

(Publikums-)Fragmentierung *online*

Analysen der Mediennutzung kommerzieller Onlinemedien in
Deutschland im Jahr 2016

Inaugural-Dissertation
zur Erlangung des Doktorgrades der Philosophie (Dr. phil.)
durch die Philosophische Fakultät der
Heinrich-Heine-Universität Düsseldorf

vorgelegt von
Inga Brentel

geboren in
Tettnang

Betreuer:
apl. Prof. Dr. Olaf Jandura
Prof. Dr. Ralph Weiß

Lindau September 2022

D61

Vorwort

Dieses Dissertationsprojekt wurde im Rahmen des Forschungsverbundes Digitale Gesellschaft NRW durch das Ministerium für Kultur und Wissenschaft des Landes Nordrhein-Westfalen gefördert. Als Teil des Promotionstandems „Angebots- und Publikumsfragmentierung online“ ist der vorliegende Kumulus der Inaugural-Dissertation im Verbund mit einer medienökonomischen Dissertation von Céline Fabienne Kampes mit dem Schwerpunkt der Angebotsfragmentierung online entstanden. Das Tandem-Promotionsprojekten startete im Winter 2017 / 2018. wurde durch die Principal Investigators Appl.-Prof. Dr. Olaf Jandura und Prof. Dr. Olexiy Khabyuk begleitet und war mit fünf weiteren Tandem-Promotionsprojekten eingebettet in das Graduiertenkolleg „Digitale Gesellschaft NRW“. Entsprechend ist die vorliegende Doktorarbeit mit dem Titel „Publikumsfragmentierung online“ in einem interdisziplinären Umfeld entstanden und von dem transdisziplinären Gedanken (medienpolitischer) Praxis geprägt.

Der stetige Blick über den fachspezifischen Tellerrand hat innovative Ansätze für die Aufbereitung der Forschungsdaten sowie die Analysen ermöglicht. Stets begleitet durch die Frage nach der Praxisrelevanz der eigenen Forschung stand im Vorfeld der Einzelstudien immer das „warum“; welchen Erkenntnisgewinn kann das eigene Forschungsprojekt für aktuelle gesellschaftliche und medienpolitische Debatten und Fragen über den fachspezifischen Mehrwert hinaus leisten. Dies hat die Ausrichtung der Analysen stark geprägt. Letztlich konnte eine Große Anfrage zum Onlinemedienmarkt in NRW angeregt werden und dabei erste Erkenntnisse zur Entwicklung medienpolitisch relevanter Fragen einfließen. So bestand das Promotionstandem „Angebots- und Publikumsfragmentierung“ aus mehr als den angefertigten Forschungsartikeln und der beiden verfassten Dissertationen. Denn die Diskurse die innerhalb des Projekts, im Rahmen des Forschungsverbundes Digitale Gesellschaft, auf Fachtagungen, veranstalteten Symposien und darüber hinaus stattfanden, gehörten ebenso dazu.

Lindau, den 30. September 2022

Übersicht der schriftlichen Dissertationsleistung

Diese kumulative Inaugural-Dissertation zur Erlangung des Doktorgrades der Philosophie (Dr. phil.) enthält eine Dachschrift, vier Forschungsartikel, eine Datenpublikation mit den aufbereiteten Daten der Media-Analyse *IntermediaPlus* für die Jahre 2014 bis 2016 sowie vier weitere Publikationen, welche das im Rahmen der Dissertation entwickelte Datenmanagementsystem zur Aufbereitung der Daten dokumentieren und beinhalten.

Die Dachschrift gibt einen Überblick über die thematischen Schwerpunkte der einzelnen Forschungsartikel (Einzelschriften) und deren wissenschaftliche Zusammenhänge. Die folgenden vier Einzelschriften sind gemeinsam mit der jeweiligen Erklärung der Eigenanteile in voller Länge in Anhang II enthalten:

Text 1

Brentel, I. & Winters, K. (2021). The Longitudinal IntermediaPlus (2014–2016): A Case Study in Structuring Unstructured Big Data. *Research Data Journal for the Humanities and Social Sciences* (6), 1-16. doi: <https://doi.org/10.1163/24523666-06010001>

Text 2

Brentel, I. & Winters, K. (2022). Processing and Documenting Big Data According to the FAIR Principles: A Communications Research Use Case. Manuskript zur Publikation eingereicht.

Text 3

Kampes, C. F., & **Brentel, I.** (2020). The German online media market: Online-born information offerings and their audiences – A shift towards digital inequalities? *World of Media. Journal of Russian Media and Journalism Studies* (4). 5-34. doi: <https://doi.org/10.30547/worldofmedia.4.2020.1>

Text 4

Brentel, I. (2022). Online media consumption in Germany: The role of political information. An analysis of German mass communication online. *Manuskript*.

Text 2 und Text 4 sind derzeit in einem peer-review Verfahren oder in Veröffentlichung und entsprechend in voller Länge als Manuskript Teil des Anhangs.

Für die Inaugural-Dissertation kann die Datenpublikation des aufbereiteten *Längsschnittes der IntermediaPlus (2014-2016)* auf Grund der Größe und des Dateiformats lediglich digital verfügbar gemacht werden. Die aufbereitete Datenquelle kann über den offen zugänglichen DBK-Katalog des GESIS Leibnitz-Instituts in Köln online abgerufen werden:

Datenpublikation

Brentel, I., Kampes, C. F., & Jandura, O. (2022). Media-Analyse-Daten: IntermediaPlus Daten von 2014 bis 2016 (MA IntermediaPlus). *GESIS Datenarchiv*, Köln. ZA5769 Datenfile Version 1.0.0, <https://doi.org/10.4232/1.13530>

Auch die vier Publikationen, welche das im Rahmen der Dissertation entwickelte Datenmanagementsystem zur Aufbereitung der *IntermediaPlus (2014-2016)* dokumentieren und die dafür nötigen Dateien und Skripte enthalten, können wegen des Dateiformats sowie deren Größe lediglich digital verfügbar gemacht werden. Sie sind nachfolgend aufgelistet und können über das open-source Repository SoWiDataNet des GESIS Leibnitz-Institut in Köln abgerufen werden.

Begleitpublikation Daten 1

Brentel, I., Jandura, O., Kampes, C. F., & Khabyuk, O. (2021). *Datenaufbereitung der MA-IntermediaPlus: Teil 1. Ein Überblick und Prozess-Schritt 1 „Strukturieren & Anpassen“*, GESIS, Cologne. SoWiDataNet Datenfile Version 1.0.0, <https://doi.org/10.7802/2339>

Begleitpublikation Daten 2

Brentel, I. (2021a). *Datenaufbereitung der MA-IntermediaPlus: Teil 2. Prozess-Schritt 2 „Import & Harmonisieren“ in und mit CharmStats*. GESIS, Cologne. SoWiDataNet Datenfile Version 1.0.0, <https://doi.org/10.7802/2340>

Begleitpublikation Daten 3

Brentel, I. (2021b). *Datenaufbereitung der MA-IntermediaPlus: Teil 3. Prozess-Schritt 3 „Harmonisierung Anwenden“ mit charmstana*. GESIS, Cologne. SoWiDataNet Datenfile Version 1.0.0, <https://doi.org/10.7802/2341>

Begleitpublikation Daten 4

Brentel, I., & Fortmann, N. (2021). *Datenaufbereitung der MA-IntermediaPlus: Teil 4. Prozess-Schritt 4 „Dokumentation erstellen“*. GESIS, Cologne. SoWiDataNet Datenfile Version 1.0.0, <https://doi.org/10.7802/2342>

Die Erklärungen der Eigenanteile zu den datenbezogenen Publikationen befinden sich ebenfalls in Anhang II.

Danksagung

Ohne die Unterstützung von vielen besonderen Menschen aus meinem beruflichen und privaten Umfeld wäre die Realisierung dieser Dissertation nicht möglich gewesen. Ihnen möchte ich Danke sagen. Ein besonderer Dank geht an ...

... meine Familie und Freunde, die gerade zum Schluss oft zu kurz kamen und mir nie einen Vorwurf daraus machten, mich unterstützten, aufbauten, wenn es nötig war und mich mit ihrem Stolz auf das Privileg der Promotion motivierten. Aber auch für die Ablenkung abseits der Promotion, wenn ich sie brauchte.

... meine Eltern für die Möglichkeit studieren zu können, die bedingungslose Unterstützung auf dem Weg der Promotion, dass sie mir immer mit Rat und Tat zur Seite standen und stehen.

... meinen Partner für die Geduld und das Zuhören bei den vielen Gedankenspielen zur Online-mediennutzung sowie die Unterstützung, das Verständnis, das Rückenfreihalten auf dem gesamten Weg der Promotion und für das Antreiben, wenn mir die Kraft ausging.

... meine Freundinnen und Kolleginnen Marike und Fabienne, dass wir den Weg der Promotion gemeinsam gehen konnten, mit allen Höhen und Tiefen; dass wir uns gegenseitig motivieren konnten, wenn es gerade nicht so lief und feiern konnten, wenn es gerade sehr gut lief. Und natürlich die fachlichen Diskussionen und Anmerkungen zu jeder Zeit.

... meine Schreibgruppe für das Zuhören, Unterstützen, Motivieren und das viele Feedback zu meinen unterschiedlichen Herausforderungen auf dem Weg der Abgabe – ihr habt mich wirklich sehr bei meiner fachlichen und persönlichen Weiterentwicklung begleitet und vorangebracht. Ein besonderer Dank gilt Simone als Initiatorin und Leitung der Schreibgruppe.

... meine Betreuer für den wissenschaftlichen Rat sowie anregende, fachliche Diskussionen, die mir argumentatorische Schwächen aufzeigten, mir damit halfen den roten Faden besser herauszuarbeiten und zu präzisieren.

... die beiden PIs für die Ermöglichung der Promotion, die Projektfinanzierung, anregende Diskussionen, die über den Tellerrand hinaus gingen und die gemeinsamen Tagungsbesuche.

... alle aus dem Graduiertenkollegs „Digitalen Gesellschaft“ für den fruchtbaren Austausch.

... meine Co-Autorinnen für die sehr gute Zusammenarbeit – ich habe viel von und mit euch gelernt.

Mein Dank gilt außerdem den anonymen Reviewern der publizierten und eingereichten Manuskripte für die Anmerkungen sowie den vielen Kolleginnen und Kollegen für anregende Diskussionen am Rande von Tagungen, Symposien, Workshops oder in der Kaffeeküche.

Dachschrift

Abstract

In einer immer pluraler werdenden Mediengesellschaft ist die Bevölkerung in ihrer Informationsgewinnung multiplen Optionen ausgesetzt, die Mediennutzung wird unübersichtlicher. Mit dem Aufkommen des Internets und seiner steigenden Bedeutung für die Informationssuche, ist die Debatte einer (Publikums-)Fragmentierung online entbrannt. So ist die Zahl der Studien, welche Fragmentierung, Filterblasen oder Echokammern untersuchen gestiegen. Eine überzeugende empirische Evidenz für die vielfach angenommene Publikumsfragmentierung online, welche häufige als logische Konsequenz eines Anstiegs der Medienangebote im Online-Kontext sowie neuer technologischer Möglichkeiten gesehen wird, bleibt jedoch aus. In Deutschland fehlen Studien zur massenmedialen Onlinenutzung, welche angesichts des Bedeutungszuwachs der Informationsnutzung als Voraussetzung für geteilte Themen und Wissensbestände sowie den darauf aufbauenden Diskurs gesehen werden kann, völlig. Ziel des hier vorliegenden Dissertationsprojektes ist es daher den ersten Teil des Fragmentierungsprozesses, gemeint ist hiermit die (gemeinsame) Mediennutzung, für kommerzielle, massenmediale Onlinemedien in Deutschland zu untersuchen. *Wie überschneiden sich demnach die (Teil-)Publika beziehungsweise lassen sich im Onlinebereich überschneidende Mediennutzungsmuster feststellen und, sofern es diese gibt, wie sehen diese aus.* Zudem können in der bestehenden Literatur zwei übergreifende Limitationen ausgemacht werden: 1) Es fehlt eine geeignete Datenbasis, die den nötigen Umfang der erhobenen Onlinemedien sowie Detailgrad und Zuverlässigkeit in der Erfassung zur Untersuchung von fragmentierter Onlinemediennutzung hat. Diese Forschungslücke wird durch Text 1 und Text 2 des Kumulus bearbeitet. 2) Studien analysieren entweder aus einer Mikro-Perspektive auf individueller Ebene die Medienselektion und -nutzung, z.B. in Form von Medienrepertoires oder Nutzungstypen (nutzerzentrierte Studien); oder untersuchen aus einer Makro-Perspektive Publikumsverteilungen (medienzentrierte Studien) sowie die Zusammensetzungen und Überschneidungen von (Teil-)Publika (publikumszentrierte Studien). Ein integrierendes Modell, in dem die verschiedenen Forschungsstränge klar definiert und in Beziehung zueinander gesetzt werden fehlt bislang. Im vorliegenden Dissertationsprojekt werden deshalb mithilfe des Mikro-Makro-Mikro Modells verschiedene Ansätze zur Analyse und Erklärung fragmentierter Mediennutzung online kontextualisiert und in den empirischen Einzelstudien anhand derselbe Datengrundlage berücksichtigt. In Text 3 wird zunächst unter Rückgriff auf die Digital Divide Forschung, eine eher nutzerzentrierte Perspektive, analysiert, inwiefern die politische Informationsnutzung sich entlang sozio-demographischer Merkmale der Nutzenden unterscheidet. Eine Makro-Perspektive wird dann in Text 4 eingenommen und analysiert, inwiefern sich individuelle Mediennutzungsmuster überschneiden sowie identifizierte Teilpublika auf Faktoren des Medienangebots und des Publikums untersucht, welche als Determinanten von Fragmentierung gelten.

Inhalt

| | |
|---|-----|
| Danksagung..... | III |
| 1 Einleitung..... | 1 |
| 2 Theoretische Grundlage und Forschungsstand | 8 |
| 2.1 <i>Fragmentierungsthese und Internet</i> | 8 |
| 2.2 <i>Fragmentierungsthese im Kontext der Integrationsfunktion von Massenmedien: gemeinsame Mediennutzung als Voraussetzung</i> | 10 |
| 2.3 <i>Konzepte von Fragmentierung und Forschungsansätze</i> | 12 |
| 2.4 <i>Schwerpunkt der Dissertation: Mediennutzungsmuster Online</i> | 15 |
| 2.5 <i>Das Mikro-Makro-Mikro Modell der Sozialtheorie: Publikumsfragmentierung als kollektives Phänomen</i> | 17 |
| 2.5.1 <i>Publikumsfragmentierung online im Kontext des Mikro-Makro-Modells</i> | 18 |
| 2.5.2 <i>Determinanten von Fragmentierung</i> | 21 |
| 2.5.3 <i>Ebenen der Publikumsfragmentierung</i> | 24 |
| 2.6 <i>Untersuchungsgegenstand der Dissertation: kommerzieller, massenmediale Einzelangebote auf Ebene inhaltlicher Gerne</i> | 29 |
| 3 Forschungsprogramm | 31 |
| 3.1 <i>Identifikation und Erschließung einer geeigneten Datenbasis</i> | 31 |
| 3.2 <i>Empirische Studien zu Publikumsfragmentierung online</i> | 33 |
| 4 Zusammenfassung, Limitationen und Ertrag der Einzelschriften..... | 38 |
| 4.1 <i>Text 1: The Longitudinal IntermediaPlus (2014–2016): A Case Study in Structuring Unstructured Big Data (Brentel & Winters 2021)</i> | 38 |
| 4.1.1 <i>Forschungslücke</i> | 38 |
| 4.1.2 <i>Ergebnisse</i> | 40 |
| 4.1.3 <i>Limitation</i> | 41 |
| 4.1.4 <i>Ertrag</i> | 43 |
| 4.2 <i>Text 2: Big-Data Management Case Study: The Process-Cycle for the Longitudinal IntermediaPlus (2014-2016) Dataset (Brentel & Winters, forthcoming)</i> | 44 |
| 4.2.1 <i>Forschungslücke</i> | 44 |
| 4.2.2 <i>Ergebnisse</i> | 45 |
| 4.2.3 <i>Limitation</i> | 49 |
| 4.2.4 <i>Ertrag</i> | 50 |

| | |
|---|-----------|
| 4.3 Text 3: The German online media market: Online-born information offerings and their audiences – A shift towards digital inequalities? (Kampes & Brentel 2020)..... | 51 |
| 4.3.1 Forschungslücke | 51 |
| 4.3.2 Ergebnisse | 52 |
| 4.3.3 Limitation | 53 |
| 4.3.4 Ertrag | 54 |
| 4.4 Text 4: Online media consumption in Germany: An analysis of German mass communication online (Brentel, forthcoming)..... | 55 |
| 4.4.1 Forschungslücke | 55 |
| 4.4.2 Ergebnisse | 56 |
| 4.4.3 Limitation | 58 |
| 4.4.4 Ertrag | 59 |
| 5 Ertrag des Kumulus, Diskussion und Ausblick..... | 60 |
| 5.1 Erkenntnismehrwert des Kumulus und Diskussion | 61 |
| 5.2 Ausblick..... | 63 |
| Literatur..... | 66 |
| Anhang I..... | I |
| Anhang II..... | I |

Abbildungsverzeichnis

| | |
|--|----|
| Abbildung 1: Schematische Darstellung des Mikro-Makro-Mikro Modell | 18 |
| Abbildung 2: Strukturierungseinheiten und -logik für die Aufbereitung der Rohdaten..... | 41 |

1 Einleitung

Jeder Transformationsprozess des Mediensystems durch aufkommen neuer Technologien ist von gesellschaftlichen und wissenschaftlichen Debatten begleitet, so auch bei Aufkommen des Internets (Wendelin 2011, 13-17; siehe auch Bächler 2017, 321; McQuail 2010, 136, 151f., 157; Neuberger 2017, 46). Zu Beginn wurde das Potenzial der Demokratisierung durch diese neue Technologie positiv hervorgehoben – im Markt wie bei der Nutzerschaft. Denn Markteintrittsbarrieren für Anbieter von Onlinemedien sind massiv gesunken; Distributionskosten wurden gesenkt, (Macht-)Monopole im Medienmarkt zunächst aufgeweicht und durch aufkommende neue Geschäfts- und Erlösmodelle sowie Akteure im Netz herausgefordert (vgl. Beck 2018, 320, 347ff., 351, 396, 400f.; Kampes, Jandura & Khabyuk 2021, 33f., 35; McQuail 2010, 156; Neuberger & Lobigs 2010, 28, 48; Stark & Kist 2020, 1140; Williams & Carpini 2011, 86f., 89f.). Auch für die Bürger wurde ein großes Potenzial zur Partizipation an Gesellschaften, zur freien Informationssuche und einer gesteigerten Repräsentation – da durch Onlinemedien, Social Media und Blogs die Möglichkeit besteht Minoritäten-Meinungen und -Perspektiven, partikulare Interessen, im öffentlichen Diskurs abzubilden – festgestellt (McQuail 2010, 151, 533; Neuberger 2017, 54, 56; Williams & Carpini 2011, 287f., 324). Hier wurde und wird ein großer emanzipatorischer Wert im Netz gesehen (Habermas 2021, 488), welcher in Bewegungen wie #blacklivesmatter, #metoo, u.a. Ausdruck findet.

