

#wenigerHashtagswirkenwissenschaftlicher

Der Zusammenhang von Tweet-Eigenschaften und wahrgenommener Wissenschaftlichkeit

Athanasios Mazarakis

ZBW – Leibniz-Informationszentrum
Wirtschaft und Christian-Albrechts-
Universität zu Kiel
a.mazarakis@zbw.eu

Isabella Peters

ZBW – Leibniz-Informationszentrum
Wirtschaft und Christian-Albrechts-
Universität zu Kiel
i.peters@zbw.eu

Jasmin Schmitz

ZB MED –
Informationszentrum
Lebenswissenschaften
schmitz@zbmed.de

Abstract

Der Beitrag berichtet die Ergebnisse einer Proof-of-Concept-Studie zur wahrgenommenen Wissenschaftlichkeit von Tweets. Im Rahmen eines online durchgeführten Experiments wurden Proband_innen in einem Fragebogen manuell thematisch-kategorisierte Tweets aus einer Tweet-Sammlung zur Bewertung vorgelegt, die aus medizinischen Hashtags generiert wurden. Es wurde ermittelt, welche Eigenschaften und Themen von Tweets einen Einfluss auf ihre wahrgenommene Wissenschaftlichkeit haben. Die Ergebnisse beruhen auf einer Analyse von 162 klassifizierten Tweets in deutscher und englischer Sprache und ihrer Bewertung durch 109 deutschsprachige Proband_innen. Die Ergebnisse zeigen unter anderem, dass Tweets mit vielen Hashtags eher als „nicht wissenschaftlich“ wahrgenommen werden. Eindeutig zugeordnet wurden von Proband_innen solche Tweets, die sich mit den Kategorien „Methodik“, „Job“ und „Spaß“ beschäftigen, wobei nur zur ersteren Kategorie zugehörige Tweets überwiegend als „wissenschaftlich“ bezeichnet wurden.

Keywords: wahrgenommene Wissenschaftlichkeit; Social Media; Twitter; Informationsverhalten

1 Einleitung und Motivation

Soziale Medien, wie Twitter oder Facebook, werden mittlerweile sehr häufig in der Wissenschaft eingesetzt, um aktuelle Forschungsergebnisse zu kommunizieren (McKendrick/Cumming/Lee, 2012) oder auch mit verschiedenen Stakeholdern zu diskutieren (Salem/Borgman/Murphy, 2016; Sugimoto et al., 2017). Allerdings besteht aufseiten der Forschenden teilweise noch immer die Befürchtung, dass Wissenschaftler_innen, welche soziale Medien aktiv für ihre Arbeit nutzen, als unprofessionell angesehen werden (Lemke et al., 2019; Osterrieder, 2013). Trotzdem twittern viele Forschende, insbesondere während wissenschaftlicher Konferenzen (Letierce et al., 2010). Die Inhalte der Tweets sind auch persönlicher Natur, wie zum Beispiel Äußerungen zu aktuellen politischen Geschehnissen oder Nachrichten (McQuate, 2017). Auch Haustein et al. (2014) konnten zeigen, dass Astrophysiker_innen größtenteils nicht die gleichen Terme in Tweets nutzen wie in Abstracts zu publizierten Artikeln. Ob dies der anderen antizipierten Nutzerschaft von Twitter geschuldet ist oder der Trennung von wissenschaftlicher Kommunikation und eher informeller Privatkommunikation, ist offen geblieben.

Auf der anderen Seite zeigen Studien (u.a. Hölzig/Hasebrink, 2020), dass sich immer mehr Menschen auf Social-Media-Plattformen über das aktuelle Tagesgeschehen und z.B. Gesundheitsinformationen (Islam et al., 2020) informieren. Die Umfrage von Hölzig und Hasebrink (2020) kam auch zu dem Ergebnis, dass etwas mehr als ein Drittel der Befragten in Sorge sind, dass sie Fakten nicht von Falschinformationen unterscheiden können. Gleichzeitig sind Wissenschaftler_innen, Ärzt_innen und Gesundheitsorganisationen die am zweithäufigsten genutzten Nachrichten- und Informationsquellen über das Coronavirus SARS-CoV-2 nach „Nachrichtenorganisationen“.

Produzent_innen und Rezipient_innen von wissenschaftlichen Social-Media-Informationen haben dementsprechend gleichermaßen ein Interesse daran, eine gewisse „Wissenschaftlichkeit“ zu demonstrieren und als „wissenschaftlich“ wahrgenommen zu werden bzw. wahrhaftige „Wissenschaftlichkeit“ zu erkennen und den Konsum von Falschinformationen zu vermeiden. Doch wie kann diese „Wissenschaftlichkeit“ in Social-Media-Postings nachgewiesen werden – insbesondere im Hinblick darauf, dass nicht alle Rezipient_innen Fachexpert_innen sein können und damit in der Lage sind, die inhaltliche Qualität eines Beitrags zu beurteilen? Im Kontext von Tweets und Twitter haben Weller, Dröge und Puschmann (2011) folgende Signale

von Wissenschaftlichkeit diskutiert: wissenschaftlicher Inhalt oder Link zu einem wissenschaftlichen Beitrag (z.B. DOI), Tweet mit einem wissenschaftlichen Hashtag (z.B. Konferenz-Hashtag) oder wissenschaftliche_r Autor_in als Autor_in des Tweets. Obwohl Studien zu Konferenz-Hashtags populär sind (z.B. Letierce et al., 2010; Lemke/Mazarakis/Peters, 2016; Mazarakis/Peters, 2015), ist das Bestimmen von wissenschaftlichen Hashtags und Autor_innen der Flaschenhals dieser Ansätze. In etablierten Formen wissenschaftlicher Kommunikation, z.B. in Zeitschriftenartikeln, Konferenzpublikationen oder Fachbüchern, signalisieren weitere, leichter erhältliche Metadaten die Wissenschaftlichkeit eines Forschungsoutputs, z.B. Veröffentlichung bei einem bekannten Verlag, Impact Factor der Zeitschrift, durchgeführtes Peer Review. Auch die optische Erscheinung („Layout-Effekt“) des Beitrags trägt dazu bei, dass Texte als „wissenschaftlich“ wahrgenommen werden (Hahn et al., 2020).

Diese explorative Studie geht u.a. der Frage nach, ob sich ein derartiger Layout-Effekt auch für Tweets, d.h. Beiträge auf Twitter, nachweisen lässt. Denn auf Social-Media-Plattformen wie Twitter gibt es (noch) keine Konventionen oder gar Templates, die eine „Wissenschaftlichkeit“ optisch unterstreichen können oder in 280 Zeichen verdeutlichen, dass es sich um einen wissenschaftlichen Tweet handelt. Es fehlen zudem die Metadaten der etablierten wissenschaftlichen Publikationsformen, was die Beurteilung der Tweets schwierig macht. Zwei Fragestellungen wurden im Rahmen der Studie bearbeitet:

1. Wann wird ein Tweet als wissenschaftlich eingestuft?
2. Welche Eigenschaften bei Tweets wirken sich auf die wahrgenommene Wissenschaftlichkeit aus?

