können, da hier Informationen über bereits gelöschte kommentierte Tweets fehlten.

Modellierung des dynamischen Netzwerks

Für dynamische Netzwerkanalysen gibt es zwei Vorgehensweisen. Einerseits kann das Netzwerkwachstum im Zeitverlauf beobachtet werden. Hierfür kommen nach und nach weitere Knoten und Kanten hinzu bis das Netzwerk zum finalen Zeitpunkt vollständig ist. In diesem Fall haben die Kanten einen Zeitstempel, der die Grenze (den Eintrittszeitpunkt) festlegt. Aus einer praxistheoretischen Perspektive interessiert uns jedoch nicht das reine Netzwerkwachstum, sondern vielmehr die Kontinuität, d.h. das Entstehen wie auch Verschwinden von Praktiken (hier kommunikativen Interaktionen), die Auswirkungen auf die Verstetigung oder Veränderung der Netzwerkstruktur im Zeitverlauf haben. Daher ist es für uns sinnvoll, Kanten mit zwei Grenzen (Ein- und Austrittszeitpunkt) zu berücksichtigen. In diesem Fall ist jeder Kante ein Zeitintervall zugeordnet. Dieses Zeitintervall konnten wir induktiv aus unseren Daten heraus bestimmen. Dafür wurden die zeitlichen Abstände zwischen einem Tweet und der Reaktion darauf (Retweet, Quote, Reply) berechnet und das dritte Quantil bestimmt. Dies liegt bei 18 Stunden. Innerhalb dieser 18 Stunden sind 75% aller Reaktionen auf einen Tweet in unserem Datensatz entstanden. Dieser Wert wurde anstelle des Mittelwertes (geprägt durch starke Ausreißer) oder Medians gewählt, um nicht professionell betreute und damit vermutlich schneller arbeitende Accounts gegenüber privat oder ehrenamtlich geführten zivilgesellschaftlichen Accounts zu bevorzugen. Auch hier gibt es methodische Limitationen, da für einzelne Interaktionsformen von Datensatz1 keine Informationen über den zeitlichen Abstand zwischen Originaltweet und Reaktion gewonnen werden konnte. Die 18 Stunden spiegeln daher tendenziell eher die Reaktionszeit des späteren Debattenverlaufs wieder. Außerdem sind Mention-Beziehungen hier nicht repräsentiert, da diese Akteur*innen adressieren und keine Reaktion auf einen bestimmten Tweet darstellen. Allen vier Beziehungstypen wurde daher das Zeitintervall von 18 Stunden beginnend mit dem Eintrittszeitpunkt zugeordnet.

Visualisierung des dynamischen Netzwerks

Die dynamische Visualisierung des Netzwerks erfolgte mit der open source software *Gephi* (Bastian et al., 2009). Für die Analyse wurde mit einem reduzierten Netzwerk (Degree ≥ 10) gearbeitet, das eine dynamische Darstellung im Videoformat erlaubte. Teil des reduzierten Netzwerks sind alle Knoten (und entsprechende Kanten), die im gesamten Zeitraum mindestens 10 kommunikative Interaktionen egal welchen Beziehungstyps aufweisen. Das reduzierte Netzwerk umfasst 9.784 Knoten (ca. 10% der Gesamtknoten) und 182.403 Kanten. Die Netzwerkdichte liegt bei 0,2% als Anteil aller realisierten im Verhältnis zu möglichen Kanten und spiegelt die üblicherweise geringe Dichte in Social Media Diskursnetzwerken wieder (González-Bailón & Wang, 2016, S. 95). Sowohl die Netzwerkstruktur als auch das Verhältnis der verschiedenen Beziehungstypen (Retweets = 53%, Mentions = 35%, Quotes = 6%, Replies = 6 %) des reduzierten Netzwerks gleichen dem Gesamtnetzwerk. Für die Identifikation von eng vernetzten Clustern im Netzwerk wurde der Multi-Level Community Detection Algorithmus nach Blondel et al. (2008) angewandt. Dieser zielt auf die Optimierung der Netzwerk-Modularität ab und eignet sich daher besonders für die Analyse großer Netzwerke (Stoltenberg et al., 2019, S. 130).

Als Layout-Algorithmus zur Netzwerkvisualisierung wurde Force Atlas2 verwendet. Dieser arbeitet fortlaufend an der Optimierung des Layouts, was eine Voraussetzung bei der Darstellung dynamischer Netzwerkentwicklungen ist (Levallois, 2013). Die Kantenfarbe im visualisierten Netzwerk entspricht dem Beziehungstyp (lila = Mention, grün = Retweet, orange = Quote, blau = Reply), die Knotengröße dem Indegree. Das dynamische Netzwerk ist in monatlichen Abständen (je 30,5 Tage) visualisiert.

Referenzen

- Bastian, M., Heymann, S., & Jacomy, M. (2009). Gephi: An Open Source Software for Exploring and Manipulating Networks. *Proceedings of the Third International ICWSM Conference*, 361–362.
- Blondel, V. D., Guillaume, J.-L., Lambiotte, R., & Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, P10008.
<https://doi.org/10.1088/1742-5468/2008/10/P10008>
- González-Bailón, S., & Wang, N. (2016). Networked Discontent: The Anatomy of Protest Campaigns in Social Media. *Social Networks*, 44, 95–104. <https://doi.org/10.1016/j.socnet.2015.07.003>
- Jackson, S. J., & Welles, B. F. (2016). #Ferguson is everywhere: Initiators in emerging counterpublic networks. *Information, Communication & Society*, 19(3), 397–418.
<https://doi.org/10.1080/1369118X.2015.1106571>
- Kearney, M. W. (2018). *Rtweet: Collecting Twitter Data. R package version 0.6.7*. <https://cran.r-project.org/package=rtweet>.
- Levallois, C. (2013). *Gephi. A Tutorial – on dynamic networks*.
https://edisciplinas.usp.br/pluginfile.php/2671054/mod_resource/content/1/gephi_tutorial_dynamics.pdf
- Stoltenberg, D., Maier, D., & Waldherr, A. (2019). Community Detection in Civil Society Online Networks: Theoretical Guide and Empirical Assessment. *Social Networks*, 59, 120–133.
<https://doi.org/10.1016/j.socnet.2019.07.001>
- Trier, M. (2010). Struktur und Dynamik in der Netzwerkanalyse. In C. Stegbauer & R. Häußling (Hrsg.), *Handbuch Netzwerkforschung* (S. 205–217). VS Verlag für Sozialwissenschaften.
https://doi.org/10.1007/978-3-531-92575-2_18

Van Dijck, J., de Winkel, T., & Schäfer, M. T. (2021). Deplatformization and the governance of the platform ecosystem. *New Media & Society*, 00(00), 14614448211045662.

<https://doi.org/10.1177/14614448211045662>