Doch nach anfänglicher Euphorie schlugen die Utopien einer schönen neuen Welt, dem Nutzen des Internets für demokratische Gesellschaften, in Dystopien um (vgl. Coleman 2017, 3ff., 28-55; van Aelst et al. 2017), wie es bei Diskursen um neue Medien üblich ist (Neuberger 2017, 46): Insbesondere durch neue Marktmonopole, wie die des Google- und Facebook-Konzerns, welche Plattformen der Informationsdistributionen dominieren und somit als neue Gatekeeper¹ oder Intermediäre bezeichnet werden (siehe u.a. Beck 2018, 348, 397; vgl. Williams & Carpini 2011, 86f.; Stark, Magin & Jürgens 2021a, 304), welchen Meinungsmacht im öffentlichen Raum zugesprochen wird (Neuberger & Lobigs 2010, 27f.; Grotz & Schroeder 2021, 228) Neue technologische Möglichkeiten wie Personalisierung und Algorithmen (Stark et al. 2021a, 314f.; vgl. Rau & Stier 2019, 402, 405; Pariser 2011; Webster 2014, Williams & Carpini 2011, 86f.), die Empfehlungssysteme füllen (Stark & Kist 2020, 1141) und vorselektieren, welchen Ausschnitt der medialen Online-Öffentlichkeit Nutzende sehen beziehungsweise vorgeschlagen bekommen (Jürgens & Stark 2017, 396; Magin et al. 2022) – sei es bei der Google-Suche, auf einer Social-Media-Plattform oder auf der Onlineseite der

¹ Der Vollständigkeit halber sei darauf hingewiesen, dass Machill, Beiler und Zenker (2007, 7) dagegen argumentieren. Sie schreiben, dass Suchmaschinen keine klassische Gatekeeper-Rolle im Netz einnehmen würden, da Sie nicht bestimmen, welche Inhalte publiziert werden, wie dies im Offline-Kontext bei Medienhäusern der Fall war und nur ein Weg zu Online-Inhalten sind.

Lieblingszeitung. Desweiteren ist die Kehrseite der „Demokratisierung“ des Marktes mit einer Ausdifferenzierung der Onlineangebote² verbunden. Neben der potenziellen Steigerung der Vielfalt (Neuberger 2017, 56) durch die Vielzahl an Onlinemedianangeboten, steigt der Selektionsdruck auf Nutzer (Stark 2013, 214). Als Folge wird eine „Zersplitterung“ des Publikums, eine vollständige (Publikums-)Fragmentierung, befürchtet (Handel 2000, 21, 29; Stark 2013, 200).

Damit einhergehend wurde die These aufgestellt, dass Nutzende sich in voneinander abgegrenzte Publika online verabschieden: sogenannte Filterblasen (Pariser 2011), die sich auf Grund der Selektion und Rezeption unterschiedlicher Onlinemedienangebote bilden, welche anhand von variierenden Interessen und Präferenzen ausgewählt oder gar „durch algorithmische Personalisierung geschaffen[e]“ (Neuberger 2017, 56) werden. Einen Schritt weiter geht die Befürchtung der Echokammern von Sunstein (2001, 2007, 2017): Hier wird vermutet, dass Menschen mit gleichen (politischen) Einstellungen diese völlig voneinander abgetrennten Teilpublika bilden. Die Selektion der Medienangebote demnach anhand der wahrgenommenen, politischen Färbung der Inhalte erfolgt und gegensätzliche Meinungen und Argumente nicht mehr rezipiert oder ausgetauscht werden können. Dadurch soll eine Verstärkung der bestehenden Meinung und Einstellung resultieren und sich die voneinander getrennten Teilpublika immer stärker radikalieren (vgl. Stoud 2010). Eine Folge, so die Annahme, eine mehr und mehr polarisierte Gesellschaft, welche mit Voranschreiten der Fragmentierung früher oder später auseinanderbricht (vgl. Stark 2013, 206f.; Rau & Stier 2019; Tewksbury & Rittenberg 2009; Tewksbury & Riles 2015; Barnidge et al. 2021, 898f.). Denn der „social glue“ (Downey & Fenton 2003; vgl. Stark 2013, 199; Webster 2005), die medienvermittelte Information, Realitäten und Werte, geht auf Grund der fehlenden gemeinsamen Mediennutzung verloren (Vlašić 2004, 67ff.; Handel 2000, 12). Als bestätigend wird eine stärker polarisierte Gesellschaft in den USA gedeutet, Ursprungsland der Filterblasen und Echokammer Thesen (vgl. Rau & Stier 2019, 400f.; Stark 2013; siehe Sunstein 2001, 2007, 2017; Pariser 2011). Aber auch in Deutschland wurde eine voranschreitende Polarisierung beobachtet, die sich beispielsweise im Aufkommen der AfD als rechtspopulistische Partei, Bewegungen wie PEGIDA, die Debatte um Flüchtende³ und sogenannte Querdenker ausdrückt (vgl. Rau & Stier 2019, 400f., 412).

² Wobei durch eine Vermehrung der Anzahl der Angebote nicht notwendigerweise die Vielfalt nach inhaltlichen Bewertungskriterien gesteigert wird. So besteht die These des „more of the same“, also einer Vervielfältigung gleicher beziehungsweise ähnlicher Inhalte (Kampes et al. 2021, 202f.; Neuberger & Lobigs 2010, 34, 90; van Aelst et al. 2017). Dies wird auch „diversity paradox“ genannt (van Cuilenburg 1998, 44). Ebenso die These, der Bildung von Oligopolen in Anbieterstrukturen, „eine Ballung von Meinungsmacht bei wenigen Anbietern beziehungsweise Intermediären“ (Neuberger 2017, 56), als (wirtschaftliche) Folge eines Ausdifferenzierungsprozesses, welcher eher mit einem Rückgang der (Akteurs-)Vielfalt im Markt assoziiert wird (vgl. Kampes et al. 2021, 43, 45, 55f.).

³ Hier sind nicht die Flüchtenden vor dem Ukraine-Krieg gemeint.

Der Fokus wissenschaftlicher Studien lag bisher auf Social Media Plattformen (siehe u.a. Beaufort 2018; Dvir-Gvirsman 2017; Blank 2017; Jürgens, Jungherr & Schoen 2011; Kümpel 2020a; Merten 2021; Merten 2020; van Dijck 2012; Weber et al. 2019; für einen Literatur-Überblick siehe auch Rau & Stier 2019; Stark et al. 2021a; Mahrt 2019; Stark 2019), die durch ihre Eigenschaften als besonders guter Nährboden potenzieller Filterblasen und Echokammern gelten (vgl. Barnidge et al. 2021, 902; Grotz & Schroeder 2021, 224). Denn Newsfeeds sind algorithmisch gesteuert. Sie werden basierend auf Empfehlungssystemen gleicher, ähnlicher oder extremerer Inhalte mit dem Ziel Reaktionen und Klicks zu maximieren gefüllt. Dies macht Negativismus, Überspitzung, boulevardeske und populistische Inhalte beziehungsweise Headlines erfolgreicher als sachliche, (vgl. Rau & Stier 2019, 408; Breyer-Mayländer 2018, 27; Range & Schweins 2007, 5-8, 60; van Aelst et al. 2017, 4, 6). Dazu ähnelt die Kommunikation über soziale Netzwerke eher einer interpersonellen Kommunikation beziehungsweise steht im Zusammenhang persönlicher, sozialer Beziehungen (vgl. McQuail 2010, 144; Rau & Stier 2019, 404) anstelle einer eher unpersönlichen massenmedialen Kommunikation (McQuail 2010, 56f.), weshalb eine größere Wirkung von auf Social Media rezipierten Inhalten, die aus dem eigenen sozialen Umfeld stammen, vermutet werden kann⁴ (vgl. u.a. Schweiger 2017, 126; Bähr & Feuer 2018, 49-74 zu Influencer Marketing; Friemel 2013, 165-172 zur Rolle von Empfehlungen bei der Medienselektion).

Bisher bleiben jedoch eindeutige, empirische Belege für Filterblasen und Echokammern aus (Stark et al. 2021a; Rau & Stier 2019; Bruns 2019b; Mahrt 2020; vgl. Barnidge et al. 2021, 899, 902). Sowohl bei Studien, die Online-Suche und Social Media Nutzung in den Blick nehmen (siehe u.a. Beaufort 2018; Bruns 2019a; Bruns et al. 2017; Bruns & Enli 2018, 145f.; Haim et al. 2018; Nechushtai & Lewis 2019; Ørmen 2019a; Puschmann 2019), als auch bei Studien, welche crossmediale Onlinemediennutzung im massenmedialen Kontext untersuchen (siehe u.a. Dubois & Blank 2018; Mahrt 2019; Mukerjee, Majó-Vázquez & González-Bailón 2018a; Trilling & Schoenbach 2013; Taneja et al. 2012; Webster & Ksiazek 2012). Gestützt sind die meisten dieser Studien auf Befragungsdaten, wodurch die Onlinenutzung auf der Selbstausskunft der Befragten beruht (Taneja et al. 2012, 956; siehe u.a. Bächler 2017, Barnidge et al. 2021; Dubois & Blank 2018; Fletcher & Nielsen 2017; Kleinen von Königslow 2016; Stroud 2010; Trilling & Schoenbach 2013; Peng & Yang 2022). Diese Datenbasis birgt jedoch Unsicherheiten durch die Erinnerungsfähigkeit, soziale Erwünschtheit und anderen verzerrenden Faktoren in sich (Mukerjee et al. 2018a, 27; Olsen 2020, 9; Ørmen 2019b, 165-172). Zudem kann nur eine sehr limitierte Anzahl der genutzten Onlinemedien

⁴ Bei Social Media greifen demnach andere Wirkungsmechanismen als in massenmedialem Kontext. So werden professionell werbende Persönlichkeiten wie Influencer oftmals als Freunde wahrgenommen (Libor 2021, 396), ihre Meinung und Empfehlung durch die parasoziale Beziehung zum Rezipienten anders wahrgenommen als übliche Werbung oder Information eines Zeitungsartikels (vgl. Enke et al. 2021, 9, 11f.; Schweiger 2017, 126)

abgefragt werden. In einer ‚high-choice‘ Medienumgebung, wie dem Internet, wird auf Grund dieser Fallstricke die Reliabilität und Validität von Daten zu Mediennutzung, die der Selbstauskunft in Befragungen basieren, in Frage gestellt (Mukerjee et al. 2018a, 27; Olsen 2020, 9; Ørmen 2019b, 1656f.; Scharnow & Bachl 2017). Bei Untersuchungen von sozialen Netzwerken wird häufig auf digitale Spuren – sogenannte click-streams – zurückgegriffen, die auf Verlinkungsstrukturen, Abonnements von Seiten und Personen, Likes, Shares, Hyperlinks und ähnlichem basieren (vgl. Rau & Stier 2019, 408f.; siehe u.a. Bruns et al. 2011; Mahrt 2019, Kap. 7; Neuberger & Lobigs 2010, 127-154; Taneja & Wu 2018; Wu, Taneja & Webster 2021; Puschmann 2019). Analysen auf tatsächlichen Verhaltensdaten zur Mediennutzung online sind selten und fehlen für Deutschland in Bezug auf Massenmedien (in Abgrenzung zu sozialen Plattformen⁵) bislang völlig. Doch bildet die Mediennutzung gerade im massenmedialen Kontext die Basis der oben genannten Befürchtungen und Theorien (vgl. Vlašić 2004, 70). Gemeinsam genutzte Medien oder eben nicht gemeinsam genutzte Medien stehen am Anfang sogenannter Fragmentierungsketten, die über nicht geteilte Medienagenden zu polarisierten, segmentierten und zersplitterten Gesellschaften führen können sollen (Handel 2000, 12; Stark 2013, 199ff.). Daher sollte zunächst wissenschaftlich geklärt werden, ob eine gemeinsame Mediennutzung online gegeben ist oder nicht, um aufbauende Vorgänge im in Frage stehenden gesamtgesellschaftlichen Fragmentierungsprozess besser evaluieren und zielgerichteter untersuchen zu können.

Die vorliegende Dissertationsarbeit leistet hierzu einen Beitrag, indem die Nutzung kommerzieller Onlinemedien in Deutschland, die als Massenmedien konzipiert sind und deren Geschäftsmodell auf der Erzeugung und Verbreitung von Inhalten basiert (siehe Wirtz 2018), für 2016 untersucht wird. Denn „Demokratie braucht eine funktionierende Öffentlichkeit. Dazu gehören professionell-journalistische Medien (...) mit entsprechend großen Reichweiten“ (Schröder 2021, 103). Zum einen kann bei Onlinemedien, die als Massenmedien konzipiert sind und deren Geschäftsmodell die Erzeugung und Verbreitung von Inhalten ist, davon gegangen werden, dass es sich um überwiegend professionell-journalistische Medien handelt, welche zumindest den Anspruch haben größere Reichweiten zu erzielen und so Teil der (politischen) Öffentlichkeit online zu sein. Zudem wird sowohl im privatwirtschaftlichen Bereich als auch Online ein erhöhtes Risiko zur (Publikums-)Fragmentierung gesehen beziehungsweise diese attestiert (vgl. Bächler 2017, 321; Grotz & Schroeder 2021, 219, 228; Handel 2000, 36). Zum anderen gefährdet eine starke Publikumsfragmentierung im Sinne von voneinander abgetrennten Teilpublika ein Zustandekommen ausreichender Reichweiten, welche zur Konstitution (politischer) Öffentlichkeit als notwendig erachtet wird. Im Fokus steht

⁵ Zur Unterscheidung von „medialer Internetnutzung“, in dieser Arbeit als massenmediale Onlinenutzung bezeichnet, und der Internetnutzung zur „Individualkommunikation“ siehe auch Grotz und Schroeder (2021, 219) unter Rückgriff auf die Kategorisierung der ARD/ZDF-Onlinestudie 2019.

daher die überlappende Mediennutzung im Gegensatz zu nicht geteilter, also fragmentierter, Mediennutzung sowie die Rolle politischer Informationsnutzung in der Onlinemediennutzung. Denn besonders politischer Information wird einerseits als hoch relevant im Kontext medial vermittelter Öffentlichkeit in Demokratien wie Deutschland erachtet (vgl. Grotz & Schroeder 2021, 212; Jarren 2014, 325; McNair 2003, 20f.; Weiß & Jandura 2017, 14). Andererseits wird die Nutzung politischer Information als gefährdet für fragmentierte Mediennutzung angesehen (vgl. Babić & Jandura 2017, 108; Bächler 2017, 320f., 327; Barnidge et al. 2021, 902; Dohle, Jandura & Vowe 2014, 425; Grotz & Schroeder 2021, 219; Steppat, Herrero & Esser 2022; van Aelst et al. 2017, 4f.). Ziel ist es in einer ersten Exploration des Untersuchungsgegenstandes ein möglichst umfassendes Bild der (gemeinsam) genutzten Onlinemedien zu geben. Dabei soll sowohl die Mikroebene, die individuelle Mediennutzung, als auch die Makroebene, der Kontext in den die individuelle Mediennutzung eingebettet ist, berücksichtigt werden. Dadurch soll ein Analyserahmen für Folgestudien geschaffen werden, indem bisherige Studien eingeordnet und in Beziehung zueinander gesetzt werden können. Zusätzlich sollen so interessante Nutzergruppen für weitere, nachfolgende Studien identifiziert werden können, um der Frage „WER eigentlich fragmentiert ist“ (Handel 2000, 31) nachzugehen. Außerdem hat die Dissertation das Ziel Hinweise auf Determinanten von (fragmentierter) Onlinemediennutzung zu finden und einen Anfangspunkt für ein Monitoring eines potenziellen Fragmentierungsprozesses zu bieten.

Um die übergeordnete, leitende Forschungsfrage bearbeiten zu können, *inwiefern Onlinemedien gemeinsam genutzt werden und wie Onlinenutzungsmuster aussehen*, ist zunächst eine geeignete Datenbasis notwendig. Denn insbesondere bei der Onlinemediennutzung ist der Rückgriff auf Befragungsdaten, wie es bisher die Regel war, auf Grund der Vielfalt des Medienangebots und die veränderten Nutzungsmöglichkeiten nicht mehr valide oder reliabel, Verhaltensdaten zur tatsächlichen Nutzung online also unverzichtbar. Daher wurde zunächst für akademische Zwecke eine geeignete Datenquelle ausgemacht, erschlossen und zur Nachnutzung bei GESIS bereitgestellt (siehe Brentel et al. 2022, ZA5769 [Datenpublikation]; Brentel et al. 2021; Brentel 2021a; Brentel 2021b; Brentel & Fortmann 2021 [Begleitpublikation 1, 2, 3 und 4]). Dies bildete den ersten Teil des Dissertationsprojektes (siehe Text 1 und 2), bevor in einem zweiten Schritt mit ersten, explorativen, empirischen Arbeiten eine Antwort auf die leitende Forschungsfrage erarbeitet werden konnte. Bei dieser Datenquelle handelt es sich um die Media-Analyse Daten der IntermediaPlus (MA *IntermediaPlus*) für die Jahre 2014 bis 2016. Sie werden im Rahmen der Mediaforschung sowie zur Bestimmung von Reichweiten und Preisen für Werbeplätze auf den Einzelseiten der kommerziellen Onlinemedien Deutschlands erhoben. Es handelt sich dabei um die umfangreichste und detaillierteste Datenquelle für Mediennutzung, insbesondere für die online Mediennutzung, die zudem fortlaufend erhoben und angepasst wird (Strippel 2021,

Stark & Kist 2020, 1146; Brentel & Winters 2021 [Text 1]). Die MA *IntermediaPlus* bietet sich als Datenquelle an, da die benötigten Informationen enthalten sind: es ist eine für die deutsche Bevölkerung ab 14 Jahren repräsentative Erhebung, es werden soziodemographische Daten sowie Lifestyle-Merkmale der Nutzenden, welche in der Publikumsforschung herangezogen werden, um die Mediennutzung zu erklären und Publika zu charakterisieren, abgebildet und die Mediennutzung mit verschiedenen Parametern erfasst. Neben Onlinenutzung werden auch TV-, Radio- und Printnutzung abgebildet, was die Nachnutzung für anschließende Forschungsarbeiten zu Fragmentierung im crossmedialen Nutzungskontext ermöglicht und die Interoperabilität dieses Dissertationsprojektes steigert. Denn mit weiterführenden Studien auf Grundlage einer gemeinsamen Datenbasis kann eine Vergleichbarkeit der Ergebnisse hergestellt werden, welche aktuell fehlt (vgl. Stark 2013, 209). Weiterführende Studien sind wiederum notwendig, um ein umfassenderes Bild von Fragmentierung sowie der Rolle von Onlinemedien in der allgemeinen Mediennutzung erhalten zu können und methodische Debatten zur Messung von Onlinemediennutzung und Publikumsfragmentierung führen zu können.