Für die Studie wurden typische Eigenschaften von Tweets, die sich auch optisch gut wahrnehmen lassen (z.B. Vorhandensein von Hashtags, Retweets, Gefällt-mir-Angaben; Bruns/Stieglitz, 2013), in einem online durchgeführten Experiment untersucht, um Hinweise darauf zu erhalten, welche Eigenschaften von Tweets einen Einfluss auf ihre wahrgenommene „Wissenschaftlichkeit“ haben. In die Auswertung der Online-Befragung wurden auch die Tweet-Inhalte eingebracht, um den Zusammenhang zwischen Thema, Tweet-Eigenschaften und „Wissenschaftlichkeit“ zu beleuchten. Für die Studie wurden Themenfelder aus der Medizin ausgewählt, da diese sowohl von Forschenden, aber auch von Patient_innen, interessierten Laien, Politiker_innen etc. aufgegriffen werden, sodass ein Datenpool zu erwarten ist, anhand dessen sich konkrete Unterschiede zwischen „Wissenschaftlichkeit“

und „Nicht-Wissenschaftlichkeit“ herausarbeiten lassen. In diesem Beitrag wird eine Proof-of-Concept-Studie präsentiert, welche die Grundlage für weiterführende Arbeiten z.B. im Bereich des Informationsverhaltens oder der Informationskompetenz bilden kann. Wir betrachten unsere Arbeiten zur wahrgenommenen „Wissenschaftlichkeit“ (Hahn et al., 2020) von Social-Media-Postings als Work-in-Progress.

2 Methoden und Daten

Im Folgenden werden die methodischen Ansätze zur Datenerhebung und -analyse sowie die erhobenen Daten beschrieben.

2.1 Datenbasis für die Online-Befragung

Die Datenbasis ist eine Tweet-Sammlung, welche zwischen April 2018 und Oktober 2018 erhoben und im Februar 2019 final bearbeitet wurde. Dafür wurden einschlägige medizinische Fachbezeichnungen in Deutsch und Englisch als Suchworte (d.h. Hashtags) auf Twitter verwendet und die gefundenen Tweets heruntergeladen. Die Hashtags entstammen einer explorativen Ad-hoc-Suchliste, bestehend aus medizinischen Hashtags. In Tabelle 1 sind die Hashtags verzeichnet, welche für die Suche verwendet wurden.

Tab. 1: Für die Suche verwendete Hashtags

Englisch	Deutsch
#ophthalmology	#augenheilkunde
#radiology	#radiologie
#midwifery	#geburtshilfe
#dermatology	#dermatologie
#orthopedics	#orthopädie
#microbiology	#mikrobiologie
#urology	#urologie
#epidemiology	#epidemiologie
#andrology	#andrologie
#oncology	#onkologie

Für die Auswahl der zu analysierenden Tweets wurden auch Antworten aufgenommen, welche auf einen vorherigen Tweet mit dem fachlich einschlägigen Hashtag geantwortet haben, selbst aber keinen Hashtag nachweisen konnten. Hierbei wird dennoch eine thematische Nähe zum Ursprungstweet angenommen.

Insgesamt wurden zum Erhebungszeitpunkt 186 Tweets ausgewählt, von denen final 162 Tweets in die anschließende Online-Befragung eingeflossen sind. Davon sind 22 Tweets in deutscher Sprache und 140 Tweets in englischer Sprache. Zum Erhebungszeitpunkt wurde der Tweet-Korpus in deutscher Sprache für die gesuchten Hashtags aus Tabelle 1 nahezu erschöpfend erfasst, da die Tweets nur in sehr geringer Anzahl auf Twitter vorhanden waren. Daher wurden zusätzlich überdurchschnittlich viele englischsprachige Tweets erhoben. Dies geschah, um ggf. relevante Sprachunterschiede zwischen den Tweets aufzuzeigen. Bei den aus dem Sample ausgeschlossenen Tweets handelt es sich primär um Antworten auf Tweets, die ohne die gleichzeitige Darstellung des Original-Tweets nicht verstanden werden. Auch Tweets, welche fachspezifische Abkürzungen und Fachtermini enthalten, sodass sie für Außenstehende nicht verständlich sind, wurden aus dem Sample entfernt.

Die Tweets wurden manuell beschrieben, um später den Einfluss spezifischer Eigenschaften bei der Einschätzung von „Wissenschaftlichkeit“ statistisch auswerten zu können. Zwei Kodierer haben dazu drei Arbeitsschritte vollzogen (s. Abb. 1). Zunächst wurden die 162 Tweets danach überprüft, ob in ihnen eine Frage gestellt wurde oder ob sie entweder ein Bild, ein Video oder einen Link enthalten haben. Es wurden außerdem die Anzahl der Retweets, der „Gefällt mir“-Angaben, der Hashtags und der @-mentions pro Tweet erfasst, welche gängige Metriken bei der Analyse von Tweets sind (Bruns/Stieglitz, 2013). Im Rahmen der Proof-of-Concept-Studie sollte untersucht werden, inwieweit diese Metriken im Hinblick auf die Bewertung von „Wissenschaftlichkeit“ eine Rolle spielen. Ein Tweet kann gleichzeitig mehrere Eigenschaften aufweisen, zum Beispiel kann er ein Bild und einen Link enthalten. Solche Tweets wurden dann in allen entsprechenden Eigenschaften annotiert, im gerade erwähnten Beispiel also sowohl in der Eigenschaft „Bild“ als auch in der Eigenschaft „Link“. Die Ergebnisse der Beschreibung sind in den Tabellen 2 und 3 dargestellt. Dreiviertel der Tweets mit den gesuchten Hashtags enthielten ein Bild, etwas weniger auch einen Link. Videos kamen praktisch nicht vor. Im Durchschnitt wurden die Tweets knapp zwei-

mal mit einer „Gefällt mir“-Angabe versehen und enthielten circa fünf Hash-tags.

Tab. 2: Vorhandensein bestimmter Eigenschaften
im analysierten Tweet-Sample, $n = 162$

	Ja	Nein
Tweet ist als Frage formuliert?	17 (10%)	145 (90%)
Tweet enthält ein Bild?	122 (75%)	40 (25%)
Tweet enthält ein Video?	4 (2%)	158 (98%)
Tweet enthält einen Link?	113 (70%)	49 (30%)

Tab. 3: Häufigkeit des Vorkommens weiterer Eigenschaften
in dem analysierten Tweet-Sample, $n = 162$

	Mittelw.	St.Abw.	Median	Min.	Max.
Anzahl der Retweets pro Tweet	0.88	2.96	0	0	27
Anzahl der „Gefällt mir“-Angaben pro Tweet	1.96	5.41	0	0	54
Anzahl der Hashtags pro Tweet	5.48	4.54	4	0	28
Anzahl der @-Erwähnungen pro Tweet	0.59	1.06	0	0	7

Um weitere Hinweise darauf zu erhalten, welche Tweet-Typen wissenschaftlich wirken, haben im zweiten Schritt die beiden Kodierer die Tweets im Rahmen einer Inhaltsanalyse verschiedenen Kategorien zugeordnet. Jeder Tweet wurde genau einer von sieben Kategorien zugeordnet (Kategorien und Beschreibungen s. u.). In acht Zweifelsfällen haben sich die Kodierer beraten und gemeinsam eine Zuweisung vorgenommen. Insgesamt wurde eine Übereinstimmung von ca. 95% erreicht (Cohens Kappa bei 0.91), was eine sehr hohe Übereinstimmung widerspiegelt.

Das Kodieren der Tweets resultierte in den folgenden sieben Kategorien:

- **Artikel** ($n = 35$) sind Tweets mit Informationen darüber, ob fachliche Artikel auf einem Fachportal oder in einer Fachzeitschrift veröffentlicht wurden, auf einer Universitätsseite, einem Blog, als Podcast usw., oder ob es sich um eine (wissenschaftliche) Nachricht auf einer Nachrichtenseite handelt. Die Studienergebnisse zeigen, dass Tweets, welche zu Artikeln auf wissenschaftlichen Nachrichtenseiten oder auf Fachportale leiten, eher als „wissenschaftlich“ betrachtet werden als Tweets, die auf Blogs, Podcasts oder Q&A-Seiten (Sonstiges) verlinken;

- **Job** ($n = 20$), bestehend aus Jobangeboten o.Ä.;
- **Meinung** ($n = 8$), was als Behauptung ohne Quellen zu verstehen ist;
- **Methodik** ($n = 18$), als Darstellung wissenschaftlicher Methodik sind Tweets kategorisiert, welche im Rahmen der wissenschaftlichen Abläufe entstanden sind, mit einer klaren Fokussierung auf wissenschaftliche Publikationen und Vorträge;
- **Publicity** ($n = 59$), bezüglich z.B. Buchvorstellungen, die Suche nach Personal, Teamvorstellungen, Werbung für einen Kongress, Kurse oder Organisationen. Die Ergebnisse zeigen, dass Tweets, die über Kongresse berichten, z.B. verstärkt als wissenschaftlich kategorisiert werden, auch wenn die meisten Tweets inhaltlich nur auf Kongresse aufmerksam machen und deshalb der Kategorie „Publicity“ zugeordnet wurden;
- **Spaß** ($n = 5$), wie z.B. sogenannte Funfacts;
- **Verständnis** ($n = 17$), mit einfachen Ergebnisdarstellungen oder allgemeinen Fragen.

Der überwiegende Teil der Tweets in dem Sample beschäftigte sich mit „Publicity“ ($n=59, 36\%$), zu denen auch Tweets zu Konferenzen gehören, gefolgt von Tweets über wissenschaftliche Artikel ($n=35, 22\%$). Die Kategorien sind nicht überlappend, jeder Tweet wurde einer Kategorie eindeutig zugeordnet. Beispiele finden sich in den Abbildungen 3, 4 und 5.

Im dritten Schritt wurden die Tweets durch die Kodierer bezüglich der wahrgenommenen „Wissenschaftlichkeit“ bewertet. Dies geschah, um eine Grundannahme des Experiments – nämlich, dass „wissenschaftliche“ und „nicht-wissenschaftliche“ Tweets von Laien und Expert_innen (d.h. Kodierern) ähnlich wahrgenommen werden – überprüfen zu können. Wie auch im tatsächlichen Experiment wurde keine Definition von „Wissenschaftlichkeit“ vorgegeben und nur die Frage gestellt: „Ist der Tweet aus Ihrer Sicht wissenschaftlich?“. Insgesamt wurde eine Übereinstimmung von ca. 94% zwischen den Kodierern erreicht, Cohens Kappa liegt bei 0.88.

Die Bewertung durch die Kodierer resultierte in folgender Verteilungshäufigkeit der Tweets im Sample: Ca. ein Viertel der Tweets wurden als „wissenschaftlich“ eingestuft und ca. drei Viertel als „nicht-wissenschaftlich“. Damit war die nötige Varianz in der wahrgenommenen „Wissenschaftlichkeit“ von Tweets für die nachfolgenden statistischen Auswertungen gegeben. Abbildung 1 gibt einen Überblick zur Sammlung und Kategorisierung des Tweet-Samples der Online-Befragung.

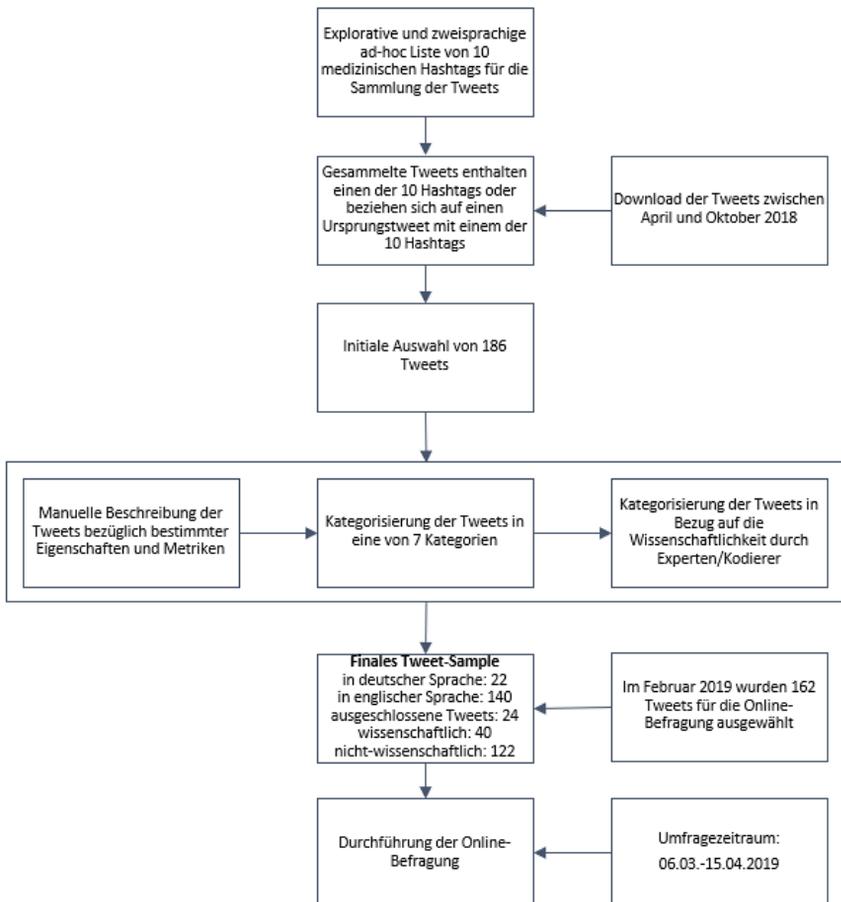


Abb. 1 Überblick zur Sammlung und Kategorisierung des Tweet-Samples der Online-Befragung