Im Onlinekontext handelt es sich bei der Media-Analyse *IntermediaPlus* um die detaillierteste Erfassung von Onlinemedien, die datenschutzkonform in Deutschland möglich ist. Die Nutzung kann bis auf jede Einzelseite, sprich jede Unter- und Unter-Unterseite, eines Onlinemedienmedienangebots wie faz.net nachvollzogen werden⁶ – nur so kann Publikumsfragmentierung online hinreichend untersucht werden. In der Erhebung erfasst werden die Medienangebote, welche durch die Anbieter gemeldet werden, die Mitglied des Konsortiums hinter den Media-Analyse Daten sind.⁷ Dies umfasst die meisten deutschsprachigen, werbebasierten Onlinemedien deutscher Anbieter, die anhand der Media-Analyse Daten Werbefläche auf ihren Onlineseiten bepreisen. Eine solche Datenbasis ist in der akademischen Forschung kaum finanzierbar und liegt entsprechend bisher in Deutschland nicht vor. Die Media-Analyse Daten stehen aber zur akademischen Nutzung nach Ablauf eines zwei jährigen Embargos kostenfrei zur Verfügung, was sie zu einer sehr wertvollen Datenquelle zum Medienkonsum in Deutschland macht (siehe hierzu auch Stark & Kist 2020; Strippel 2021). Selbstverständlich haben die MA *IntermediaPlus* Daten auch Schwächen,

⁶ Eine Erfassung von Nutzungsdaten in Kombination mit Information über Nutzende ist auf Ebene von URLs, wodurch artikelspezifische Nutzung modelliert werden kann, nicht mehr möglich. Denn die Anonymität der Nutzenden beziehungsweise Befragten kann dann auf Grund von „digitalen Fingerabdrücken“ nicht gewährleistet werden.

⁷ Die Daten werden durch die ag.ma zur Verfügung gestellt. Das Konsortium besteht aus ag.ma, welche die Gesamtkoordination der Media-Analyse verantwortet, der AGF/GfK, die vor allem die TV-Daten beisteuern, sowie der agof, welche die Erfassung der Onlinemediennutzung organisiert. Ein Joint-Venture aus Anbietern, Werbetreibenden und Media-Agenturen sorgt dabei nach dem „Checks and Balances“ Prinzip durch ihre unterschiedlichen Interessen bei der Messung von Reichweiten, Unique Users, etc. für eine Qualitätssicherung dieser (siehe die Webseiten der Beteiligten Organisatoren ag.ma-mmc.de, agof.org und agf.de sowie Hofsäss, Engel & Kellerwessel 2020 für weitere Information).

welche die hier vorliegende Dissertation flankieren und Analysen limitieren. So ist die Eingrenzung auf bestimmte Teilbereiche der Analysen zur Mediennutzung online auch durch die verfügbare Datenlage geleitet und beispielsweise Nutzung öffentlich-rechtlicher Onlineangebote exkludiert, die beobachtete Nutzung auf die erfassten Onlinemedien begrenzt und die Aktualität der Daten eingeschränkt, da das aktuellste, aufbereitete Jahr 2016 ist. Trotz dieser Limitation ist es die umfassendste Datenquelle, die für Deutschland vorliegt (vgl. Stark & Kist 2020, 1146) und kann durch den Detailgrad der Erfassung der Onlinenutzung einen großen Mehrwert für die Kommunikationswissenschaft bieten (vgl. Webster & Ksiazek 2012, 51; Mukerjee 2021a, 9; Stark & Kist 2020, 1139). Daher bildet die Erschließung dieser wertvollen Datenquelle zur akademischen Nutzung die Grundlage dieses Dissertationsprojekts, um Publikumsfragmentierung online in Deutschland anhand erster explorativer, empirischer Analysen möglichst umfassend und mit größtmöglichem Detailgrad untersuchen zu können.

Zunächst wird in Kapitel zwei der vorliegenden Rahmenschrift der kumulativen Dissertation in einem ersten Teil der Forschungsstand zu (Publikums-)Fragmentierung, insbesondere im Online Kontext, sowie die gesellschaftliche Relevanz beschrieben. Dabei werden verschiedene Konzepte der Fragmentierungsforschung, unterschiedliche Analyseansätze sowie Untersuchungsebenen dieser vorgestellt. Daran anschließend werden die Schwerpunkte der hier vorliegenden Doktorarbeit gesetzt und definiert: fragmentierte beziehungsweise überlappende Mediennutzungsmuster online. In einem zweiten Teil folgt eine theoretische Rahmung für das hier vorliegende Dissertationsprojekt anhand des Mikro-Makro-Mikro Modells der Sozialtheorie. Hierin werden die Determinanten von Publikumsfragmentierung (online) verortet. Zudem werden verschiedene Ebenen der Publikumsfragmentierung vorgestellt und daraus resultierend der Untersuchungsgegenstand der Doktorarbeit definiert: die Nutzung kommerzieller, massenmedialer Onlineangebote auf Ebene der thematischen Genre, um eine an Inhalten ausgerichtete Publikumsfragmentierung analysieren zu können. In Kapitel drei wird das Forschungsprogramm des Dissertationsprojekts in Anlehnung an die theoretische Rahmung durch das Mikro-Makro-Mikro Modell vorgestellt und die konkreten forschungsleitenden Fragen der Einzelschriften formuliert. Sodann werden im vierten Kapitel die Einzelschriftenvorgestellt: Die jeweilige konkrete Forschungslücke wird aufgezeigt und die Ergebnisse des jeweiligen Beitrags zusammengefasst, Limitationen und der Ertrag für das Dissertationsprojekt der jeweiligen Einzelschrift diskutiert. Abschließend wird in Kapitel 5 der Ertrag des Kumulus durch die Verknüpfung der Einzelschriften aufgezeigt, der Erkenntnismehrwert für Wissenschaft und Politik diskutiert und ein Ausblick gegeben.

2 Theoretische Grundlage und Forschungsstand

Zunächst wird erläutert, was unter Fragmentierungsthese verstanden wird, sowie die gesellschaftliche Relevanz dieser und die Notwendigkeit von empirischen Studien für Deutschland, gerade im Kontext des Internets, kurz beschrieben. Aufbauend auf den fünf Dimensionen von Vlašić zur Integrationsfunktion von Massenmedien wird die Relevanz sowie das Erkenntnisinteresse der vorliegenden Dissertation expliziert. Dem folgt eine Strukturierung des Forschungsfelds anhand unterschiedlicher Konzepte für Fragmentierung. Nach einer ersten Eingrenzung anhand der (Begriffs-)Konzepte von Fragmentierung für die Dissertation erfolgt die Ableitung des Forschungsschwerpunkts. Daran schließt sich das theoretische Modell für die Analyse von Publikumsfragmentierung in der hier vorliegenden Arbeit an. Hier wird Mediennutzung theoretisiert, Einflussfaktoren auf diese sowie mögliche Determinanten von Publikumsfragmentierung ausgemacht und in das theoretische Modell eingebunden. Abschließend wird der Untersuchungsgegenstand des Dissertationsprojektes anhand der Ebenen der Publikumsfragmentierung bestimmt und konkretisiert.

2.1 Fragmentierungsthese und Internet

„Die Fragmentierungsthese ist zweifelsohne einer der zentralen Thesen in der Kommunikationswissenschaft“, schreibt Birgit Stark (2013, 200; vgl. Thimm & Anastasiadis 2017, 244; Peng & Yang 2022, 2271). Doch besteht „weder theoretisch noch methodischer Konsens über die zentralen Konzepte“ und es liegen „wenig belastbare empirische Belege für die These“ vor, so Stark weiter (vgl. Stier & Rau 2019; Stark et al. 2021a; Bruns 2019b; Peng & Yang 2022, 2271, 2273). Die meisten dieser empirischen Belege stammen aus den USA und sind nicht ohne weiteres für Deutschland übertragbar (Stark 2013, 213; vgl. auch Rau & Stier 2019, 399ff.; Fletcher & Nielsen 2017, 477, 482f.; Kleinen von Königslow 2016, 274; Mukerjee 2021a), da die USA als ein Spezialfall mit stark partei-politisch gefärbten Medienmarkt (partisan media) (Steppat et al. 2022, 85). Als Fragmentierungsthese wird mit Anstieg der Angebotsauswahl, einer Ausdifferenzierung des Medienmarktes, eine immer stärkere Individualisierung der Mediennutzung verstanden (McQuail & Deuze 2020, 597; Stark 2013, 201, 214; Webster & Ksiazek 2012, 1), die in eine „völlige Zersplitterung des Publikums“ münden kann (Stark 2013, 201; McQuail 2010, 444f.). Damit verbunden ist die Annahme sowie Sorge, eines Rückgangs gemeinsam genutzter Medienangebote und somit in „gemeinsam medial vermittelten Erfahrungen“, was wiederum zu einem „Rückgang gemeinsamer Gespräche und Kontakte mit Andersdenkenden“ führen und schließlich die gesellschaftliche Integration bedrohen kann (Handel 2000, 12; siehe auch Holtz-Bacha & Peiser 1999, 42-43). Denn massenmedial vermittelter Diskurs, die Herstellung von (politischer) Öffentlichkeit, wird als wichtige Stütze von Gesellschaften, vor allem in Demokratien gesehen (McNair 2003, 20f.;

Weiß & Jandura 2017, 104; Maletzke 1980; Stark, Rußmann & Magin 2021b, 4ff.; Jarren 2014, 320). In Deutschland spiegelt dies beispielsweise Artikel fünf des Grundgesetzes wieder (vgl. Grotz & Schroeder 2021, 212) und gilt in besonderem Maße für den öffentlich-rechtlichen Rundfunk. So stellt das Bundesverfassungsgericht im Rahmen der verschiedenen Rundfunk-Urteile immer wieder dessen öffentliche Aufgabe, die Rolle als Integrationsfaktor in der Gesellschaft, auch über rundfunkähnliche Kommunikationsdienste wie den Verbreitungskanal des Internets (BVerfGE 74, 297; BVerfGE 83, 283), klar (vgl. Neuberger 2017, 48f., 52)⁸. Spätestens seit dem Medienstaatsvertrag 2020 ist dann eine „Angleichung der regulativen Standards zwischen ‚alter‘ und ‚neuer‘ Medienwelt“ (Grotz & Schroeder 2021, 216) erfolgt. Der öffentliche Auftrag wird also auch den privatwirtschaftlichen Medienorganisationen in gewissem Maße zugeschrieben und erfüllt (vgl. Jarren 2014, 317). Denn einerseits haben Massenmedien eine Informationsfunktion, wodurch Meinungsbildung ermöglicht werden soll, damit mündige Bürger im gesellschaftlichen und politischen Prozess partizipieren können (vgl. Beaufort 2018, 915; Grotz & Schroeder, 2021, 211f.; McNair 2011, 18f.; Neuberger 2017, 52; Peng & Yang 2022, 2272, 2285; van Aelst et al. 2017, 4f.; Vlašić 2004, 37; Weiß & Jandura 2017, 13; Williams & Carpini 2011, 286f.). Andererseits wird ihnen eine Kontroll- und Kritikfunktion zugeschrieben (Grotz & Schroeder, 2021, 211f., 221; Holtz-Bacha 1997, 15f.; McNair 2011, 20; Vlašić 2004, 170f.). Als „vierte Gewalt“ im demokratischen System sollen sie Politik kritisch begleiten und entsprechend unabhängig von politischen Akteuren und Interessen agieren können.⁹ Empirische Untersuchungen zu Fragmentierung basieren, besonders in Deutschland, häufig auf Studien zur TV-Nutzung (Stark 2013, 208; vgl. Bächler 2017, 319f.; siehe u.a. Hasebrink & Popp 2006; Holtz-Bacha & Peiser 1999), wo die Fragmentierungsliteratur ihren Ursprung findet (siehe Webster & Wakschlag 1983 als Pionierstudien aus den USA). Belege für eine Zersplitterung des Publikums fehlen jedoch

⁸ Darauf fußt der öffentlich-rechtliche Rundfunk sowie medienpolitische Maßnahmen in Deutschland, um eine freie Willens- und Meinungsbildung anhand professioneller, journalistischer Inhalte zu sichern; Gerade für den öffentlich-rechtlichen Rundfunk werden hier hohe Maßstäbe angelegt, da dieser nicht primär Werbefinanziert ist. Doch auch werbebasierte Informationsmedien werden von politischer Seite beobachtet (siehe KEK, ALM, Medienanstalten u.a.). Bei Bedarf wird der Werbemarkt flankiert sowie Zusammenschlüsse von Medienhäusern unter Berücksichtigung der nutzungsbasierten Reichweite (u.a. anhand des KEK-Konzentrationsberichts sowie dem medienvielfaltsmonitor.de) nicht zugelassen, sollten diese zu einer zu großen Konzentration und damit Meinungsmacht einzelner Marktteilnehmer führen (siehe hierzu auch Grotz & Schroeder 2021, 212-221).

⁹ Siehe McQuail 2010, 161-185 für einen Überblick zu Theorien und Historie des Spannungsfelds zwischen (Massen-)Medien und Gesellschaft unter normativen Gesichtspunkten, spricht Medien als Institutionen öffentlichen Interesses und „vierte Gewalt“ (ebd., S. 168f.) sowie Grotz und Schroeder 2021, 221-225.

(Mangold, Vogelsang & Scharkow 2017, 707). Studien zur Internetnutzung in Deutschland sind nach wie vor selten (Stark 2013, 212f.) und beziehen sich meist auf Social Media Nutzung oder crossmediale Nutzung, wobei lediglich die Online-Nutzung etablierter Medienhäuser inkludiert ist oder die Onlinenutzung stark aggregiert abgefragt wird (vgl. Stark et al. 2021a; Rau & Stier 2019; Mahrt 2020; siehe u.a. Fletcher & Nielsen 2017; Puschmann 2019). Studien mit geeigneter Datenbasis zur Untersuchung von Publikumsfragmentierung im Netz in Deutschland sind daher wichtig, um Thesen wie die Bildung von Filterblasen und Echokammer im Internet besser einordnen zu können. Auch um evaluieren zu können, inwiefern im deutschen Kontext die Funktion von Massenmedien im Internet auf Grundlage einer hinreichend überlappenden Mediennutzung als Voraussetzung dieser gegeben ist.

2.2 Fragmentierungsthese im Kontext der Integrationsfunktion von Massenmedien: gemeinsame Mediennutzung als Voraussetzung

Wie bereits oben skizziert, wird Massenmedien in modernen Gesellschaften eine Integrationsfunktion zugeschrieben (Maletzke 1980; siehe auch Jarren 2000; Neuberger 2017). Dies gilt in besonderem Maße für demokratische Systeme wie Deutschland, da Medien hier eine Kontrollfunktion (vgl. Grotz & Schroeder, 2021, 212; Holtz-Bacha 1997, 15f.; Williams & Carpini 2011, 285) und in besonderem Maße eine Vermittlerfunktion als Intermediäre zwischen politischem System und Bürger:innen zukommt (McNair 2011, 19f.; Meier & Jarren 2001, 156f.) (z.B. im Kontext der Aggregationsfunktion und Akkumulationsfunktion). Publikumsfragmentierung stellt die Erfüllung dieser Integrationsfunktion von Massenmedien in Frage (Bächler 2017, 321; Magin et al. 2022), was die Untersuchung von Publikumsfragmentierung auch im Internet relevant macht. Besonders, da durch den Aufstieg des Internets eine Beschleunigung der Fragmentierung in Teilöffentlichkeiten gesehen wird (Grotz & Schroeder 2021, 228).

Vlašić fasst die Integration durch Massenmedien in einem Modell mit fünf Dimensionen zusammen: (1) *Bereitstellung von gemeinsamen Themen und Wissensbasis*, (2) *Ermöglichen von Repräsentation*, (3) *Konstituieren von (politischer) Öffentlichkeit*, (4) *Vermittlung gemeinsamer Normen und Werte*, (5) *Konstruktion von Realität* der eigenen Lebenswelten, aber auch durch die Selbst- und Fremdbeobachtung anhand massenmedial vermittelter Inhalte (Vlašić 2004, 67; siehe auch „Spiegelfunktion“ nach Luhmann 2005, 172; Wessler & Rinke 2016, 641). Damit strukturiert und expliziert er die Prämissen über die Funktionen von Massenmedien zur Erfüllung als Integrationsfaktor von (demokratischen) Gesellschaften. Denn der Anspruch an Massenmedien und massenmedial vermittelter Diskurs, variiert mit dem Verständnis von Demokratie und damit einhergehend der „Öffentlichkeit“, die durch

Massenmedien hergestellt werden soll (siehe auch Bächler 2017)¹⁰. Oftmals bleiben die Erwartungen an Massenmedien hinsichtlich ihrer integrativen Funktion dabei implizit und schwammig (Vlašić 2004, 67).