2.2 Online-Befragung

Die Online-Befragung wurde zwischen dem 6. März 2019 und dem 15. April 2019 durchgeführt, nachdem Teilnahmeaufrufe auf den Institutsseiten und auf Twitter und Facebook veröffentlicht wurden. Dazu wurde den Proband_innen jeweils eins von acht Sets mit je 20 unterschiedlichen Tweets aus dem Sample mit 162 Tweets angezeigt. Die Sets sollten dann von den Proband_innen bewertet werden. Die Proband_innen wurden gefragt, ob der angezeigte Tweet ihrer Meinung nach „wissenschaftlich“ ist (Antwortmög-

lichkeit: „Ja“ oder „Nein“). Im Anschluss an die Befragung erhielten die Proband_innen die Möglichkeit, ihre Votings noch einmal zu überprüfen. Mehrfachteilnahmen waren möglich, da die vorgegebenen Sets unterschiedliche Tweets beinhalteten und die Sets randomisiert verteilt wurden. Wir haben in unserer Studie keine Hinweise darauf, dass Mehrfachteilnahmen vorgekommen sind. In der Studie wurde mit Sets gearbeitet, damit eine möglichst gleichverteilte Anzeige aller Tweets stattfindet. Bei einer vollständig randomisierten Tweet-Anzeige und unter Beibehaltung der Abfrage von 20 unterschiedlichen Tweets je Proband_in wäre sonst eine mindestens vierstellige Proband_innenanzahl notwendig gewesen, um valide Ergebnisse zu erhalten. Die Proband_innen haben immer 20 Tweets bewertet, um den Zeitaufwand und die Abbruchquote gering zu halten.

Insgesamt haben 109 deutschsprachige Proband_innen an der Studie teilgenommen. Das durchschnittliche Alter der Proband_innen ist 43 Jahre, der Median liegt bei 41 Jahren, die Altersspanne reicht von 25 bis 65 Jahren. Der größte Teil der befragten Personen ist weiblich ($n = 47$, 43%), 32 Personen sind männlich (29%) und 30 (28%) Personen haben keine Angaben zum Geschlecht gemacht.

3 Ergebnisse

Zuerst werden die Ergebnisse zum Zusammenhang von Tweet-Eigenschaften und wahrgenommener „Wissenschaftlichkeit“ präsentiert, bevor auf die Tweets eingegangen wird, die als eindeutig „wissenschaftlich“ und eindeutig „nicht-wissenschaftlich“ kategorisiert wurden.

3.1 Tweet-Eigenschaften und wahrgenommene „Wissenschaftlichkeit“

Die Ergebnisse wurden über Visualisierungen wie Boxplots und deskriptive Tabellen ausgewertet. Außerdem wurden mehrere Spearman-Korrelationen durchgeführt, da die Ergebnisse nicht normalverteilt sind und durch Aggregation auf Ordinalskalenniveau vorliegen, womit die Voraussetzungen für die Durchführung von Spearman-Korrelationen erfüllt sind.

Insgesamt wurden im Durchschnitt die Tweets zu 41% (Akzeptanzwert = 0.41) als „wissenschaftlich“ von den Proband_innen eingestuft (Median = 0.40 und Standardabweichung = 0.29). Dies entspricht dem aggregierten Durchschnitt aller Bewertungen bezüglich der wahrgenommenen „Wissenschaftlichkeit“ je Tweet. Wurde also beispielsweise ein Tweet von 20 Proband_innen bewertet, wobei jedoch nur sechs Proband_innen den Tweet als „wissenschaftlich“ wahrgenommen haben, dann hat dieser Tweet einen Akzeptanzwert von 0.30, bzw. von 30%.

Um die Grundannahmen des Experiments mit den Ergebnissen zu vergleichen, wurde die wahrgenommene „Wissenschaftlichkeit“ der Tweets durch die Kodierer mit der wahrgenommenen „Wissenschaftlichkeit“ der Tweets durch die Proband_innen verglichen. Da die Voraussetzungen für die Verwendung eines parametrischen Verfahrens nicht gegeben sind (nicht vorhandene Normalverteilung), wurde ein Mann-Whitney-U-Test gerechnet. Der nicht-parametrische Mittelwertvergleich zeigt für die Gruppe mit „wissenschaftlich“ wahrgenommenen Tweets (Mdn = 0.69 = 69%) gegenüber der Gruppe mit „nicht-wissenschaftlich“ wahrgenommenen Tweets (Mdn = 0.32 = 32%) einen statistisch signifikanten Unterschied mit $U = 799.00$, $z = -6.38$, $p < 0.001$. Dies bedeutet, dass die vermeintlich „nicht-wissenschaftlichen“ Tweets und die vermeintlich „wissenschaftlichen“ Tweets von den Proband_innen ebenfalls so bewertet wurden wie von den Kodierern. Die Grundannahme bezüglich der ausgewählten Tweets wurde bestätigt, „wissenschaftliche“ und „nicht-wissenschaftliche“ Tweets werden von Laien und Expert_innen (d.h. Kodierern) ähnlich wahrgenommen. Durch dieses Ergebnis sind nun auch die weiteren statistischen Analysen möglich.

Bezogen auf die ursprüngliche Forschungsfrage, nämlich die „wissenschaftliche“ Wahrnehmung von Tweets, kann ein statistisch signifikantes Ergebnis berichtet werden. Die Anzahl der Hashtags korreliert statistisch signifikant negativ mit $r_s = -0.23$, $p = 0.003$ mit der Klassifikation der Proband_innen, dass ein Tweet als „wissenschaftlich“ eingeschätzt wird. Um dieses Ergebnis nochmals zu überprüfen, wurde ein Mann-Whitney-U-Test gerechnet. Hierzu wurden alle Tweets, welche häufiger als 50% als „wissenschaftlich“ bewertet wurden, in eine Gruppe kategorisiert, Tweets mit weniger als 51%-Bewertung in eine zweite Gruppe hinzugefügt. In Bezug auf Hashtags zeigt der nicht-parametrische Mittelwertvergleich für die Gruppe mit mehrheitlich „wissenschaftlich“ wahrgenommenen Tweets (Mdn = 3.00) gegenüber der Gruppe mit nicht-mehrheitlich „wissenschaftlich“ wahrgenommenen Tweets (Mdn = 5.00) ebenfalls einen statistisch signifikanten Un-

terschied mit $U = 2070.50$, $z = -3.34$, $p = 0.001$. Je mehr Hashtags vorhanden sind, desto weniger „wissenschaftlich“ wirkt der Tweet. Andere Korrelationen mit Tweet-Eigenschaften – wie z.B., ob ein Tweet als Frage formuliert ist, die Anzahl der @-Erwähnungen oder ob ein Video vorhanden ist – führten zu Null-Korrelationen.