Die fünf zentralen Dimensionen der Integrationsfunktion sind auf unterschiedlichen Ebenen angesiedelt, die nicht ganz trennscharf sind und ineinander übergehen (vgl. Vlašić 2004, 68). Diese lassen sich auf Mikroebene, also der des Individuums, welches Medien nutzt, ebenso wie auf Makroebene, sprich der Ebene des sozialen Umfelds eines Individuums sowie weiter abstrahiert auf Ebene der Gesellschaft, beziehen. So ist die *Bereitstellung von Themen* beim Individuum und damit auf der Mikroebene zu verorten. Denn es ist eine „konkrete Funktion für den Einzelnen, indem sie Gesprächsstoff bietet für die Alltagskommunikation“ (Vlašić, 2004, 67). Während die *Repräsentation* zwar auf Ebene des Individuums angesiedelt ist, jedoch bereits in die Ebene des sozialen Umfelds diffundiert. Damit reicht es in die Makroebene hinein, wenn beispielsweise die eigene Lebenswirklichkeit sowie Interessen des sozialen Umfeldes, eine gesellschaftliche Gruppe, mit der sich ein Individuum identifiziert, als angemessen repräsentiert empfunden wird. Klar auf einer Makroebene anzusiedeln ist die *Konstitution der Öffentlichkeit*, bei welcher „[e]ine angemessene Repräsentation bzw. Präsenz der Individuen und ihrer organisierten Interessen (...) ein wesentlicher Bestandteil der politischen Öffentlichkeit“ ist (Vlašić 2004, 67). Dem übergeordnet liegen die Normen und Werte, welche durch (politische) Öffentlichkeit ausgehandelt werden und Teil medial-vermittelter Inhalte sind. Sie prägen das Wertesystem einer Gesellschaft und bilden damit eine übergeordnete Struktur in der Medien produziert und durch die diese beeinflusst werden. „Über allem schließlich steht die Frage nach der Konstruktion von Realität in einer Gesellschaft“ und damit die fünfte Dimension der Integrationsfunktion durch Massenmedien (Vlašić 2004, 68).¹¹

Die Fragmentierungsthese knüpft auf all diesen Ebenen als Störfaktor an. Je nach Fragmentierungsbegriff und Studienschwerpunkt kann die Untersuchung einer dieser fünf Dimensionen der Integrationsfunktion auch im Zentrum der Analysen stehen – beispielsweise die *Ermöglichung der Repräsentation* bei der Untersuchung von bereitgestellten Medieninhalten (siehe u.a. Neuberger & Lobigs 2010, 65-96; Kösters et al. 2021; Kösters 2020; Weiß, Kösters & Mahrt 2020) oder die *Bereitstellung gemeinsamer Themen* durch die Untersuchung der Themenagenden von einzelnen Nutzenden, Nutzergruppen oder (rezipierter) Medieninhalte (siehe u.a. Gehrau & Goertz 2010; Haas & Brosius 2013; Lee 2007;

¹⁰ Es kann insgesamt zwischen drei Traditionen des Öffentlichkeitsverständnisses unterschieden werden, welchen unterschiedliche Demokratievorstellungen zugrunde liegen: das liberale, deliberale und agnostisch-partizipatorische Öffentlichkeitsverständnis (Wessler & Rinke 2016; Jandura & Friedrich 2014; McQuail 2010, S. 175-185; zum liberal-repräsentative und partizipatorischem Öffentlichkeitsverständnis im Online-Kontext siehe Beaufort 2021, 70-78).

¹¹ Für eine ausführliche Beschreibung und Erläuterungen der einzelnen Integrationsfunktionen siehe Vlašić 2004, S. 69-85.

Magin et al. 2022; Mahrt 2019; Neuberger & Lobigs 2010, 70-125; Stark 2014; Yuan 2011). Die Integrationsfunktion von Massenmedien steht bei diesem Dissertationsprojekt jedoch nicht im Zentrum der empirischen Analysen. Vielmehr geht es hier darum zu untersuchen, ob die Voraussetzung, dass eine gemeinsame Wissensbasis durch geteilte Mediennutzung online hergestellt werden kann, erfüllt werden kann. Denn „[e]ine - wenn nicht sogar *die* - grundlegende Definition der Integrationsfunktion ist die Vorstellung, dass Massenmedien gemeinsame Themen bereitstellen, über die sich die Mitglieder der Gesellschaft verständigen können“ (Vlašić 2004, 70). So ermöglichen gemeinsame Themen und gemeinsames Wissen erst die Herstellung von (politischer) Öffentlichkeit (Vlašić 2004, 68; vgl. Magin et al. 2022). „Dabei dienen mediale Inhalte sowohl informativer (Katastrophen-Nachrichten etc.) als auch unterhaltamer Natur (Talk-Shows, Promi-Geschichten) als kommunikative Grundlage vor allem bei Interaktionen“ (Vlašić 2004, 70). Dadurch formen geteilte Themen und Wissensbestände soziale Realität (ebd. 68, 72; vgl. Magin et al. 2022) in welche Individuen wiederum eingebettet sind, sich in ihr verorten und durch sie beeinflusst werden. Dies entspricht der Vorstellung eines Mikro-Makro-Modells, bei welchem Individuen in (strukturelle) Kontexte eingebettet sind, die ihr individuelles Handeln determinieren. Durch individuelles Handeln wird wiederum im Aggregat der (strukturelle) Kontext beeinflusst und geformt (Coleman 1990).

2.3 Konzepte von Fragmentierung und Forschungsansätze

Im Zusammenhang mit Fragmentierung stehen demnach viele Themenfelder. So werden im Kontext der Fragmentierung von den Funktionen von Massenmedien in (demokratischen) Gesellschaften, dem Medienmarkt, -angebot und Anbieterstrukturen, über Mediennutzung, Inhalte und deren Wirkung, hin zu Auswirkungen (nicht) rezipierter Inhalte für moderne Gesellschaften viele Themen diskutiert und beforscht. Entsprechend interdisziplinär und teilweise unübersichtlich ist das Forschungsfeld. Zur Strukturierung identifiziert Handel (2000, 21ff.) drei Begriffskonzepte für Fragmentierung, welche verschiedene Schwerpunkte im Erkenntnisinteresse repräsentieren sowie unterschiedliche Implikationen für die Auswahl der Untersuchungsebenen und Studienanlage haben: (1) Die *Struktur der Gesellschaft*, bei der die Auswirkungen fragmentierter Mediennutzung im Mittelpunkt stehen. Häufig werden sie im Zusammenhang mit Segmentierungstendenzen diskutiert (siehe z.B. Weiß 2013), also einem „Zerfall der Gesellschaft in sozio-kulturelle Segmente (Wöste, 1999, S. 589)“ (Handel 2000, 21) mit einhergehender, steigender gesellschaftlicher Polarisierung. (2) Die *Struktur des Medienangebots*, welche sich primär mit einer Angebotsfragmentierung beschäftigen, sprich der Ausdifferenzierung und Multiplikation von Angeboten, ebenso wie die Ausbreitung und Spezialisierung in immer kleinteiligere Marktsegmente (Handel 2000, 22). Webster und

Ksiazek (2012), würden dies als medienzentrierte Forschung fassen, welche jedoch auch das Publikum als Kenngröße für die Reichweite einbeziehen und nicht vollständig distinktiv zu Handels dritten Begriffskonzept ist: (3) Die *Muster der Mediennutzung*, welches die Struktur des Publikums, eine mögliche *Publikumsfragmentierung*, fokussiert (Handel 2000, 22) und der hier vorliegenden Arbeit zugrunde liegt. Denn bestimmte Muster der Mediennutzung sind letztlich die Voraussetzung dafür, ob Medienangebote gemeinsam oder getrennt voneinander genutzt werden und entsprechend Überschneidungen in der Mediennutzung und dem Publikum gegeben sind oder voneinander getrennte, fragmentierte Teilpublika zu beobachten sind (Handel 2000, 22f.). Zudem kann die gemeinsame Mediennutzung als Voraussetzung für die Integrationsfunktion durch Massenmedien als zentrale Funktion in modernen (demokratischen) Gesellschaften gesehen werden (Vlašić 2004, 70). Das macht die Untersuchung des ersten Schritts der Fragmentierungsthese, die Mediennutzung, so relevant und ist Fokus des vorliegenden Dissertationsprojekts.

Studien mit dem Fragmentierungskonzept der Muster der Mediennutzung greifen häufig auf die Bestimmung von Medienrepertoires, die Identifikation von Medien-Menüs oder von Clustern des Publikums zurück, um eine Überschneidung in der Mediennutzung zu evaluieren (siehe Handel 2000, 25-31). Webster und Ksiazek (2012) differenzieren hier noch weiter und unterscheiden neben *medienzentrierten* auch zwischen *nutzer-* und *publikumszentrierten* Studien¹², welche Fragmentierung als Muster der Mediennutzung fassen. Medienzentrierte Studien nehmen die Verteilung des Publikums über die ausdifferenzierten Medienangebote, die „Entropie“ (Handel 2000, 24f.), als eine Kenngröße von (Angebots-)Fragmentierung in den Blick: so wird in Long-Tail Studien¹³, die insbesondere für das Medium Internet typisch zur Beschreibung der Struktur des Medienangebotes sind, die Reichweite der Angebote für die Analysen berücksichtigt (Webster & Ksiazek 2012, 42ff.; siehe Kampes 2020a; Kampes et al. 2021, Text 5: 81-106). Fokus solcher Studien ist jedoch die Struktur des Medienangebots, eine Angebotsfragmentierung, welche durch Mediennutzungsmuster beeinflusst wird.

Bei Studien, die auf dem Fragmentierungskonzept der Muster der Mediennutzung im Sinne der Publikumsfragmentierung basieren, ist vor allem die *nutzerzentrierte Perspektive* verbreitet (Webster & Ksiazek 2012, 45; Stark 2013, 210f.): Im Forschungsfokus stehen häufig Medienrepertoires, also “subsets of available media that individuals use on a day-to-day basis” (Webster & Ksiazek 2012, 45). Zum einen wird der Umfang der Repertoires, also die Anzahl der genutzten Medien, untersucht. Zum anderen deren inhaltliche Zusammensetzung

¹² siehe Stark 2013 (S. 209-214) für einen ausführlichen Überblick zu Studien im Internetzeitalter, die nach der Systematik von Webster & Ksiazek (2012) strukturiert sind; ebenso wie Mukerjee und Kollegen (2018a) für eine ausführliche Historie der Literatur mit einer publikumszentrierten Perspektive. Ein aktuellerer Literaturüberblick findet sich zudem u.a. bei Rau & Stier 2019, 406-409 sowie Stark und Kollegen 2021a, 308-314.

¹³ Diese Studien gehen auf die Long-Tail Theorie von Anderson (2006) zurück.

(Webster & Ksiazek 2012, 45; Handel 2000, 25ff.). Ein weiterer nutzerzentrierter Analyseansatz ist die Abweichung in der Nutzung von der durchschnittlichen Nutzung, welche für jede Nutzerin und jeden Nutzer berechnet wird, um den Grad der Überschneidung in den Nutzungsmustern zu bestimmen (siehe Handel 2000). Die Struktur des konkreten Medienangebots, welche genutzt werden, sowie die Verteilung der Nutzenden auf verschiedene Medienangebote, wird dabei eher außenvorgelassen.

Als *publikumszentriert* können solche Studien bezeichnet werden, die sowohl die Struktur des Medienangebots als auch die Individuelle Nutzung in ihre Betrachtung miteinbeziehen (Webster & Ksiazek 2012, 45f.). Darunter fallen Untersuchungen zu Medien-Menüs, also der Zusammensetzung von Repertoires ganzer Nutzergruppen, bei denen auch die Überschneidung des Publikums in einzelnen Medien betrachtet wird (siehe Handel 2000, 30). Ebenso zählen Studien dazu, welche die „Clusterung des Publikums“ (Handel 2000, 29) analysieren. Hier wird die individuelle Nutzung zu Nutzungsmustern auf aggregiert und die Verteilung über verschiedenen Medienangebote analysiert, um Aufschluss über die Überlappung der Publika einzelner Medienangebote zu gewinnen. So können sowohl zentrale Medienangebote mit einer hohen Publikumsüberschneidung identifiziert werden, als auch Medienangebote, die von bestimmten Publikumsgruppen exklusiv genutzt werden. Solche *publikumzentrierten* Studien „scale-up“ to the larger issues of how the public allocates its attention across media“ (Webster & Ksiazek 2012, 45). Dies erfolgt im Zusammenhang mit Onlinemediennutzung überwiegend anhand von sozialen Netzwerkanalysen¹⁴ (Webster & Ksiazek 2012, 46f.)¹⁵ und entspricht dem Ansatz der hier vorliegenden Arbeit.

Kleinen-von Königslow (2016, 257-259) fügt zur Konzeptionierung von Publikumsfragmentierung in der Online-Umgebung neben Angebot, Anbieter und Publikum eine weitere Perspektive auf Fragmentierung hinzu: Die genutzten bzw. bereitgestellten (*Medien-)Inhalte* (vgl. Mahrt 2019, 28). Hier werden geteilte Themenagenden, Frames und Positionen der einzelnen Medienangebote analysiert (siehe z.B Kösters 2020; Lee 2007; Weiß et al. 2020). Denn zwei unterschiedliche Medienangebote können große inhaltliche Überschneidungen haben und dadurch gegebenenfalls trotz getrennter Nutzung für geteilte Inhalte sorgen¹⁶. So kann die inhaltliche Komponente eine Brücke zu ansonsten getrennten

¹⁴ siehe Friemel 2013, 104-106 für eine Erklärung der grundlegenden Logik sozialer Netzwerkanalysen sowie die Seiten 107-123 für eine Adaption in der Publizistikwissenschaft sowie Strippel et al. 2018, 17ff. für den Einsatz von Netzwerkanalysen in der Kommunikationswissenschaft.

¹⁵ siehe auch Ksiazek 2011, Mukerjee et al. 2018a, und Mukerjee 2021a, 2021b sowie beispielsweise die Studien von Taneja et al. 2012 und Olsen 2020; für methodische Debatten zur Messung von Publikumsüberschneidung siehe Webster & Taneja 2018, Mukerjee et al. 2018b und Brentel *forthcoming* (Text 4).

¹⁶ Dabei spielt häufig auch die Qualität der untersuchten Berichterstattung eine Rolle sowie die interne vs. externe Vielfalt von Themen, Frames und Positionen, welche von Medienangeboten bereitgestellt werden (siehe z.B. Kösters et al. 2020; Neuberger & Lobigs 2010; Weiß et al. 2020). Die externe Vielfalt des angebotenen Medienangebots beziehungsweise genutzter Medienangebote wird so evaluiert und

Teilpublika schlagen und die integrative Funktion von Medien ermöglichen. Ebenso spielt der strukturelle bzw. politische Bias in den Inhalten der Medienangebote eine große Rolle (vgl. Steppat et al. 2022). Dies ist vor allem bei der Betrachtung von Polarisierungstendenzen in der Gesellschaft, welche der (exklusiven) Nutzung gegensätzlicher Medienangebote zugeschrieben werden, der Fall: Wenn von „Enklaven“, „Echokammern“, „red media – blue media“ und der gleichen gesprochen wird, bei denen die Nutzung von „like-minded media“ und die Abspaltung vom Gesamtpublikum in „like-minded“ Gruppen im Fokus des Forschungsinteresses stehen (siehe z.B. Sunstein 2007). In beiden Fällen wird in der Regel mit Inhaltsanalysen gearbeitet.

Die Hinzunahme der Betrachtungsebene der Medieninhalte verdeutlicht, dass es sich bei den unterschiedlichen Konzepten von Fragmentierung keineswegs um völlig losgelöste Forschungsfelder handelt (siehe Handel 2000, 30f.). So kann mit einem Methodenmix gearbeitet werden, um unterschiedliche Betrachtungsebenen in die Analyse zu Fragmentierung einzubeziehen (siehe z.B. Mahrt 2019) oder ein Teilaspekt, wie hier, fokussiert werden. Zudem wird häufig eine fragmentierte Struktur des Medienangebots als Ursache von fragmentierter und segmentierter Mediennutzung sowie daraus folgender Polarisierung und immer stärker segmentierte und auch stratifizierte Gesellschaft gesehen (Steppat et al. 2022). Studien mit verschiedenen Begriffskonzepten von Fragmentierung und entsprechend unterschiedlichen Forschungsschwerpunkten arbeiten zum Teil mit Grundannahmen, die in die Bereiche der jeweils anderen Forschungsfelder von Fragmentierung reichen, sich zum Beispiel auf Erkenntnisse dieser Studien beziehen.

2.4 Schwerpunkt der Dissertation: Mediennutzungsmuster Online

Dieses Dissertationsprojekt basiert auf dem Begriffskonzept der Fragmentierung als Muster der Mediennutzung, das den meisten Studien im Kontext von Fragmentierung zugrunde liegt (Handel 2000, 22). Denn eine fragmentierte Mediennutzung, also eine Nutzung anderer Medienangebote und Inhalte im Sinne einer Publikumsfragmentierung, ist der erste Schritt der Fragmentierungskette (vgl. Stark 2013, 199ff.; Handel 2000, 12). Als (Publikums-)Fragmentierung wird die „Zersplitterung des Publikums in viele Teilpublika, die

trägt zur Beurteilung von fragmentierten Mediennutzungsmustern bei: Zum einen, wie vielfältig sind unterschiedlich genutzte Medienangebote tatsächlich? Zum anderen, nutzen unterschiedliche Nutzergruppen und voneinander getrennte Publikumsgruppen tatsächlich unterschiedliche Inhalte oder kann durch Überschneidung in den Inhalten verschiedener Medienangebote eine Gemeinsamkeit in rezipierten Inhalten, Themen und Meinungen von ansonsten getrennten Segmenten hergestellt werden? Sprich können (genutzte) Medienangebote anhand ihrer inhaltlichen Überschneidung eine „bridging role“ zukommen (vgl. Barnidge et al. 2022, 916f.).

Unterschiedliches nutzen und nur noch selten zu einem großen Publikum zusammenkommen“ (Holtz-Bacha & Peiser 1999, 41; siehe auch Stark 2013, 200; McQuail & Deuze 2020, 597; Schulz 1999, 92) verstanden. Im Gegensatz dazu wird eine geteilte, nicht-fragmentierte, Mediennutzung als Voraussetzung zur Erfüllung der Integrationsfunktion von Massenmedien gesehen (vgl. Vlašić 2004, 70; Handel 2000, 12, 22f.), welche für den gesellschaftlichen Zusammenhalt als zentral erachtet wird (vgl. Stark 2013, 200f.; Downey & Fenton 2003; Jarren 2000; Webster 2005). (*Publikums-*)*Fragmentierung* wird daher in dieser Arbeit, ähnlich wie bei Handel (2000, 23), als ein Kontinuum verstanden: die vollständige Publikumsfragmentierung als ein Extrempunkt, der einen Zustand nicht überlappender Nutzung von einzelnen Medienangeboten – also keine Überschneidung von Publika – beschreibt (siehe Webster & Ksiazek 2012; Mukerjee et al. 2018a; vgl. Handel 2000, 32). Wobei eine vollständige Überschneidung von Publika bzw. Überlappung der Mediennutzung den Gegenpol bildet – spricht ein Zustand ohne jegliche Publikumsfragmentierung. Da zur Beschreibung des Zustands von Publikumsfragmentierung die Überschneidung bzw. nicht-Überschneidung von Publika ausschlaggebend ist, ist das Konzept der Mediennutzungsmuster für die hier vorliegende Arbeit so wesentlich. Folglich wird Publikumsfragmentierung unter anderem anhand der Überschneidung in Publika gemessen, welche als Resultat der Mediennutzungsmuster von Nutzenden verstanden wird.

Mit der Schwerpunktsetzung auf die *Muster der Mediennutzung* als Untersuchungsgegenstand von Publikumsfragmentierung, verortet sich diese Arbeit in der Mediennutzungsforschung. Für die konkreten Analysen sind demnach (Aus-)Wirkungen von Fragmentierung auf die *Struktur der Gesellschaft* ausgeklammert. Das heißt mögliche Folgen fragmentierter Mediennutzungsmuster, wie beispielsweise Abweichungen in Themenagenden, Polarisierungstendenzen in Meinungen und Einstellungen verschiedener Publika, werden nicht analytisch untersucht. So wird auch auf die inhaltsanalytische Betrachtung von genutzten Medienangeboten verzichtet.¹⁷ Da die Dissertation im Rahmen eines interdisziplinären Forschungsprojektes aus Sozialwissenschaften und Medienökonomie entstand, steht zwar die *Struktur des Medienangebots* nicht im Fokus, wird jedoch bei den Analysen sowie der Aufbereitung der Daten mitgedacht – ganz im Sinne eines *publikumszentrierten* Ansatzes nach Webster und Ksiazek (2012). Bisher ist im Online-Kontext auf Grund fehlender empirischer Evidenz unklar, ob Massenmedien Online – in Abgrenzung zu Social Media – fragmentiert genutzt werden oder aber eine Fragmentierung auf dieser ersten Ebene des

¹⁷ Zum einen, weil mit den verfügbaren Daten nicht auf Artikelebene nachvollzogen werden kann, welcher Artikel genutzt beziehungsweise tatsächlich gelesen wurde. Zum anderen, da eine inhaltsanalytische Betrachtung genutzter Onlinemedien vor allem dann sinnvoll ist, wenn ein Bezug zu Themenagenden der Nutzenden hergestellt wird, um Einblick in die (Aus-)Wirkungen möglicher fragmentierter Nutzung zu erhalten (siehe z.B. Mahrt 2019). Die Betrachtungsebene der genutzten Inhalte wird jedoch im Sinne thematischer Genrezuordnungen berücksichtigt.

Fragmentierungsprozesses ausbleibt. In dem vorliegenden Projekt ist folglich das übergeordnete Interesse herauszufinden:

1. *Inwiefern gibt es Überlappungen in den (Teil-)Publika bzw. überschneidende Mediennutzungsmuster online in Deutschland? Das heißt, inwiefern werden Onlinemedien gemeinsam genutzt und welche Rolle spielt politische Information dabei?*

Denn die gemeinsame Nutzung, der geteilte Zugriff auf die gleiche Informationsquelle – also ein Online-Medienangebot – wird hier als Voraussetzung für die Herstellung einer gemeinsamen Wissensbasis und Themen gesehen (vgl. auch Vlašić 2004, 70; Handel 2000, 12).

Insbesondere die (geteilte) Nutzung von politischer Information wird als substantiell zur Herstellung politischer Öffentlichkeit in einem demokratischen System wie Deutschland erachtet. Entsprechend wird ein besonderes Augenmerk auf die (geteilte) Nutzung politischer Information gelegt.

2. *Gibt es Teilpublika im Gesamtpublikum? Und wenn ja, wie bzw. was konstituiert diese? Sprich, welche Faktoren lassen sich für voneinander abweichende Medienmuster identifizieren und determinieren Publikumsfragmentierung online in Deutschland.*
3. *Wie kann eine geeignete Datenbasis zur Untersuchung von Publikumsfragmentierung online aussehen? Und wie kann diese nachhaltig (FAIR) aufbereitet und für akademische Forschung zugänglich gemacht werden?*

2.5 Das Mikro-Makro-Mikro Modell der Sozialtheorie: Publikumsfragmentierung als kollektives Phänomen

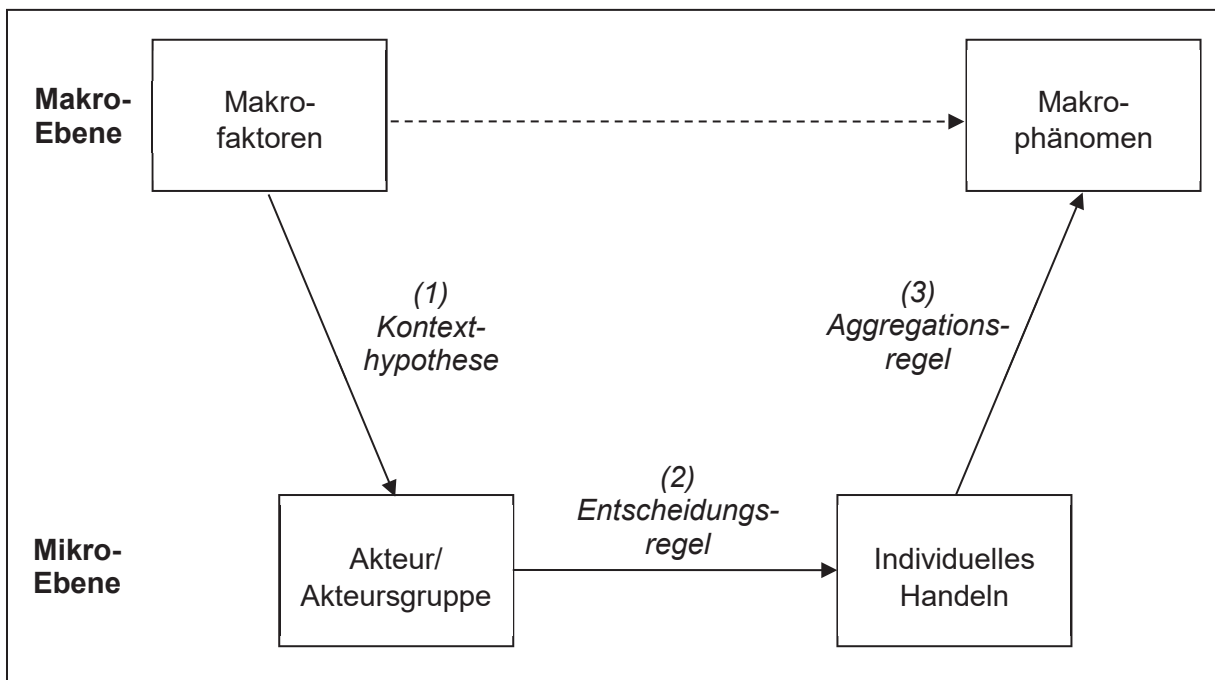
Zur Beantwortung dieser Fragen ist die Betrachtung von Publikumsfragmentierung online in dem strukturellen Kontext, in welchem die Mediennutzung stattfindet, notwendig. Zwar sind weder Markt noch Gesellschaft im Fokus der hier vorliegenden Dissertation, spielen aber als Kontextfaktoren von Nutzenden und ihrer Medienauswahl – der Mediennutzung – eine Rolle¹⁸. So können Einflussfaktoren der Mediennutzung auch Determinanten von Publikumsfragmentierung sein.

Das Fragmentierungsphänomen wird als ein Makrophänomen – ein Ergebnis kollektiven Handelns – begriffen, welches aus der Interaktion von Struktur und Akteuren entsteht (vgl.

¹⁸ Wie bereits die Konzeptualisierung der verschiedenen Fragmentierungsbegriffe (Handel 2000), die Fragmentierungshypothese (Handel 2000; Stark 2013) und das Zwiebelmodell der Mediennutzung (Schweiger 2007) zeigten, ist das Fragmentierungsphänomen und die Mediennutzung von Individuen eingebettet in die Markt- und Gesellschaftsstrukturen – eine Folge und Faktor von Wandlungsprozessen in Markt und Gesellschaft zugleich (Handel 2000).

Giddens' theory of structuration (1984); Webster & Ksiazek 2012, 40; Webster 2018, 97). Die Coleman'sche Badewanne (Coleman 1990) bzw. dessen Fortführung als Makro-Mikro-Makro-Modell der soziologischen Erklärung (Esser 1993) bietet sich daher als Heuristik und theoretischen Rahmen (Sozialtheorie) zur Untersuchung der Publikumsfragmentierung online an. Denn dieses Modell wurde zur Erklärung kollektiver Phänomene entwickelt und verschreibt sich einem methodologischen Individualismus (Greve, Schnabel & Schützeichel 2008, 7f.; Friemel 2013, 112f.). Das heißt, es bietet als Sozialtheorie einen Rahmen für die Erklärung und Beschreibung von Kollektivphänomen, wie der Publikumsfragmentierung online, anhand individuellen Handelns, hier der Mediennutzung von Individuen. Dies wird in Abbildung 1 schematisch dargestellt.

Abbildung 1: Schematische Darstellung des Mikro-Makro-Mikro Modell



Quelle: eigene Darstellung nach Coleman (1990) & Greve et al. (2008, 8)

2.5.1 Publikumsfragmentierung online im Kontext des Mikro-Makro-Modells

In Bezug auf die Beschreibung bzw. Erklärung von Publikumsfragmentierung online als ein Makrophänomen kollektiven Handelns, bedeutet dies, dass Makrofaktoren wie Strukturen des Medienmarkts oder der Gesellschaft Einfluss auf die Medienauswahl von Individuen auf der Mikroebene haben. Aus dieser Medienauswahl Einzelner bilden sich Mediennutzungsmuster heraus, welche zusammen wiederum den Status der Publikumsfragmentierung ergeben.

hierbei handelt es sich um eher direkte Zusammenhänge der einzelnen Mikro-Makro-Faktoren im Sinne der Sozialtheorie, wie Abbildung 1 zeigt.

Im Fall der Publikumsfragmentierung online werden für die hier vorliegende Arbeit die Strukturen des online Medienmarktes als Makrofaktoren der Medienauswahl von Individuen definiert. Denn sie geben einen Rahmen, die Auswahl an Onlinemedienangeboten, vor, welche die Medienauswahl von Individuen determinieren. Hier können mit Hilfe von Kontexthypothesen (1) Determinanten des Medienmarktes und -struktur von Publikumsfragmentierung online identifiziert werden (siehe Kapitel 2.5.2.2).

Individuen treffen innerhalb dieses Rahmens an Onlinemedienangeboten, die zur Verfügung stehen, eine Auswahl und zeigen ein bestimmtes Mediennutzungsverhalten, das in Mediennutzungsmuster resultiert. Diese Medienauswahl wird neben den Makrofaktoren, wie der Struktur des Medienmarkts, auch anhand von individuellen Einflussfaktoren getroffen. Diese können struktureller oder situativer Natur sein, welche auf Mikroebene angesiedelt sind. Anhand von Entscheidungsregeln (2), welche sich unter anderem aus Theorien der Mediennutzungsforschung bzw. des Medienverhaltens ableiten lassen, können weitere Determinanten von Publikumsfragmentierung online, die der Nutzenden bzw. des Publikums, identifiziert werden (siehe Kapitel 2.5.2.1).

Mit Hilfe von Aggregationsregeln (3) kann dann das individuelle Handeln, die Mediennutzungsmuster Einzelner, zu einem kollektiven Phänomen, dem Auftreten bzw. dem Grad von Publikumsfragmentierung online, von Mikro- zu Makroebene zusammengefasst werden. Wie diese Aggregation erfolgt ist zum einen abhängig von der Definition von Publikumsfragmentierung – hier die Überschneidung von Mediennutzungsmustern. Zum anderen ist es eine methodische Frage, wie eine Überschneidung in der Mediennutzung online gemessen werden soll (siehe Ksiazek 2011; Mukerjee et al. 2018a, 2018b; Taneja & Webster 2018 und Text 4 für verschiedene Ansätze hierzu).

Die Mediennutzungsmuster, also die Verteilung von Nutzenden auf verschiedene Medienangebote wirkt sich wiederum auf die Makrofaktoren der Mediennutzung aus. Es handelt sich aber eher um einen indirekten Zusammenhang, wie auch Abbildung 1 zeigt (vgl. auch Coleman 1990, 8f., 11f., 19ff.; Greve et al. 2008, 8f.). So kann das kollektive Phänomen eine Veränderung des Kontextes bedingen (Coleman 1990, 20f.). Im Fall der Publikumsfragmentierung (online) wird beispielsweise angenommen, dass die Verteilung der Aufmerksamkeit des Publikums Marktdynamiken beeinflusst und damit Fragmentierungstendenzen verstärken kann, da Anbieter und Werbemarkt sich an der Nachfrage der Nutzerschaft orientieren (vgl. Webster 2018; Handel 2000, 66; Grotz & Schroeder 2021, 216; McQuail 2010, 423). Ebenso wird in der Fragmentierungsthese davon ausgegangen, dass das Vorhandensein bzw. Fehlen überlappender Nutzungsmuster durch

den Grad des geteilten Wissens und der Themen sowie der damit verbundenen Konstitution (politischer) Öffentlichkeit (vgl. Vlašić 2004, 67ff., 70) moderne Gesellschaften prägt. In der hier vorliegenden Arbeit spielt dieser Aspekt auf Grund des gesetzten Fokus keine Rolle in den Analysen. So wird der prozesshafte Charakter der Publikumsfragmentierung nicht näher untersucht oder Wandlungsprozesse auf Makroebene, wie Struktur der Gesellschaft oder des Medienmarktes, weitere beleuchtet.

Unter Rückgriff auf das Mikro-Makro-Modell als theoretische Rahmung können Kontextfaktoren, welche als wichtige Einflussfaktoren der Mediennutzung und damit Determinanten von Publikumsfragmentierung (online) gelten, in die Analysen der Dissertationsarbeit eingebunden werden. Gleichzeitig liegt jedoch der Fokus bei den Analysen zu Publikumsfragmentierung online als Kollektivphänomen ganz im Sinne des methodologischen Individualismus (Greve et al. 2009, 7f.; Friemel 2013, 112f.) auf der Betrachtung der Mediennutzungsmuster Einzelner. Die Heuristik, welche das Mikro-Makro-Modell bietet, eignet sich zudem für den eher explorativen Charakter der empirischen Studien dieses Dissertationsprojektes. So können und sollen sowohl situative als auch strukturelle Determinanten verschiedener Dimensionen und Betrachtungsebenen (Markt, Gesellschaft, Individuum) im Zusammenhang mit der Kontexthypothese (siehe Abbildung 1) berücksichtigt werden¹⁹. Gleiches gilt für unterschiedliche Theorien des individuellen Handelns, den Entscheidungsregel in Abbildung 1. So können für die individuelle Medienauswahl und -nutzung beispielsweise die theoretischen Ansätze des Selective Exposure (siehe u.a. Knobloch-Westerwick 2014), Kognitiver Dissonanz (Festinger 1957), Uses and Gratification (Katz, Blumler & Gurevitch 1973; Palmgreen 1984), die Mood-Management Theorie (Zillmann 1985), Theorien der Sozialen Identität (Georgiou 2006; Storey 1999) oder das theoretische Konzept von Issue Publics (Converse 1964; siehe Mukerjee 2021a), eine Interessenfragmentierung (Handel 2000, 64f.; vgl. Waldherr et al. 2021, 164), zur Erklärung bestimmter Mediennutzungsmuster herangezogen werden (vgl. Webster 2018 zu „audience behaviour“ und Schweiger 2007 zu Mediennutzung). Einzelstudien mit unterschiedlichen theoretischen Ansätzen und ggf. verschiedenen Begriffskonzepten von Fragmentierung können dadurch kontextualisiert und (systematisch) in Bezug zueinander gesetzt werden. Zudem können Forschungsarbeiten mit verschiedenen Auffassungen bezüglich der Aggregationsregel, also dem Übersetzen von individuellem Handeln zu kollektiven Phänomenen, in dieses theoretische Modell integriert werden. Entsprechend können Mediennutzungsmuster für das Verständnis bzw. die Messung von Publikumsfragmentierung unterschiedlich konfiguriert und über dieses Dissertationsprojekt hinaus, z.B. für Medienrepertoires, erweitert werden (siehe u.a. Handel 2000, 24ff.).

¹⁹ siehe Schweiger 2007, 313-328 für einen Theorie-Überblick zu kontextbezogener Mediennutzung.

2.5.2 Determinanten von Fragmentierung

Mit der Flexibilität gegenüber verschiedenen theoretischen Erklärungen individuellen Handelns eignet sich das Mikro-Makro-Modell in besonderem Maße für die hier vorliegende, explorative Studie zur Publikumsfragmentierung online, welche in der Projektanlage ergebnisoffen ist: So ist nicht das Ziel eine bestimmte Theorie zu Mediennutzungsverhalten auf die Publikumsfragmentierung online anzuwenden, sondern das beobachtete Online-Nutzungsverhalten auf überschneidende vs. unterschiedliche Nutzungsmuster zu untersuchen. Dabei sollen mögliche Determinanten von Fragmentierung ausgemacht und geprüft werden, welche der Theorien von (ungleicher) Mediennutzung zur Erklärung der empirischen Befunde dienen können (vgl. Mukerjee 2021a). Dennoch ist eine gewisse Auswahl von möglichen Determinanten der Publikumsfragmentierung (online) nötig sowie eine Theoretisierung dieser. Zunächst werden die Determinanten auf Seiten des Publikums skizziert, welche bei der Betrachtung von Publikumsfragmentierung (online) eine Rolle spielen könnten. Dies umfasst zum einen Einflussfaktoren auf die Mediennutzung von Individuen, welche eher indirekt durch Strukturen der Gesellschaft, aber auch dem sozialen Umfeld wirken. Zum anderen werden Strukturen des Medienmarktes als mögliche Einflussfaktoren auf die Mediennutzung beschrieben, welche als Determinanten von Publikumsfragmentierung (online) in Verdacht stehen. Dabei werden verschiedene theoretische Ansätze, welche als Kontexthypothesen dienen können, genannt. Diese theoretischen Ansätze werden jedoch nicht näher ausgeführt, da sie lediglich dazu dienen mögliche, sinnvolle Determinanten von Publikumsfragmentierung zu identifizieren und Scheinkorrelationen besser ausschließen zu können.

2.5.2.1 Determinanten des Publikums: Faktoren der individuellen Mediennutzung auf Mikroebene

Als Determinanten von Fragmentierung²⁰ gelten beispielsweise allgemeine Individualisierungstendenzen auf Ebene der Gesellschaft, welche sich in der Mediennutzung niederschlagen können (Handel 2000, 59, 67). Als Faktor des sozialen Umfelds können die Sozialisation und damit einhergehende Werte, Einstellungen und auch Mediennutzungsgewohnheiten gesehen werden (Handel 2000, 60ff.). Ebenso wie soziale Identitäten, welche beispielsweise Einfluss auf Interessen, inhaltliche Präferenzen sowie die Medienwahrnehmung und -bewertung haben können (Mukerjee 2021a, 13f).

²⁰ siehe Handel 2000, 59-66 für eine ausführlichere Erklärung der einzelnen Determinanten von Fragmentierung und deren Zusammenhänge. Sie ordnet die einzelnen Determinanten zum Teil etwas anders zu, doch betont dabei, jedoch die fließenden Übergänge und den „fortlaufenden kreislaufförmigen Prozeß“, sodass sich „Reihenfolge als auch Einflußrichtungen der Determinanten herausbilden und ändern“ können (Handel 2000, 66).

Selbstverständlich sind Werte, Einstellungen, Interessen und inhaltliche Vorlieben einer Person letztlich auf Ebene des Individuums anzusiedeln. Ebenso ist die Soziodemographie, allen voran Alter, Geschlecht, Bildung, Einkommen und sozialer Status Merkmale des Individuums. Sie gelten als vielversprechende individuelle Faktoren der Mediennutzung und im Kontext von Publikumsfragmentierung gesehen werden (Handel 2000, 60f., 67; Webster & Ksiazek 2012, 45; vgl. Schweiger 2007, 269-278; Peiser & Jandura 2015; Peng & Yang 2022, 2273-2276) und stehen bei dem hier vorliegenden Dissertationsprojekt gemeinsam mit inhaltlichen Präferenzen, welche anhand der genutzten inhaltlichen Genres ausgemacht werden, im Fokus der Einzelstudien.