Da sich in der anfänglichen manuellen Kategorisierung der Tweets schon gezeigt hatte, dass Tweets häufig nicht nur einer Eigenschaft wie z.B. „enthält eine Abbildung“ zugeordnet werden konnten, wurde auch der Zusammenhang zwischen den verschiedenen Eigenschaften statistisch (über Spearman-Korrelationen) überprüft. Die Anzahl der Retweets korreliert mit einem statistisch signifikanten Wert von $r_s = 0.70$, $p < 0.001$ positiv mit der Anzahl der Favorisierungen („Gefällt mir“-Funktion bzw. Herz-Funktion). Je häufiger ein Tweet retweetet wird, desto mehr Gefallen findet er in der Community. Das Vorhandensein eines Links im Tweet korreliert mit $r_s = -0.21$, $p = 0.007$ negativ mit der Anzahl der Favorisierungen („Gefällt mir“-Funktion“ bzw. Herz-Funktion) und negativ mit der Anzahl an Retweets ($r_s = -0.21$, $p = 0.008$). Links sorgen also eher dafür, dass ein Tweet nicht favorisiert wird bzw. nicht retweetet wird. Das Alphaniveau wurde konservativ mit der Bonferroni-Methode korrigiert und beträgt für alle berichteten Ergebnisse 0.006. Derartige Auswertungen lassen Hinweise darauf zu, ob sich Tweet-Eigenschaften gegenseitig verstärken und damit als Bündel Einfluss auf die wahrgenommene „Wissenschaftlichkeit“ nehmen (was in zukünftigen Studien überprüft werden soll).

Die Ergebnisse der statistischen Auswertung zu den sieben Tweet-Kategorien sind als Boxplots in Abbildung 2 dargestellt. Es ist zu erkennen, dass Tweets, welche in der Kategorie „Methodik“ ($M = 0.76$, $SD = 0.19$ und ein Median von 0.80) enthalten sind, am ehesten als „wissenschaftlich“ anerkannt werden (hier benannt als „Akzeptanz“). Tweets in den Kategorien „Artikel“ ($M = 0.53$, $SD = 0.25$ und ein Median von 0.57) und „Verständnis“ ($M = 0.54$, $SD = 0.19$ und ein Median von 0.50) werden im Durchschnitt mindestens zur Hälfte als „wissenschaftlich“ bewertet (Akzeptanzwerte ab 50%). Solche Tweets, welche den Kategorien „Spaß“ ($M = 0.07$, $SD = 0.08$ und ein Median von 0.07), „Meinung“ ($M = 0.35$, $SD = 0.23$ und ein Median von 0.43), „Publicity“ ($M = 0.32$, $SD = 0.26$ und ein Median von 0.30) und „Job“ ($M = 0.15$, $SD = 0.16$ und ein Median von 0.10) zugeordnet wurden, werden mehrheitlich als „nicht-wissenschaftlich“ angesehen. Eine starke Streuung wie z.B. in den Kategorien „Artikel“ oder „Publicity“ deutet auf eine grundlegende Ambivalenz bei der Beurteilung hin.

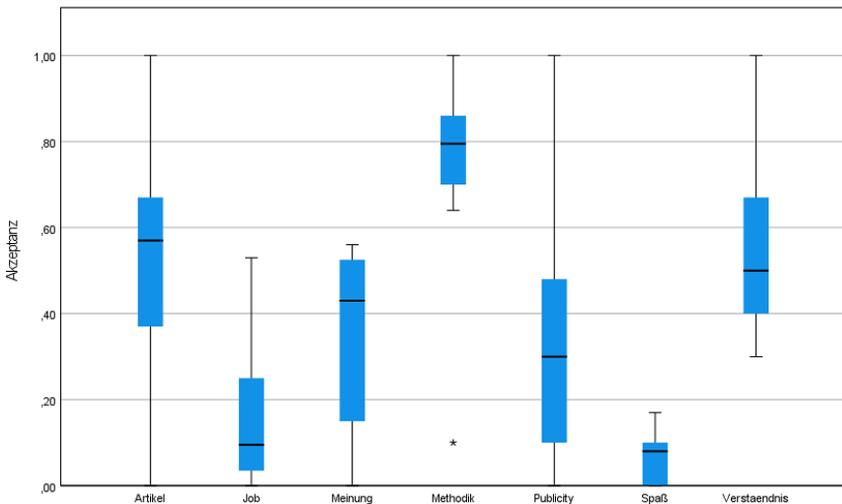


Abb. 2 Boxplots-Analyse der Tweets im Hinblick auf die Bewertung der „Wissenschaftlichkeit“ in Zusammenhang mit der Kategorisierung; x-Achse: Kategorien; y-Achse: Proband_inneneinschätzung: „wissenschaftlich“ ja/nein; * = Ausreißer

3.2 Vergleich von Tweets, die als eindeutig „wissenschaftlich“ und eindeutig „nicht-wissenschaftlich“ kategorisiert wurden

Im Folgenden werden zur Veranschaulichung bestimmte Tweets aus der Online-Befragung dargestellt, welche entweder sehr häufig (Abb. 3) oder nie (Abb. 4 u. 5) als „wissenschaftlich“ wahrgenommen wurden. Es werden nur Tweets dargestellt, die mindestens von zehn Proband_innen bewertet wurden.

Im direkten Vergleich beider Gruppen von Tweets fällt auf, dass Tweets, welche als „wissenschaftlich“ betrachtet werden, tatsächlich Aspekte aus der wissenschaftlichen Tätigkeit beschreiben. Das bedeutet, entweder wird auf eine wissenschaftliche Publikation, einen wissenschaftlichen Bericht oder medizinische und lebenswissenschaftliche Kongresse hingewiesen oder verlinkt.

Im Gegensatz dazu sind Tweets, welche als „nicht wissenschaftlich“ betrachtet werden, thematisch eher heterogen. Zwar wird der Themenbereich „Wissenschaft“ im weitesten Sinne angesprochen, z.B. in Form von wissenschaftlichen Jobangeboten sowie der Nennung oder Vorstellung von Institutionen oder Personal, jedoch fehlt in diesen Tweets der direkte Bezug zu wissenschaftlichen Ergebnissen bzw. zu den Aspekten aus der wissenschaftlichen Tätigkeit. Sobald aber dieser Bezug vorhanden ist, steigt die „Wissenschaftlichkeit“ der Tweets an. So werden Tweets, die auf Kongresse auf-

merksam machen, eher als „wissenschaftlich“ bewertet, da auf Kongressen wissenschaftliche Ergebnisse präsentiert werden.