Gesellschaftliche Wandlungsprozess wie die Individualisierung, die Sozialisation oder auch die Zugehörigkeit zu sozialen Gruppen und Identitäten können vor allem als Kontexthypothesen dienen, sodass eine theoretische Verbindung von den eher gesellschaftlichen Kontextfaktoren und dem Individuum, dem Akteur, hergestellt werden kann. So werden beispielweise geschlechterspezifische Unterschiede bei der Mediennutzung, wie divergierende inhaltliche Präferenzen oder die Nutzungsintensität, u.a. auf die Sozialisation von Mädchen und Jungen zurückgeführt (siehe Text 3 und 4 für weitere Ausführungen zu den Determinanten der Soziodemographie und inhaltliche Präferenzen). Diese individuellen Faktoren sind eher strukturelle Faktoren eines Akteurs, wie die Einstellungen und Geschmack, beeinflussen aber auch eher situative Faktoren, wie Bedürfnisse und Stimmungen, in denen es zu einer bestimmten Mediennutzung kommt. Diese strukturellen und situativen Faktoren spielen dann bei dem konkreten Entscheidungsprozess der Mediennutzung zusammen, wie die theoretischen Erklärungen der Mediennutzung (Uses and Gratification, Mood-Management, Selective Exposure, u.a.) veranschaulichen. Webster nennt diese Faktoren in seinem Theoriebeitrag zu Publikumsverhalten „Pull“-Faktoren der Mediennutzung und zeigt gleichzeitig auf, dass Nutzende verschiedene Strategien nutzen, um mit dem erhöhten Selektionsdruck im Netz umzugehen²¹: „As people attempt to “pull” what they want from the unimaginably vast offerings of digital media, they often rely upon user information regimes like search and recommendation systems“ (Webster 2018, 100). Dies veranschaulicht noch einmal die Dualität von Markt und Publikum bei der Medienauswahl und folglich im Kontext der Publikumsfragmentierung online (vgl. Webster & Ksiazek 2012; Taneja et al. 2012; Stark 2013).

²¹ So bezeichnen Webster und Ksiazek Medienrepertoires als Coping-Strategie von Nutzenden (Webster & Ksiazek 2012, 45) und auch Stark (2013, 207, 214) nennt die spezialisierte Mediennutzung eine Selektionsstrategie, welche sich beispielsweise anhand von Gewohnheiten, Präferenzen, Empfehlungssystemen, u.ä. orientieren kann.

2.5.2.2 Determinanten des Medienmarktes: Faktoren der individuellen Mediennutzung auf Makroebene

Ebenfalls als Determinanten von Fragmentierung gelten Merkmale des Medienmarktes und -angebots. Zum einen Medientechnik und Gerätespezifika, wie eben die eher interaktiven Rückkanäle (vgl. Handel 2000, 67), die im Netz möglich sind wie bei Sozialen Medien, oder aber die Medienrezeption über Desktop-Geräte im Unterschied zu mobilen Endgeräten wie Smartphones oder Tablets. So wird beispielsweise eine eher oberflächliche Rezeption und ein „Scannen“ der Inhalte und Newsfeeds über mobile Endgeräte wie Smartphones vermutet. Mobile Endgeräte ermöglichen die Rezeption von Onlinemedien in völlig anderen situativen Kontexten als feststehende PCs (vgl. Görland et al. 2017, 248f., Struckmann & Karnowski 2016, 309) und Nutzungspraktiken unterscheiden sich (Karnowski 2019, 87), was neben der Medienselektion auch Einfluss auf die Medienwahrnehmung haben kann. Des Weiteren wird das verfügbare Medienangebot durch den Markt bestimmt: die Ausdifferenzierung von Online-Angeboten (Titelvermehrung) und Spezialisierung dieser in Themen und zugeschnitten auf bestimmte Zielgruppen, werden als sehr einflussreiche Determinanten von Publikumsfragmentierung gesehen (vgl. Handel 2000, 67; Benoit & Billings 2020). Häufig wird in eben dieser Ausdifferenzierung und Spezialisierung in Verbindung mit dem gesellschaftlichen Phänomen der Individualisierung die Ursache von Publikumsfragmentierung online gesehen (siehe u.a. Webster & Ksiazek 2012; Stark 2013, 206f.; Handel 2000, 57ff.; Stunstein 2007; Pariser 2011; vgl. McQuail 2010, 90, 432, 444; Schweiger 2007, S. 46f., 287, 302ff.; Jarren 2000, 34). Außerdem sind sogenannte Aufmerksamkeitsaktionen durch Medien eine weitere Gruppe von Faktoren der Publikumsfragmentierung (Handel 2000, 67). Diese werden auch als „Push“-Faktoren des Medienmarktes bezeichnet (Webster 2014; Webster 2018, 100ff.). Gemeint sind zum Beispiel Werbemaßnahmen, Empfehlungssysteme, Personalisierung von Inhalten und Werbung, „guiding people’s choices“ und „to orchestrate much more aggressive ways to construct audiences (...) who might not otherwise be interested“ (Webster 2018, 99). Sie haben also marktseitig zum Ziel die Aufmerksamkeit des Publikums auf die eigenen Medienangebote und -inhalte zu lenken (vgl. Habermas 2021, 494). Denn Unique User, Page Impressions und ein zielgruppenspezifisches Publikum sind auf dem Online-Werbemarkt die Währung durch die gute Preise für Werbeflächen erzielt werden können (vgl. Taneja 2020) und die produzierten Inhalte (re-)finanziert.

Als Kontexthypothese zum Individuum dienen genau diese Marketing-Mechanismen: Durch Targeting, Personalisierung und Spezialisierung von Medienangeboten online werden gezielt inhaltliche Interessen und Präferenzen von Nutzende angesprochen und in der „Programmgestaltung“ berücksichtigt. Das heißt Nutzungskaskaden werden durch Empfehlungssysteme und Verlinkungen von Anbietern gleitet, um Nutzende im eigenen

Kosmos des (Werbe-)Angebots zu halten. Genauso soll die User Experience gesteigert werden, damit Nutzende wieder zum Angebot zurückkehren (vgl. bspw. Taneja 2020). Das bedeutet, Medienangebote werden so spezialisiert, konzipiert und optimiert, dass sie zum einen die Bedürfnisse, Einstellungen, Präferenzen und Interessen von Nutzer(gruppen) befriedigen. Zum anderen, um ein positives Medienerlebnis für die angesprochenen Nutzende zu schaffen und damit Bedürfnisse und Präferenzen für die nächste Nutzungssituation zu kreieren bzw. zu bestärken, sodass Nutzende zu Angeboten zurückkehren (vgl. Stark et al. 2021a, 304f.). Dies hat, so eine verbreitete These, eine immer größere Ausdifferenzierung und Spezialisierung zugeschnitten auf Publikumssegmente zur Folge (McQuail 2010, 444; Webster & Ksiazek 2012, 48ff.; Schweiger 2007, 303f.; Handel 2000, 57f.; Tewksbury & Rittenberg 2009, 195; Tewksbury 2005, 332ff.). Diese Publikumssegmente folgen häufig den Marketing- und Werbe-relevanten Zielgruppen, die in der Regel auf den gängigen soziodemographischen Merkmalen, Alter, Geschlecht, Bildung und Einkommen (als Indikator für die Kaufkraft) potenzieller Kund:innen basieren (Schweiger 2007, 270ff.; Webster 2005, 369; vgl. Stark & Kist 2020, 1145; Tewksbury 2005; Brentel, *forthcoming* [Text 4]). So soll Publikumsfragmentierung online durch privatwirtschaftliche Interessen des kommerziell orientierten Medienmarktes gefördert werden: Damit Werbenden ermöglicht wird “to profile individuals [...] creating specialized audiences for different products, thereby ‘segmenting’ what was hitherto a ‘whole’ national audience (Anderson, 2008; Turow, 1998, 2011)” (Mukerjee, 2021a, p. 3).

2.5.3 Ebenen der Publikumsfragmentierung

Die Analyse und Bestimmung von Mediennutzungsmustern kann auf unterschiedlichen Untersuchungsebenen erfolgen (siehe Handel 2000, 33-42). So unterscheidet Handel insgesamt zwischen neun Untersuchungsebenen. Bei den ersten fünf Ebenen, Kommunikationskanäle, Mediengattungen, Rundfunksender und Printtitel, Sendungen bzw. Sendungsausschnitte sowie Wahrnehmung handelt es sich, „um eine Differenzierung des Betrachtungsgegenstandes bei der Fragmentierung und um eine Annäherung an das tatsächlich rezipierte“ (Handel 2000, 35). Die übrigen vier Ebenen – Raumdimension, Zeitdimension, Inhalte und Genre sowie das Rundfunksystem – können als übergeordnete Betrachtungsebenen verstanden werden, welche auf jede der zuvor genannten Differenzierungsgrade angewendet werden können (Handel 2000, 35). Nachfolgend werden die neun Ebenen der Publikumsfragmentierung nach Handel (2000, 33-42) kurz skizziert und auf den Onlinekontext übertragen, um den Untersuchungsgegenstand der Mediennutzungsmuster für die hier vor liegende Arbeit systematisch weiter einzugrenzen und

den Schwerpunkt der Dissertation in den Gesamtkontext der Publikumsfragmentierungsforschung einordnen zu können:

- *Kommunikationskanäle*: damit ist das „Sehen“, „Hören“ und „Lesen“ gemeint (Handel 2000, 34). Durch die steigende Medienkonvergenz online (Schweiger 2007, 343ff., 41; Stark & Kist 2020, 1139f.; Neuberger 2009, 23f.) ist diese Differenzierung für den Online-Bereich weniger zielführend – denn beim Internet handelt es sich um einen Multimedia-Kanal (Handel 2000, 34).
- *Mediengattungen*: sind definiert als TV, Hörfunk, Zeitschrift, Zeitung und Internet – bei der hier vorliegenden Arbeit wird die Mediengattung Internet untersucht.
- *Rundfunksender und Printtitel*: umfassen konkrete Gesamtangebote; einzelner Mediengattungen. Im Onlinekontext entspricht dies dem Onlineangebot einer Medienmarke, also dem was in der Regel zwischen „www.“ und „.de“, „.com“, „.org“, etc. steht. Gemeint sind also Webseiten als Gesamtes, wie zum Beispiel RTL.de, BILD.de, ZEIT.de, t-online.de, gutefrage.net, usw..
- *Sendungen, Sendungsausschnitte und -schnipsel*: Dabei handelt es sich um eine weitere, strukturelle Verfeinerung. Gemeint sind hier zum einen Sendungen oder einzelne Printausgaben, wobei in der Fragmentierungsforschung vor allem TV-Sendungen betrachtet werden und Zeitungsausgaben keine Rolle spielen (Handel 2000, 34). Das online-Pendant hierzu sind Unterseiten eines Gesamtangebots, das Einzelangebot: z.B. die Politik- oder Sport-Rubrik eines Onlineangebots, also dessen Unterseiten, wie „bild.de / politik“ oder „bild.de / sport“ sowie weitere Unterseitenstrukturen wie „bild.de / sport / fußball“.

Zudem kann eine weitere Differenzierung im Onlinekontext vorgenommen werden, die etwa mit einem Sendungsausschnitt oder Schnipsel im TV-Kontext vergleichbar ist: konkrete Medienprodukte wie einzelne Artikel, Videos, Meldungen o.ä.. Die Nutzung dieser sind Online lediglich durch die Erhebung einzelner URLs oder mit Hilfe eines Eye-Trackings messbar. Dies wäre für die Analyse einzelner Nutzungsepisoden denkbar, ist für die Erfassung von Mediennutzungsmuster online auf Grund der Vielfalt und daraus resultierenden Datenmenge jedoch in der Breite der Onlinemediennutzung wenig sinnvoll und sollte auf Fallstudien bzw. die Betrachtung der Nutzung einzelner Angebote oder spezieller Nutzergruppen begrenzt werden. Folglich spielt der Differenzierungsgrad von konkreten Medienprodukten für die vorliegende Arbeit keine Rolle.

- *Wahrnehmung*: hierbei handelt es sich weniger um die Wahrnehmung hinsichtlich einer Medienwirkung als um die Aufmerksamkeit, mit der Medien oder konkrete Medieninhalte rezipiert werden. Denn durch unterschiedlich aufmerksame Nutzung, die beispielsweise durch das Involvement von Individuen determiniert ist, „kann nicht

davon ausgegangen werden, daß von allen dasselbe wahrgenommen und aufmerksam verfolgt wird“ (Handel 2000, 35). In der hier vorliegenden Arbeit wird diese Ebene nicht weiter berücksichtigt. Denn für den Fragmentierungsbegriff dieser Dissertation sind nicht überschneidende Themenagenden von Individuen oder Nutzergruppen, sondern die (nicht) Überschneidung von Publika bzw. die Überlappung von Mediennutzungsmustern als beobachtetes Nutzungsverhalten entscheidend.

- *Raumdimension*: bildet die erste übergeordnete Betrachtungsebene und beschreibt keine weitere (strukturelle) Differenzierung der Untersuchungsebenen der Publikumsfragmentierung. Für das Medium Internet spielt die Raumdimension keine große Rolle mehr (Stark & Kist 2020, 1138; Neuberger 2009, 22-26), da die Distribution von Medienangeboten im Netz global und unabhängig von geografischen Grenzen funktioniert. Die Raumdimension kann aber helfen Onlinenutzung weiter einzugrenzen, z.B. bei einem Vergleich der Nutzung lokaler vs. regionaler Onlineangebote (siehe Olsen 2020).
- *Zeitdimension*, welche ebenfalls im Internetzeitalter an Relevanz verloren hat (Stark & Kist 2020, 1138; Neuberger 2009, 22-26), da Medienangebote „on-demand“ genutzt werden können, unabhängig von Raum und Zeit²². So kann Online „die Gemeinsamkeit der Mediennutzung mit den anderen Zuschauern zeitversetzt wieder hergestellt und der Fragmentierungsgrad in einem Zeitrahmen dadurch verkleinert“ (Handel 2000, 39) statt, wie beispielsweise bei TV, dadurch erst hergestellt werden.

Entsprechend spielen bei der Analyse von Publikumsfragmentierung online die Referenzebenen der Raum- und Zeitdimension (siehe Handel 2000, 36-39) eine untergeordnete Rolle (vgl. Webster 2014, u.a. 4; Rau & Stier 2019, 403; Stark 2013, 214) bzw. werden nicht mehr als „fragmentierenden“ Faktoren gesehen und hier eher ausgeklammert.

- *Inhalte und Genre*: Diese Fragmentierungsebene beschreibt unterschiedliche bzw. überlappende Nutzung anhand verschiedener thematischer Inhalte und Genre, die den Interessen eines Individuums entsprechen. In der Konsequenz wenden sich die einen Nutzenden „anderen Genres oder inhaltlichen Bereichen zu[wendet] als die anderen“ (Handel 2000, 35). Die Folge sind unterschiedliche Themenverständnisse des Gesamtpublikums „ohne gemeinsame thematische Anknüpfungspunkte, [sodass] der zu Integration notwendige Diskurs in der Gesellschaft abbricht“ (Handel 2000, 35). Im Kontext des Internets wird diese Ebene häufig als entscheidender Faktor von Fragmentierung gesehen (vgl. u.a. Sunstein 2001, 2007, 2017; Pariser 2011; Mukerjee

²² Dennoch kann die Zeitkomponente bei der Betrachtung bestimmter Medienereignisse (Live-Events wie die Fußball-WM, Wahlen, u.a.) auch im Netz eine Rolle spielen; da in der vorliegenden Dissertation jedoch keine spezifischen (Medien-)Ereignisse betrachtet werden, ist die Zeitdimension für die Betrachtung der Publikumsfragmentierung ausgeklammert.

2021a), insbesondere in Bezug auf eine stärkere Ausdifferenzierung von Onlinemedienangeboten. So spielt diese Betrachtungsebene in der hier vorliegenden Arbeit eine wichtige Rolle bei der Analyse der Publikumsfragmentierung online.

- *Rundfunksystem*: beschreibt die Betrachtung von öffentlich-rechtlichen Medienangeboten in Abgrenzung zu privatrechtlichen, kommerziellen Medienangeboten oder die Kombination von beidem. Diese übergeordnete Betrachtungsebene hat auch im Onlinekontext weiterhin Bestand – Vergleiche der Nutzung zwischen den Rundfunksystemen im Netz fehlen bislang jedoch.

Die Betrachtung von Publikumsfragmentierung wird bei der hier vorliegenden Arbeit auf kommerzielle Onlineangebote begrenzt. Denn gerade in den privatwirtschaftlichen Marktmechanismen werden in der Fragmentierungsforschung Gefahren für eine ansteigende Fragmentierung sowie Segmentierung des Gesamtpublikums gesehen (vgl. Schweiger 2007, 303f.). Eine steigende Angebotsfragmentierung, die einer Publikumsfragmentierung tendenziell vorausgeht (vgl. McQuail 1997, 138; McQuail 2010, 444ff.; Tewksbury 2005, 332; Babić & Jandura 2017, 105), resultiert Online vor allem aus den gesunkenen Marktbarrieren sowie dem damit verbundenen Wettbewerbsdruck möglichst viel Aufmerksamkeit, also Nutzung vorwiegend gemessen als Klicks oder Unique Users (Hofsäss et al. 2020), zu generieren. Denn produzierte (online) Inhalte werden bei kommerziellen Medien meist durch Werbung finanziert (vgl. McQuail 2010, S. 402; Stark & Kist 2020, 1149; Webster & Ksiazek 2012, 40). Eine (inhaltliche) Spezialisierung für Publikumssegmente wird, gerade bei privatwirtschaftlichen Medienangeboten, als Folge gesehen, um auf dem Werbemarkt attraktiv zu sein (vgl. McQuail 2010, S. 224; Schweiger 2007, 303f.; Stark et al. 2021a, 304f.; Taneja 2020). Eine Ausdifferenzierung und (inhaltliche) Spezialisierung online als treibender Faktor einer Publikumsfragmentierung wird also vor allem im kommerziellen, privatwirtschaftlichen Mediensektor erwartet (vgl. Bächler 2017, 321).