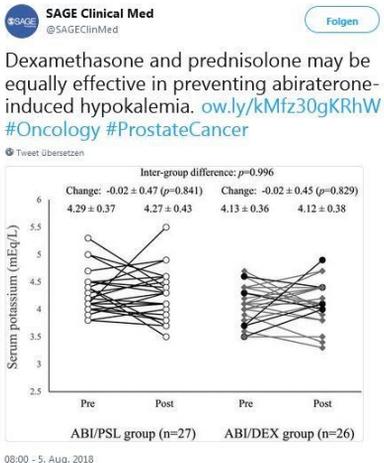
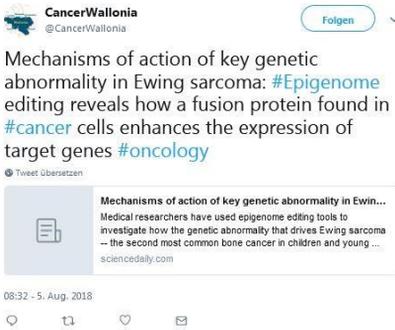
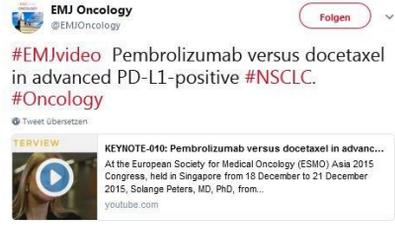
 <p>08:00 - 5. Aug. 2018</p>	 <p>08:32 - 5. Aug. 2018</p>
<p>100% Akzeptanz der ProbandInnen https://twitter.com/SAGEClinMed/status/1026120800175370241</p>	<p>93% Akzeptanz der ProbandInnen https://twitter.com/CancerWallonia/status/1026128705230385152</p>
<p>Dieser Tweet wurde von allen befragten Personen als „wissenschaftlich“ bewertet. Der Link im Tweet führt zu einem wissenschaftlichen Artikel. Der Text im Tweet gibt ein Resultat der Publikation wieder.</p>	<p>Der Text gibt ein fachliches Resultat wieder, der Link verlinkt auf einen Artikel in Sciedaily.com, eine wissenschaftliche Nachrichtenseite.</p>
 <p>00:47 - 5. Aug. 2018</p>	 <p>07:20 - 5. Aug. 2018</p>
<p>90% Akzeptanz der ProbandInnen https://twitter.com/IllingworthInfo/status/1026011901761925120</p>	<p>90% Akzeptanz der ProbandInnen https://twitter.com/EMJOnco/status/1036257481897455623</p>
<p>Im Tweet selbst steht ein Auszug des Titels einer Zeitschriftenpublikation. Der Link leitet zur Publikation weiter. Im Link-Erklärungstext ist ein Auszug aus dem Abstract zu lesen.</p>	<p>Der Tweet-Text ist ein Auszug des Titels einer Keynote. Der Link verweist auf ein Video, welches eine Vorschau auf die Keynote gibt.</p>

Abb. 3 Beispielhafte Tweets, die eindeutig als „wissenschaftlich“ in der Online-Befragung bewertet wurden (Kategorisierungen: oben links: Methode; oben rechts: Artikel; beide unten: Methode)

<p>Lokmanya Hospitals @LIVHospitals Folgen</p> <p>Hydration during winter is especially essential as it fights inflammation and keeps your joints working fit & fine. If you too are looking for a comprehensive joint treatment, visit #LokmanyaHospitals: goo.gl/a4U8iu #Health #Healthy #Pune #Orthopedics #NoPainThisWinter</p>  <p>21:30 - 25. Okt. 2018</p> <p>1 „Gefällt mir“-Angabe</p>	<p>Q Laboratories, Inc. @QLaboratories Folgen</p> <p>@QLaboratories was honored to be named a Finalist for the Deal Maker Award by @ACGCincinnati @ACGGlobal bit.ly/2kiYxC #laboratory #microbiology #chemistry #analytical</p>  <p>08:13 - 23. Mai 2018</p> <p>1 „Gefällt mir“-Angabe</p>
<p>0% Akzeptanz der ProbandInnen</p> <p>https://twitter.com/LMhospitals/status/1055677982961004547</p> <p>Der Text beschreibt den Zusammenhang zwischen Hydration und Entzündungen, um dann zur eigentlichen Werbung überzuleiten. Folgt man dem Link, kommt man auf die Internetseite der Institution, welche auch im Bild anhand des Logos erkennbar ist, bzw. welche diesen Tweet gepostet hat.</p>	<p>0% Akzeptanz der ProbandInnen</p> <p>https://twitter.com/QLaboratories/status/999307365588619264</p> <p>Der Tweet weist auf eine Auszeichnung hin, welche die postende Institution bekommen hat. Im Bild ist der Award zu sehen. Der Link leitet auf die offizielle Pressemitteilung der Institution anlässlich einer Preisverleihung.</p>

Abb. 4 Beispielhafte Tweets, die eindeutig als „nicht wissenschaftlich“ in der Online-Befragung bewertet wurden (Kategorisierungen: beide: Publicity)

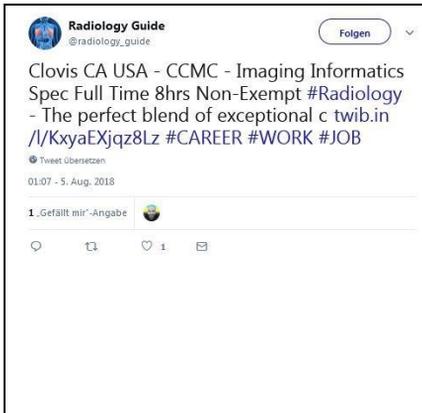
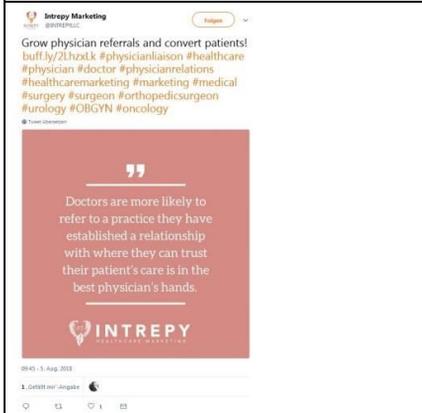
	
<p>0% Akzeptanz der ProbandInnen https://twitter.com/radiology_guide (Account ist derzeit gesperrt)</p>	<p>0% Akzeptanz der ProbandInnen https://twitter.com/bonniejaddario/status/1026142076071759872</p>
<p>Der Tweet ist eine Jobanzeige. Hinter dem Link verbirgt sich eine Jobplattform.</p>	<p>Der Tweet ist ein Spendenaufruf und erklärt, wie dieser funktioniert, mit einem Bild, auf der die postende Person zu sehen ist (letzteres lässt sich nur durch Interpretation des Tweets erschließen).</p>
	
<p>0% Akzeptanz der ProbandInnen https://twitter.com/INTREPYLLC/status/1027234266709745666</p>	<p>0% Akzeptanz der ProbandInnen https://twitter.com/JenaHall/status/1055656894960975872</p>
<p>Der Link des Tweets leitet auf einen Artikel auf der Webseite der zu postenden Institution weiter. Ein Aufforderungssatz, gefolgt von Schlagwörtern in Form von Hashtags, leitet das Thema ein. Ein Bild, welches ein Zitat darstellen soll, beinhaltet einen Satz, der so nicht im Artikel hinter dem Link zu finden ist.</p>	<p>Der Tweet weist auf ein Team hin, welches auf dem Bild abgebildet ist.</p>

Abb. 5 Beispielhafte Tweets, die eindeutig als „nicht wissenschaftlich“ in der Online-Befragung bewertet wurden (Kategorisierungen: oben links: Job; oben rechts: Publicity; beide unten: Publicity)

4 Diskussion der Ergebnisse

Die Antworten der Online-Befragung geben im Zusammenhang mit den Tweet-Eigenschaften Aufschluss darüber, welche Tweets als „wissenschaftlich“ wahrgenommen werden und welche nicht. In der Detailanalyse der Tweets, die seitens der Proband_innen weitestgehend eindeutig als „wissenschaftlich“ und „nicht-wissenschaftlich“ bewertet wurden, weisen diejenigen Tweets, die als „wissenschaftlich“ (primär aus der Kategorie „Methodik“) bewertet werden, eine stärkere Fokussierung auf Aspekte aus der wissenschaftlichen Tätigkeit auf und beinhalten Informationen zu Publikationen, Abbildungen aus Publikationen oder von Kongressen.

Als „nicht-wissenschaftlich“ eingestufte Tweets (primär aus den Kategorien „Job“ und „Spaß“) haben eine größere thematische Breite. Bei allen übrigen Kategorien gibt es eine große Streuung der Bewertungen. Eine mögliche Erklärung für dieses Ergebnis könnte sein, dass die Kategorien nicht ausreichend trennscharf sind. Dagegen spricht allerdings das quantitative Ergebnis, da keine bemerkenswerten Unterschiede in der Standardabweichung existieren. Es ist eher anzunehmen, dass Tweets aus den anderen Kategorien („Verständnis“, „Publicity“, „Meinung“ und „Artikel“) sehr ambivalent bewertet werden und hier eine sehr hohe Subjektivität vorhanden ist. Hierbei ist zu beachten, dass Tweets, welche fachspezifische Abkürzungen und Fachtermini enthielten, sodass sie für Laien nicht verständlich sind, aus dem Sample entfernt wurden. Es besteht aber die Möglichkeit, dass diese Tweets eventuell als „wissenschaftlich“ wahrgenommen werden, gerade weil man die Fachtermini nicht versteht (Weisberg et al., 2007).

Die durchgeführte Korrelationsanalyse und der nicht-parametrische Mittelwertvergleich zeigen zudem, dass Tweets mit vielen Hashtags als „nicht-wissenschaftlich“ wahrgenommen werden. Schaut man sich die beispielhafte Gegenüberstellung der Tweets in den Abbildungen 3, 4 und 5 an, dann zeigt sich, dass als „nicht-wissenschaftlich“ eingestufte Tweets häufig mehr als drei Hashtags verwenden, während als eindeutig „wissenschaftlich“ kategorisierte Tweets maximal drei Hashtags oder weniger verwenden.

Eine wichtige Einschränkung der Studie liegt in der Ausgestaltung der Online-Befragung begründet. In der Online-Befragung hatten die Proband_innen nur Zugriff auf ein Bild (Screenshot) des Tweets und mussten den Tweet ohne die Prüfung von z.B. weiterführenden Links im Hinblick auf seine „Wissenschaftlichkeit“ bewerten. Das Linkziel ist jedoch ein wichtiger

Bestandteil eines Tweets und wurde dementsprechend auch in die manuelle Klassifikation durch die Kodierer einbezogen. Für die Einschätzung eines Tweets spielt möglicherweise der potenzielle Zugriff auf die Ressource oder die Linkbezeichnung, auf die verwiesen wird, ebenfalls eine Rolle. Durch die fehlende Möglichkeit, wie auf Twitter, den Link anzuklicken, war es Proband_innen nicht direkt möglich, eine eventuell vorhandene subjektive Einschätzung bezüglich des Links zu überprüfen. Dies könnte sich auf die Ergebnisse ausgewirkt haben, sodass zukünftige Studien idealerweise diese Funktionalität gewährleisten sollten.

Eine weitere Einschränkung dieser Studie ist auch die relativ geringe Anzahl an Bewertungen je Tweet. Dies könnte teilweise die fehlende Eindeutigkeit bzw. die hohe Varianz bestimmter Ergebnisse erklären. Die quantitative Unausgeglichenheit zwischen den sieben Tweet-Kategorien in dieser Studie lässt sich mit der Wahl der Hashtags für die Suche nach Tweets erklären. Es ist zu vermuten, dass die ausgewählten Hashtags eher dazu genutzt werden, bereits sehr „wissenschaftliche“ Themen und Tweets zu beschreiben.

Während der größte Teil der präsentierten Tweets in englischer Sprache vorliegt (86%), die Proband_innen hingegen deutschsprachig sind und auch der Fragebogen in deutscher Sprache verfasst wurde, ist es denkbar, dass an dieser Stelle ein Bias vorliegen könnte. Zwar konnten wir in unserer Analyse keine ungewöhnlichen Ergebnisse bezüglich der Sprache entdecken, trotzdem ist es möglich, dass eine Studie mit einer einsprachigen Ausrichtung der Tweets, bzw. mit rein englischsprachigen Proband_innen, zu anderen Ergebnissen gekommen wäre.

Schließlich wäre noch zu bedenken, ob nicht schon allein die tweetende Person als ausschlaggebendster Faktor für Wissenschaftlichkeit wahrgenommen wird, also ein sogenannter „Guru-Effekt“ auftritt (Sperber, 2010), und diesbezüglich die Tweets auch entsprechend identifiziert werden. Dies würde allerdings ein anderes Forschungsdesign benötigen, um diese interessante Fragestellung experimentell zu untersuchen, z. B. die Verwendung von Tweets von bekannten und weniger bekannten Wissenschaftler_innen.

5 Zusammenfassung und Ausblick

Die vorgestellte Studie hatte sich zum Ziel gesetzt, die „Wissenschaftlichkeit“ von Tweets zu medizinischen Themen in Abhängigkeit von verschiede-

nen Tweet-Eigenschaften (z.B. Anzahl Retweets, Tweet-Typ) über eine Online-Befragung zu untersuchen. Die Ergebnisse beruhen auf einer Analyse von 162 klassifizierten Tweets und ihrer Bewertung durch 109 Proband_innen.