In öffentlich-rechtlichen Medienangeboten hingegen wird eher eine integrative Funktion als eine spaltende gesehen. Zum einen haben sie den gesetzlichen Auftrag ein Medienangebot für die gesamte Gesellschaft zu schaffen. Zum anderen haben Studien weltweit gezeigt, dass sogenannte public-service Medien einen hohen Überschneidungsgrad in der Nutzung haben (Fletcher & Nielsen 2017, 483). Dies ist in Deutschland, im TV-Kontext, in besonderem Maße der Fall. So gelten die Nachrichtenangebote von ARD und ZDF als integrative Insel der Mediennutzung und weisen den höchsten Grad an Publikumsüberschneidung auf (Handel 2000, 53; Mahrt 2019, 67). In Deutschland wird die Repräsentation von öffentlich-rechtlichen Angeboten online im internationalen Vergleich jedoch eher als gering repräsentiert beurteilt (Fletcher & Nielsen 2017, 483). Online bieten die öffentlich-rechtlichen TV-Sender Mediatheken an (ARD und ZDF Mediathek), auf denen Sendungen bzw. Spartensender „live“

oder „on-demand“ rezipiert werden können. Außerdem werden weiterführende Informationen begleitend zu Sendern und Sendungen angeboten (z.B. tagesschau.de, phoenix.de, deutschlandfunk.de) sowie Social-Media-Marken aufgebaut (z.B. funk, mithinkx). Die Nutzung dieser Online-Angebote, gemessen an Reichweiten, wird jährlich im Rahmen der Studien-Massenkommunikation veröffentlicht und zum Teil in der akademischen Forschung weiter analysiert. Eine Erweiterung der Betrachtung von Publikumsfragmentierung um den öffentlich-rechtlichen Sektor wäre also aufbauen auf die hier vorliegende Dissertation denkbar. Doch vor dem Hintergrund, dass im privatwirtschaftlichen, kommerziellen Teil des Rundfunksystems ein erhöhtes Fragmentierungspotenzial gesehen wird, liegt der Fokus in dieser Dissertation auf der Nutzung bzw. Mediennutzungsmuster kommerzieller, Onlineangebote.

„Je feiner und detaillierter die Nutzung von Medienbotschaften betrachtet wird (...), desto exakter erfaßt man den potentiellen Medieninput für spätere gemeinsame Gespräche“ (Handel 2000, 35) und damit den Grad der Fragmentierung. Die Ebene der Einzelangebote als Online-Pendant zu *Sendungen* wird hier daher als notwendige Differenzierung für die empirischen Analysen dieses Dissertationsprojektes erachtet. Denn nur so kann die inhaltliche (Aus-)Differenzierung und Vielfalt (im Sinne von Vielzahl) des kommerziellen Onlinemedienmarktes abgebildet werden, welche vor allem als treibender Faktor von Publikumsfragmentierung im Onlinekontext gesehen wird: Durch den „growth in digital media (...) patterns of consumption become more widely distributed“ (Webster & Ksiazek 2012, 1), da mit der größeren verfügbaren Angebotsauswahl der Selektionsdruck und Freiheitsgrad in der Medienauswahl der Individuen steigt (Stark 2013, 214; Webster & Ksiazek 2012, 1). Eine Aggregation auf Ebene der Gesamtangebote oder gar der Mediengattung Internet als Ganzes würde unterschiedliche Nutzungsmuster von Onlineangeboten auf Grund der strukturellen Unschärfe verwischen (vgl. Mukerjee 2021a, 9; Webster & Ksiazek 2012, 51) und den Mehrwert einer singulären Betrachtung der Onlinemediennutzung obsolet erscheinen lassen. Gleichzeitig wird die Ebene der Einzelangebote als hinreichend detailscharf zur Analyse von Publikumsfragmentierung online gesehen. Eine weitere Differenzierung einzelner Medienprodukte wird hingegen für eine erste, möglichst umfassende Exploration der Onlinenutzung von kommerziellen Massenmedien, auch auf Grund der zu erwartenden Datenmenge und damit verbundenen Notwendigkeit zur Aggregation, nicht für zielführend erachtet²³.

²³ Zudem birgt eine Erfassung der Mediennutzung auf diesem Detailgrad gerade in Verbindung mit weiteren Daten der Nutzenden beziehungsweise Befragten enorme Datenschutzprobleme nach aktueller Rechtslage. Dies macht solche Analysen ohnehin schwierig beziehungsweise würde sie auf diesem Detailgrad stark limitieren, um die Anonymität und damit den Datenschutz hinsichtlich der Befragten zu wahren.

Weiterhin wird für die empirischen Analysen der Schwerpunkt neben der Ebene der Einzelangebote, als Online-Pendant zu Sendungen, auf die der Inhalte und Genre gelegt. Denn für Publikumsfragmentierung im Onlinekontext ist die die (Aus-)Differenzierung und (thematische) Spezialisierung der Online-Angebote im Verdacht Fragmentierungstendenzen zu fördern (Webster & Ksiazek 2012, 1; Stark 2013, 206f.; Mukerjee 2021a, 3; Handel 2000, 67; Tewksbury & Rittenberg 2009). Nicht zu Letzt, weil Nutzende durch den gestiegenen Freiheitsgrad in der Medienauswahl der Selektive-Exposure-Theorie zu Folge nur noch Medienangebote nutzen könnten, die ihren Interessen und Meinungen entsprechen (vgl. u.a. Rau & Stier 2019; Nelson & Webster 2017; Steppat et al. 2022) – die Grundlage der Filterblasen- und Echokammern-These (Pariser 2011; Sunstein 2007).

2.6 Untersuchungsgegenstand der Dissertation: kommerzieller, massenmediale Einzelangebote auf Ebene inhaltlicher Gerne

Bei der hier vorliegenden Dissertation stehen Mediennutzungsmuster im Sinne der Publikumsfragmentierung im Mittelpunkt. Folglich basieren die empirischen Analysen auf der Komponente der Mediennutzungsmuster, welche sich wiederum aus der Messung einzelner Nutzungsepisoden ergibt. Wie bereits oben skizziert, ist die Arbeit in der Nutzungsforschung zu verorten. Dementsprechend werden zwar Wirkungsbeziehungen von den Einflussfaktoren Individuum, soziales Umfeld und Gesellschaft mit Mustern der Mediennutzung im Sinne der Fragmentierungsthese angenommen, jedoch nicht empirisch überprüft. Auch stehen eher langfristige Prädispositionen des Individuums im Fokus der hier vorliegenden Analysen, wie die allgemeinen Merkmale Geschlecht, Alter und Bildung (Schweiger 2007, 269-278) – ebenso wie das Einkommen, welches die Ressourcenausstattung und sozio-ökonomischen Status einer Person beeinflusst (siehe Kapitel 2.5.2).

Zusammenfassend lässt sich der Schwerpunkt für den Untersuchungsgegenstand dieses Dissertationsprojektes wie folgt definieren: Als Komponente der Mediennutzung bildet aufbauend auf das Zwiebelmodell nach Schweiger (2007, 30) die Mediennutzungsmuster den Untersuchungsgegenstand. Für die Betrachtungsebene werden die empirischen Analysen auf die Mediengattung Internet limitiert. Durch diese Schwerpunktsetzung, soll eine umfassende, explorative Vorgehensweise ermöglicht werden, um Einblicke mit detailtiefe in Online-Mediennutzungsmuster zu erlangen. Diese fehlen bisher in der Forschungsliteratur und die Notwendigkeit von detaillierteren Betrachtungsebenen wurde bereits gefordert (Webster & Ksiazek 2012, 51; Mukerjee 2021a, 9; Stark & Kist 2020, 1139). Dabei interessieren nur Medienangebote, deren Geschäftsmodell auf der Produktion und Bereitstellung von Inhalten

basiert (in Abgrenzung zu Social-Media Plattformen, E-Commerce, Gaming, etc.²⁴). Denn insbesondere für Deutschland sind Studien zu Onlinenutzungsmustern außerhalb der Social-Media Nutzung selten (Stark 2013; Mahrt 2019). Mediennutzungsmuster im Kontext von Publikumsfragmentierung online von Medien mit einem Content-Geschäftsmodell sind jedoch von höchster Relevanz, da sie die Information, den Inhalt, für die Erfüllung der Integrationsfunktion von Massenmedien bereitstellen.

Für die nötige Detailschärfe, den Differenzierungsgrad von Fragmentierung, liegt der Schwerpunkt auf Ebene der Einzelangebote (dem Online-Pendant zu Sendungen). Um auch einer inhaltlichen Differenzierung des Onlinemedienmarktes und gegebenenfalls daraus resultierenden, inhaltlich determinierten Nutzungsmustern online gerecht zu werden, werden ebenfalls die inhaltlichen Genre der Einzelangebote in den Analysen berücksichtigt und bilden einen Schwerpunkt. Vor allem politische Informationsnutzung im Kontext zur allgemeinen Onlinemediennutzung ist von besonderem Interesse für die medial-vermittelte Integration in demokratischen Systemen wie Deutschland (Jarren 2014, 325; McNair 2003, 20f.) und stehen im Verdacht durch fragmentierte Nutzung gekennzeichnet zu sein (vgl. Babić & Jandura 2017, 108; Bächler 2017, 320f., 327; Barnidge et al. 2021, 902; Dohle, Jandura & Vowe 2014, 425; Grotz & Schroeder 2021, 219; Steppat et al. 2022; van Aelst et al. 2017, 4f.). Desweiteren werden ausschließlich kommerzielle Angebote, also der privatwirtschaftliche Rundfunk, für die Analysen betrachten. Zum einen determiniert durch die Datenverfügbarkeit. Zum anderen wird im privatwirtschaftlichen, kommerziellen Medienmarkt auf Grund der wirtschaftlichen Interessen, angeheizt durch den Werbemarkt, eine stärkere Angebotsfragmentierung gesehen als bei öffentlich-rechtlichen Angeboten (vgl. Grotz & Schroeder 2021, 219). Das Fragmentierungspotenzial bei Mediennutzungsmustern von kommerziellen Onlineangeboten ist demnach höher (vgl. Handel 2000, 36; Bächler 2017, 321) und folglich Schwerpunkt der Analysen hier vorliegenden Dissertation zur Publikumsfragmentierung online in Deutschland.

²⁴ siehe hierzu Wirtz (2018, 307-309): Er definiert fünf Geschäftsmodelle: Connection, (E-)Commerce, Games, Context und Content. Diese Typisierung gehört zu den meist genutzten der Medienökonomie.

3 Forschungsprogramm

Ziel des hier vorliegenden Dissertationsprojektes ist es den ersten Teil des Fragmentierungsprozesses (vgl. Handel 2000, 12), die Voraussetzung für geteilte Themen und Wissensbestände (vgl. Vlašić 2004, 70), – die (gemeinsame) Mediennutzung – für kommerzielle, massenmediale Onlinemedien in Deutschland zu untersuchen. Denn die Fragmentierungsthese ist für Onlinemedien zwar weit verbreitet, es fehlt aber an empirischer Evidenz und Studien zur massenmedialen²⁵ Onlinemediennutzung bzw. -Nutzungsmustern in Deutschland (vgl. Stark 2013, 200, 212f.; Stark et al. 2021a; Rau & Stier 2019). Daher werden anhand eines explorativen Vorgehens Nutzungsmuster online auf Gemeinsamkeiten und Unterschiede – in der Nutzung wie auch bei den Nutzenden – analysiert, um ein möglichst umfassendes Bild (gemeinsam) genutzter, kommerzieller, massenmedialer Onlinemedien zu erhalten. Schließlich stehen kommerzielle, massenmediale Onlineangebote auf Grund privatwirtschaftlicher Marktmechanismen wie Advertising, Targeting und Personalisierung am meisten unter Verdacht Publikumsfragmentierung zu fördern und fragmentiert genutzt zu werden.

Dies erfolgt mit dem Ziel, Erkenntnisse darüber zu gewinnen, ob und inwiefern eine (Publikums-)Fragmentierung in der Onlinemediennutzung in Deutschland zu beobachten ist. Mit dem Zweck und vor dem Hintergrund der Integrationsfunktion von Massenmedien Einblicke zu erlangen, ob die Voraussetzung der (ausreichend) geteilten Nutzung online gegeben ist, sodass über massenmediale, kommerzielle Onlineangebote geteilte Themen und eine Wissensbasis existieren und sich politische Öffentlichkeit konstituieren kann. Dabei ist es auch das Ziel durch die Beschreibung von Nutzenden und Onlineangeboten mit bestimmten Nutzungsmustern Hinweise auf Determinanten von Publikumsfragmentierung online zu finden. Gleichzeitig soll sich durch die Exploration von überlappenden Onlinenutzungsmustern vs. nicht geteilter Onlinenutzung der Frage angenähert werden, „WER eigentlich fragmentiert ist“ (Handel 2000, 31) und ob Individuen gleichermaßen von Publikumsfragmentierung betroffen sind. Hier spielt insbesondere die Nutzung politischer Information eine Rolle, welche zur Erfüllung der Informationsfunktion durch Massenmedien (online) zentral ist und der bei ungleicher Nutzung die Folge eines „democratic divides“ nachgesagt wird (Norris 2001).

3.1 Identifikation und Erschließung einer geeigneten Datenbasis

Zur Umsetzung dieses Vorhabens musste zunächst eine geeignete Datenbasis geschaffen werden. Denn Daten zur Onlinemediennutzung mit dem 1) nötigen Detailgrad der gemessenen Onlinemedienangebote zur Untersuchung von Publikumsfragmentierung online und 2)

²⁵ In Abgrenzung zu Social Media Nutzung.

ausreichender Anzahl an Medienangeboten sowie 3) an Personen für repräsentative Analysen und Aussagen in Bezug auf Deutschland fehlten bislang. Folglich sieht das Forschungsprogramm der hier vorliegenden Dissertation zunächst die Schaffung einer geeigneten Datenbasis vor mit der leitenden Forschungsfrage:

Wie kann eine geeignete Datenbasis zur Untersuchung von Publikumsfragmentierung online aussehen und gewonnen werden?

Diese wird durch die erste Einzelschrift (Text 1) des hier vorliegenden Kumulus beantwortet. In Text 1 wird die Datenquelle, die Media-Analyse *IntermediaPlus* für die Jahre 2014 bis 2016, erläutert, der Mehrwert für das Fach sowie weitere mögliche Anwendungsbereiche genannt und die Aufbereitung zur Nachnutzung der dem Dissertationsprojekt zu Grunde liegenden Datenbasis *IntermediaPlus (2014-2016)* kurz beschrieben. Außerdem sind die aufbereiteten Daten im DBK-Datenkatalog bei GESIS publiziert und archiviert (siehe ZA5769).

Eine große Rolle bei der Erschließung der Datenquelle der Media-Analyse *IntermediaPlus* spielte die nachhaltige, den FAIR-Standards²⁶ entsprechende, Aufbereitung dieser. Denn es ist klar, dass das hier vorliegende Dissertationsprojekt das Fragmentierungsphänomen nicht in Gänze untersuchen kann. Es ist lediglich ein erster Schritt zur Exploration des Standes und der Beschreibung von Publikumsfragmentierung online, definiert als (überlappende) Muster der Mediennutzung. Sodass die Voraussetzung der Integrationsfunktion durch Massenmedien und die erste Ebene eines Fragmentierungsprozesses in Deutschland besser evaluiert werden kann. Folglich wurde die Datenbasis nicht nur für das hier vorliegende Forschungsprojekt aufbereitet, sondern auch für eine breite Nachnutzung von wissenschaftlichen Studien mit dem Ziel das Fragmentierungsphänomen im Kontext von Onlinemedien zu ergründen. So bietet die aufbereitete *IntermediaPlus (2014-2016)* durch die enthaltenden Daten Anknüpfungspunkte für weiterführende Forschung von Fragmentierungsprozessen, aber auch für andere Studien rund um crossmediale Mediennutzung. Demnach sind die aufbereiteten Daten der *MA IntermediaPlus (2014-2016)* interoperabel und erweiterbar: Sowohl zeitlich für eine längsschnittliche Betrachtung und beispielsweise ein Monitoring des Publikumsfragmentierungsprozesses online, als auch für weitere Datenbestände, um zum Beispiel den Einfluss von Werten und Einstellungen auf die (geteilte) Mediennutzung zu erforschen. Um dies zu erreichen lautete die leitende Forschungsfrage zur Erschließung und Aufbereitung der Daten:

²⁶ FAIR steht dabei für Findable, Accessible, Interoperable und Reusable (Wilkinson et al. 2016; Mons et al. 2017; siehe Text 1 und 2).

Wie kann die Datenquelle der Media-Analyse IntermediaPlus nachhaltig, also interoperabel, aufbereitet und für akademische Forschung und Nach-Nutzung auffindbar sowie zugänglich gemacht werden?

Im Rahmen der zweiten Einzelschrift (Text 2) wird daher ein Datenmanagementsystem für die Media-Analyse *IntermediaPlus* als unstrukturierte big-data Datenquelle beschrieben. Die Entwicklung dieses Datenmanagementsystem erfolgte als Teil dieses Dissertationsprojektes unter Einhaltung der FAIR-Prinzipien. Hierzu gibt es bislang im Fach der Kommunikations- und Medienwissenschaft kein vergleichbares Vorhaben, weshalb zunächst ein Datenmanagement den FAIR-Prinzipien folgend und Herausforderungen von big-data entsprechend innoviert werden musste. Das entwickelte Datenmanagementsystem beinhaltet die Konzeptionelle Aufbereitung der *IntermediaPlus* Daten für die Jahre 2014, 2015 und 2016, mit Implikationen für die Erschließung anderer big-data-Quellen, ebenso wie die Erarbeitung einer Prozesskette mit technischen Schnittstellen zur lückenlosen, halb-automatisierten Datenaufbereitung sowie ein Datenbanksystem, welches ein effizientes Arbeiten mit der big-data Datenquelle erlaubt. Anhand der Begleitpublikationen zur Datenaufbereitung und -harmonisierung der Media-Analyse *IntermediaPlus* bei SoWiDataNet des GESIS Leibnitz-Instituts in Köln (siehe Brentel et al. 2021; Brentel 2021a; Brentel 2021b; Brentel & Fortmann 2021 [Begleitpublikation 1, 2, 3 und 4]) kann der Prozess nicht nur genau nachvollzogen und repliziert werden, sondern auch für andere big-data Datenquellen nachgeahmt werden. Dadurch sollen die Nachnutzung und die Erweiterung der Datenbasis durch das Fach angeregt werden und ein Beitrag zur Debatte im Fach bezüglich eines FAIRen Datenmanagements mit einem best-practice Anwendungsfall geleistet werden.

3.2 Empirische Studien zu Publikumsfragmentierung online

Auf Basis der nachhaltigen und interoperabel aufbereiteten Daten der *IntermediaPlus* (2014-2016) folgen im zweiten Teil des hier vorliegenden Forschungsprojektes die inhaltlichen Fragestellungen und die daraus resultierenden, empirischen Einzelstudien des Kumulus. Die übergeordnete, leitende Forschungsfrage ist hier:

Inwiefern gibt es Überlappungen in den (Teil-)Publika bzw. überschneidende Mediennutzungsmuster online in Deutschland? Und, wie sehen Onlinenutzungsmuster aus?