Die Ergebnisse lassen sich wie folgt zusammenfassen: Tweets mit vielen Hashtags werden eher als „nicht-wissenschaftlich“ wahrgenommen, wobei hier maximal drei Hashtags noch positiv sind, eine höhere Anzahl hingegen korreliert negativ mit der wahrgenommenen „Wissenschaftlichkeit“. Diese Erkenntnis beruht einerseits auf der Analyse der Abbildungen 3, 4 und 5, andererseits liegt der Median der Anzahl der Hashtags für die Gruppe mit mehrheitlich „wissenschaftlich“ wahrgenommenen Tweets ebenfalls bei drei Hashtags. Da allerdings der Median der Anzahl der Hashtags für die Gruppe mit mehrheitlich „nicht-wissenschaftlich“ wahrgenommenen Tweets bei fünf Hashtags liegt, könnten auch vier Hashtags noch akzeptabel sein. Hierbei ist auch der zu Beginn angemerkte Layout-Effekt zu erwähnen, da die Anzahl der Hashtags durchaus einen starken Einfluss auf die optische Erscheinung der Tweets hat.

Eindeutig zugeordnet wurden von Proband_innen offenbar solche Tweets, die sich mit den Kategorien „Methodik“, „Job“ und „Spaß“ beschäftigen, wobei nur zur ersten Kategorie zugehörige Tweets überwiegend als „wissenschaftlich“ bezeichnet, die beiden anderen Kategorien hingegen als eindeutig „nicht-wissenschaftlich“ kategorisiert wurden. Tweets, die sich den Kategorien „Artikel“ und „Verständnis“ zuordnen lassen, liefern kein eindeutiges Ergebnis, enthalten aber ebenfalls mehrheitlich Tweets, die eindeutig als „wissenschaftlich“ bewertet wurden. Dies ist insbesondere dann der Fall, wenn die Tweets auch auf Aspekte aus der wissenschaftlichen Tätigkeit verweisen oder diese zum Gegenstand haben, wie zum Beispiel Tweets im Kontext von einem Kongress mit einer Fokussierung auf wissenschaftliche Publikationen und Vorträge.

Zukünftig sind weitere Studien geplant. Hierbei bieten sich mindestens zwei Anpassungen an. Einerseits können die medizinischen Fachbezeichnungen ersetzt oder ergänzt werden um Hashtags zu Krankheiten (z.B. Diabetes). Fachbezeichnungen wurden in dieser Studie deswegen verwendet, damit die thematische Breite der Hashtags und Tweets nicht zu groß wird. Weitere Untersuchungen sind notwendig, um die Validität dieses Ansatzes zu überprüfen. Andererseits kann auch ein nicht-medizinisches Setting in Betracht gezogen werden. Durch ein solches Setting besteht die Möglichkeit, die Ergebnisse dieser Studie zu generalisieren.

Literaturverzeichnis

- Bruns, A.; Stieglitz, S. (2013): Towards more systematic Twitter analysis: metrics for tweeting activities. *International Journal of Social Research Methodology: Theory and Practice* 16 (2), 91–108.
- Haustein, S.; Bowman, T. D.; Holmberg, K.; Peters, I.; Larivière, V. (2014): Astrophysicists on Twitter: An in-depth analysis of tweeting and scientific publication behavior. *Aslib Journal of Information Management* 66 (3), 279–296.
- Hölig, S.; Hasebrink, U. (2020): Reuters Institute Digital News Report 2020 – Ergebnisse für Deutschland. Arbeitspapiere des Hans-Bredow-Instituts | Projektergebnisse Nr. 50. Hamburg: Verlag Hans-Bredow-Institut. https://www.hans-bredow-institut.de/uploads/media/default/cms/media/66q2yde_AP50_RIDNR20_Deutschland.pdf
- Islam, M. S.; Sarkar, T.; Khan, S. H.; Mostofa Kamal, A.-H.; Hasan, S. M. M.; Kabir, A. ... Seale, H. (2020). COVID-19–Related Infodemic and Its Impact on Public Health: A Global Social Media Analysis. *The American Journal of Tropical Medicine and Hygiene*. <http://www.ajtmh.org/content/journals/10.4269/ajtmh.20-0812>
- Lemke, S.; Mazarakis, A.; Peters, I. (2016): Characteristics of Twitter usage at scientific conferences. In: *Proceedings of the 18th General Online Research Conference* (pp. 34–35). Dresden: Deutsche Gesellschaft für Online-Forschung (DGOF).
- Lemke, S.; Mehrazar, M.; Mazarakis, A.; Peters, I. (2019): “When You Use Social Media You Are Not Working”: Barriers for the Use of Metrics in Social Sciences. *Frontiers in Research Metrics and Analytics* 3, 39. <https://doi.org/10.3389/frma.2018.00039>
- Letierce, J.; Passant, A.; Breslin, J.; Decker, S. (2010): Understanding how Twitter is used to spread scientific messages. In: *Proceedings of Web Science Conf. 2010*, April 26–27, 2010, Raleigh, NC, USA.
- Mazarakis, A.; Peters, I. (2015): Science 2.0 and Conference Tweets: What? Where? Why? When? *Electronic Journal of Knowledge Management* 13 (4), 269–282. <http://www.ejkm.com/issue/download.html?idArticle=730>
- McKendrick, D. R. A.; Cumming, G. P.; Lee, A. J. (2012): Increased use of Twitter at a medical conference: A report and a review of the educational opportunities. *Journal of Medical Internet Research* 14 (6), e176. <https://doi.org/10.2196/jmir.2144>
- McQuate, S. (2017): What all those scientists on Twitter are really doing. *Nature* 12, e0175368. <https://doi.org/10.1038/nature.2017.21873>

- Osterrieder, A. (2013): The value and use of social media as communication tool in the plant sciences. *Plant Methods* 9 (1), 26. <https://doi.org/10.1186/1746-4811-9-26>
- Salem, J.; Borgmann, H.; Murphy, D. G. (2016): Integrating Social Media into Urologic Health care: What Can We Learn from Other Disciplines? *Current Urology Reports* 17 (2), 13. <https://doi.org/10.1007/s11934-015-0570-2>
- Sperber, D. (2010): The Guru Effect. *Review of Philosophy and Psychology* 1 (4), 583–592. <https://doi.org/10.1007/s13164-010-0025-0>
- Sugimoto, C. R.; Work, S.; Larivière, V.; Haustein, S. (2017). Scholarly use of social media and altmetrics: A review of the literature. *Journal of the Association for Information Science and Technology* 68 (9), 2037–2062.
- Weisberg, D. S.; Keil, F. C.; Goodstein, J.; Rawson, E.; Gray, J. R. (2007): The Seductive Allure of Neuroscience Explanations. *Journal of Cognitive Neuroscience* 20 (3), 470–477.
- Weller, K.; Dröge, E.; Puschmann, C. (2011): Citation Analysis in Twitter: Approaches for Defining and Measuring Information Flows within Tweets during Scientific Conferences. In M. Rowe, M. Stankovic, A.-S. Dadzie, M. Hardey (Hrsg.), *Making Sense of Microposts (#MSM2011), Workshop at Extended Semantic Web Conference (ESWC 2011), Crete, Greece*. Heidelberg: Springer. http://ceur-ws.org/Vol-718/paper_04.pdf