Ganz im Sinne des methodologischen Individualismus des Mirko-Makro-Modells (Greve et al. 2008, 4), wird mit der ersten empirischen Einzelstudie dieses Dissertationsprojektes (Text 3) die Nutzung politischer Information auf Mikroebene in den Blick genommen, um das Makrophänomen der Publikumsfragmentierung online zu ergründen: Denn insbesondere die

(geteilte) Nutzung von politischer Information wird als substanziell zur Herstellung politischer Öffentlichkeit in einem demokratischen System wie Deutschland erachtet (vgl. McNair 2003, 20f.; Weiß & Jandura 2017, 14; Jarren 2014, 317, 320f.; Jarren 2000, 22ff.) und die damit einhergehenden Integrationsfunktionen von Massenmedien. Entsprechend wird in einem ersten Analyse-Schritt ein besonderes Augenmerk auf die (geteilte) Nutzung politischer Information gelegt.

Die Analyse der politischen Informationsnutzung online folgt dabei den Überlegungen aus der Digital Divide Forschung, um Hinweise für ungleiche Onlinenutzungsmuster und damit eine mögliche Publikumsfragmentierung online für die Analyse zu erhalten. Die Digital Divide Forschung beschäftigt sich mit einer Ungleichheit in der Onlinenutzung und damit einhergehenden abweichenden, nicht überlappenden Mediennutzungsmustern auf Grund unterschiedlicher Ressourcenausstattung der Nutzenden, welche durch die Strukturen der Gesellschaft beeinflusst sind. Basierend auf der Knowledge Gap Theorie ist die Grundannahme der Digital Divide Forschung, "that the information flow is not normally homogeneous as a consequence of the underlying social structure of society" (Bonfadelli 2002, 68). Das heißt, es wird von einer Homogenisierung entlang der soziodemographischen Merkmale von Nutzenden als langfristige Einflussfaktoren, determiniert durch Strukturen der Gesellschaft, in den Publika politischer Information ausgegangen. Sprich nicht-geteilte Mediennutzungsmuster online vermutet. Demnach nutzen Menschen mit einer besseren Ressourcenausstattung, sowohl finanziell als auch kognitiv, mehr politische Information und tun dies gerade online effektiver als Personen mit geringeren Ressourcen, so die These und Rückschlüsse bisheriger empirischer Studien zum Digital Divide (vgl. Zillien & Hargittai 2009, 288; Ragnedda, Ruiu & Addeo 2020). Dabei steht die politische Informationsnutzung in Abgrenzung und im Vergleich zu unterhaltungsorientierter Nutzung von Medien im Fokus (vgl. Bonfadelli 2002). Als Folge dieser ungleichen Nutzung auf Mikroebene wird eine Verstärkung der bestehenden Wissensklüfte der Nutzenden zwischen sozialen Gruppen – sprich unterschiedliches Wissen – vermutet (Tichenor et al. 1970), welches letztlich zu unterschiedlicher Beteiligung an politischer Öffentlichkeit und politischer Partizipation, auf der demokratische Systeme wie Deutschland basieren, führt. Ein „democratic divide“, die Ausgrenzung weniger ressourcenstarker sozialer Gruppen, wird als Resultat befürchtet (vgl. Norris 2001). Das heißt, aus nicht-geteilten Mediennutzungsmustern online wird, wie bei der klassischen Fragmentierungsthese, ein auseinanderdriften, also eine polarisierende Wirkung (demokratischer) Gesellschaften als Folge gesehen. Als besonders determinierend für online Mediennutzung werden die soziodemographischen Merkmale Bildung, Einkommen, sozio-ökonomischer Status und Alter gesehen.

Die Grundlegende Prämisse des Digital Divides ist, dass durch das Aufkommen des Internets neue, ungleiche Nutzungsmuster entstanden sind. Bislang ist jedoch nicht erforscht, inwiefern diese spezifische Form der Publikumsfragmentierung²⁷ ein Onlinephänomen ist oder aus dem Offlinekontext reproduziert wird. Es ist aber für die Evaluation und Einschätzung möglicher Folgen unterschiedlicher Nutzungsmuster, wie der Vertiefung von Wissensklüften und dem „Abhängen“ sowie schließlich der Exklusion bestimmter sozialer Gruppen in Bezug auf politische Öffentlichkeit im Sinne des Digital Divides relevant, ob es sich um eine Verschärfung bekannter Mechanismen handelt, die Reproduktion dieser oder neue (fragmentierende) Mechanismen. Folglich lautet die leidende Forschungsfrage des Einzelbeitrags in Text 3:

Inwiefern handelt es sich bei ungleicher Nutzung politischer Information entlang soziodemographischer Merkmale um reproduzierte Nutzungsmuster, die in der Forschung bereits aus dem Offline-Kontext bekannt sind und wobei handelt es sich um neue Fragmentierungsmuster der Onlinemediennutzung?

Dabei bildet die Einbettung in das soziale Umfeld, die Zugehörigkeit zu einer bestimmten sozialen Gruppe, gemessen an der Soziodemographie der Nutzenden, sowie die individuellen Fähigkeiten und Ressourcen die Determinanten für die Onlinemediennutzung und möglicher Publikumsfragmentierung online. Folglich beeinflussen die Gesellschaft und das soziale Umfeld die Nutzenden als individuelle Akteure (vgl. Schweiger's Zwiebelmodell der Mediennutzung 2007, 30f.). Im Sinne der Kontexthypothese des Mikro-Makro-Modells bestimmt das soziale Umfeld, ebenso wie die Gesellschaft, z.B. im Rahmen der Mediensozialisation, unterschiedliche Kenntnisse sowie Fähigkeiten hinsichtlich der Medienkompetenz und Medienbewertung (vgl. Schweiger 2007, 31, 297-301). Als Entscheidungsregel kann für die Medienauswahl und daraus resultierender Onlinemediennutzung (sowie die Wahrnehmung rezipierter Inhalte) beispielsweise das theoretische Konzept des „digital capital“ (siehe Ragnedda et al. 2020) dienen, um (neue) digitale Ungleichheiten in der Mediennutzung und daraus resultierende Wissensklüfte anhand der individuellen Ressourcen von Nutzenden zu erklären²⁸. Wird dieses individuelle Onlinenutzungsverhalten für bestimmte gesellschaftliche Gruppen bzw. entlang spezifischer soziodemographischer Merkmale aggregiert, können Mediennutzungsmuster dieser Gruppen

²⁷ Gemeint ist die ungleiche Nutzung von politischer Information entlang der soziodemographischen Merkmale von Nutzenden.

²⁸ Selbstverständlich spielen dabei auch weiterhin theoretische Konzepte zur Erklärung von Mediennutzung eine Rolle. Das Konzept des „Digital Capital“ wurde von Ragnedda und Kollegen (2020) im Forschungsfeld des Digital Divides in Anlehnung an Bourdieu's Arbeiten zum sozialen Kapital (1984; vgl. van Deursen & van Dijk 2014, 509f.) entwickelt. Damit handelt es sich entsprechend um einen spezifischen theoretischen Ansatz für die Digital Divide Forschung, die sich auf drei Ebenen bezieht: Zugang zum Internet, Internetnutzung und Wahrnehmung von Inhalten beziehungsweise Wirkung des Online rezipierten Inhalt hinsichtlich des Wissens (first, second and third digital divide, siehe Ragnedda et al. 2020).

als eine spezifische Form von Publikumsfragmentierung online untersucht werden. Bei der vorliegenden Einzelstudie wird dabei auch die Intensität der Nutzung, die Nutzungshäufigkeit, miteinbezogen. Denn in der Digital Divide Forschung wird davon ausgegangen, dass Nutzende, die politische Information regelmäßig nutzen, über ein höheres (politisches) Wissen verfügen und politische Information effizienter nutzen (Zillien & Hargittai 2009, 288) und entsprechende Folgen von fragmentierter Nutzung unterschiedlicher sozialer Gruppen hinsichtlich der Integrationsfunktion von Massenmedien stärker ausfallen können.

Die Nutzenden für diese erste empirische Studie beispielhaft für zwei Genre-Kategorien analysiert und auf Ungleichheiten in ihren Nutzungsmustern entlang der soziodemographischen Merkmale, welche bisher als vielversprechende Erklärungsfaktoren von Mediennutzung gelten (vgl. Schweiger 2007, 269-278; Peiser & Jandura 2015; Zillien & Hargittai 2009; van Deursen & van Dijk 2014, 511f.; Bächler 2017, 327) und damit auch Determinanten von Publikumsfragmentierung (Handel 2000, 60f., 67; Webster & Ksiazek 2012, 45), untersucht. Die Herleitung der beiden zu vergleichenden Genre-Kategorien erfolgt aus der Marktperspektive. Die Nutzung der zwei typischen Fälle der „most different“ (vgl. Hopkin 2002, 292ff.) Genre-Kategorien hinsichtlich der Markteigenschaften wird verglichen, da hier ein größtmöglicher Unterschied in Mediennutzungsmustern erwartet werden kann²⁹. Dies sind online-born Medienangebote vs. Medienangeboten mit Ursprung im Offlinemedienmarkt wie die Onlineversionen von Print-Zeitungen und -Zeitschriften (siehe Text 3). Sollte Publikumsfragmentierung online auf Grund der Marktdynamiken im Online-Sektor größer sein, müssten sich Unterschiede in den Mediennutzungsmustern bei diesen beiden Genres besonders hoch sein. Denn neben der Soziodemographie von Nutzenden, gelten auch die Marktdynamiken im Netz, welche sich besonders bei online-born Medienangeboten zeigen (vgl. Text 3), als potenzielle Determinanten von (Online-)Fragmentierung.

In einer weiteren empirischen Einzelstudie dieses Kumulus (Text 4) wird Publikumsfragmentierung online in der ganzen inhaltlichen Breite hinsichtlich der Publikumsüberschneidungen und überlappender Mediennutzungsmuster analysiert. Die leitende Forschungsfrage hierbei ist:

Gibt es Teilpublika im Gesamtpublikum? Wenn ja, wie bzw. was konstituiert diese und inwiefern ist eine Überlappung dieser Teilpublika zu beobachten?

Im Sinne des methodologischen Individualismus wird auch hier die Mikroebene berücksichtigt und individuellen Nutzung zu Nutzungsmustern des Publikums zusammengefasst. Anhand überlappender Nutzungsmuster von Individuen wird mithilfe der Methodik der

²⁹ Dadurch werden auch mögliche Determinanten des Medienmarktes auf Publikumsfragmentierung in die Analyse miteinbezogen und folgt demnach dem zuschauerzentrierten Ansatz der hier vorliegenden Arbeit, während die meisten Digital Divide Studien einem eher nutzerzentrierten Ansatz folgen.

Netzwerkanalyse analysiert inwiefern sich Teilpublika identifizieren lassen. Ein Teilpublikum ist dabei definiert als eine Gruppe von Nutzenden mit überlappenden Mediennutzungsmustern, welche keine oder kaum überlappende Nutzungsmuster mit Individuen anderer Teilpublika aufweist. Die Aggregation individueller Nutzungsmuster erfolgt auf Basis eines innovativen methodischen Ansatzes, bei welchem die Bedeutung eines Einzelangebotes in der Nutzung einer Person anhand des Repertoire-Umfangs und der Nutzungshäufigkeit berücksichtigt wird (siehe Text 4).

Außerdem wird die Zusammensetzung der gemeinsam genutzten Onlinemedien hinsichtlich des thematischen Genres und der (geteilten) Nutzung politischer Information untersucht (siehe Text 4). Ebenso werden die Nutzenden dieser Teilpublika hinsichtlich ihrer soziodemographischen Merkmale auf Gemeinsamkeiten innerhalb der Teilpublika, also einer Gruppe Nutzender mit überlappenden Mediennutzungsmustern, sowie Unterschiede zwischen den Teilpublika analysiert. So sollen explorativ und datengetrieben Hinweise auf Determinanten von Publikumsfragmentierung online in Deutschland identifiziert werden. Durch den theoretischen Ansatz der News Reading Publics (vgl. Mukerjee 2021a) werden beobachtete, überlappende Nutzungsmuster nach Erklärungsfaktoren die aus den verschiedenen Kontexthypothesen und Entscheidungsregeln, die in der Mediennutzungsforschung üblicherweise herangezogen werden, resultieren können. Aus einer publikumszentrierten Perspektive werden dafür sowohl Merkmale des Individuums als auch des Onlinemedienangebots berücksichtigt und untersucht, welche sich als Faktoren voneinander abweichender Medienmuster identifizieren lassen. Diese werden als Hinweise auf Determinanten von Publikumsfragmentierung online in Deutschland interpretiert.

4 Zusammenfassung, Limitationen und Ertrag der Einzelschriften

Nachfolgend werden die Einzelschriften jeweils kurz zusammengefasst, um einen Überblick über diese zu geben. Dazu wird jeweils kurz erläutert, was die Forschungslücke ist, an welche der jeweilige Einzelbeitrag anknüpft, und die Ergebnisse zusammengefasst. Abschließend folgt für jeden Einzelbeitrag eine kurze Einordnung anhand der Limitation der Studien sowie eine Erörterung des Ertrags der Studie. Die vier Einzelstudien sind darüber hinaus gemeinsam mit der jeweiligen Erklärung der Eigenanteile in voller Länge in Anhang II zu finden.

4.1 Text 1: *The Longitudinal IntermediaPlus (2014–2016): A Case Study in Structuring Unstructured Big Data (Brentel & Winters 2021)*

Vor dem Hintergrund des Dissertationsprojektes wurde das Fehlen geeigneter Daten als Forschungslücke identifiziert. Daher wird im ersten Text des Kumulus die Forschungsfrage, wie eine geeignete Datenbasis zur Untersuchung von Publikumsfragmentierung online aussehen kann, beantwortet und die Datenquelle der Media-Analyse *IntermediaPlus*, deren Aufbereitung und der harmonisierte Datensatz als Ergebnis dessen skizziert.

Ergänzt wird diese Einzelschrift des Kumulus daher durch die Datenpublikation der aufbereiteten Datenquellen der *IntermediaPlus (2014-2016)*, welche bei GESIS Leibniz-Instituts in Köln publiziert und archiviert ist:

- Brentel, Inga, Kampes, Céline Fabienne, & Jandura, Olaf (2022). Media-Analyse Daten: IntermediaPlus Daten von 2014 bis 2016 (MA IntermediaPlus). *GESIS, Köln*. ZA5769 Datenfile Version 1.0.0, <https://doi.org/10.4232/1.13530>.

4.1.2 Forschungslücke

Bisher liegen keine geeigneten Daten vor, um Nutzungsmuster online in der nötigen Detailtiefe zu untersuchen und dabei zumindest ein Marktsegment wie den kommerziellen, privatwirtschaftlichen Rundfunk umfassend zu repräsentieren. Die meisten Studien zur Onlinemediennutzung basieren auf einer begrenzten Auswahl von Webseiten. So umfassen beispielsweise recht umfassende Studien die hundert (siehe u.a. Neuberger & Lobigs 2010, 137ff.; Ng & Taneja 2019; Wu et al. 2021) oder tausend (Wu & Taneja 2016) am meisten genutzten Webseiten. Oftmals werden Daten zur Internetnutzung in Deutschland jedoch durch Befragungen ermittelt: Es wird nach den wichtigsten Informationsquellen gefragt, wie wichtig Onlinemedien zur Informationssuche sind und welche Typen von Medienangeboten online

(TV- und Radioanbieter online, Zeitungen online, Zeitschriften online, Social Media etc.) als Nachrichtenquelle dienen (siehe u.a. Hölzig, Hasebrink & Behre 2021; Kleinen von Königslow 2016; Fletcher & Nielsen 2017; Bächler 2017; Magin et al. 2022). Ein anderer Ansatz, der zur Medienauswahl für Studien zur Onlinemediennutzung verwendet wird, ist die Erfassung der Online-Pendants der Medienangebote, welchen offline die größte Relevanz zugesprochen wird oder eben eine hohe Online-Reichweite haben (siehe u.a. Trilling & Schoenbach 2013; Mahrt 2019; Kleinen von Königslow 2016). Jedoch stellt sich die Frage, ob diese „klassische[n] Instrumente zur Erfassung von Mediennutzung noch zeitgemäß sind“, um „die Nutzungsrealität adäquat zu erfassen und systematische Muster zu identifizieren“ (Stark & Kist 2020, 1139).

Tracking-Daten gelten als zuverlässiger, wenn es darum geht Onlineverhalten wie die Mediennutzung abzubilden (vgl. Fletcher & Nielsen 2017, 484; Mukerjee et al. 2018a, 27, 46; Ørmen 2019b, 1657; van Atteveldt & Peng 2018, 8) und erlauben neue Möglichkeiten in der Untersuchung verschiedener Fragestellungen (Waldherr et al. 2021 164). Auch werden sogenannte Klick-Streams und Verlinkungsstrukturen zwischen Medienangeboten zur Analyse von Nutzungsmustern online verwendet (vgl. Rau & Stier 2019, 408f.; siehe u.a. Bruns et al. 2011; Mahrt 2019; Taneja & Wu 2018; Wu et al. 2021; Puschmann 2019). Dabei handelt es sich in der Regel um die Gesamtangebote einer Webseite statt der Einzelangebote, also der Differenzierung nach Unter-Seiten³⁰. Insbesondere im Kontext von Publikumsfragmentierung online wird aber ein deutlich größerer Detailgrad als Gesamtangebote oder Medienmarken in der Erfassung der Mediennutzung benötigt (vgl. Webster & Ksiazek 2012, 51; Stark & Kist 2020, 1139). Vor allem vor dem Hintergrund der Ausdifferenzierung und Spezialisierung des Medienangebots im Internet und der Sorge nach einer Publikumsfragmentierung entlang dieser ist die Detailtiefe von Einzelangeboten unerlässlich. Denn „because news consumers nowadays often cherry-pick articles from multiple sites rather than relying on a single source of news (Costera Meijer & Groot Kormeling, 2015)“ ist „a more sophisticated measurement of news consumption“ (van Atteveldt & Peng 2018, 83) bestehend aus einer Kombination von Tracking und Befragungsdaten geboten.

³⁰ Eine Ausnahme bilden einige wenige Studien zu Nutzungsepisoden. Für die Analyse von Nutzungsepisoden wurde meist genau getrackt, welche Seiten bei einem Online-Besuch genutzt wurden (z.B. Taneja et al. 2012; Wu & Taneja 2021). Dafür werden oftmals Teilnehmende rekrutiert, welche dann spezielle Cookies o.ä. installieren, wodurch für einen festgelegten Zeitraum ihre Online-Tätigkeiten erfasst und an die Forschenden übermittelt werden. Durch die DSGVO-Verordnung im Mai 2018 sind Daten mit diesem Detailgrad (auf URL-Ebene) jedoch nicht mehr ohne weiteres zu veröffentlichen, da die Anonymität der Nutzenden gewährleistet werden muss (vgl. van Atteveldt et al. 2019b, 3938; Peter et al. 2020, 9, 11).

4.1.2 Ergebnisse

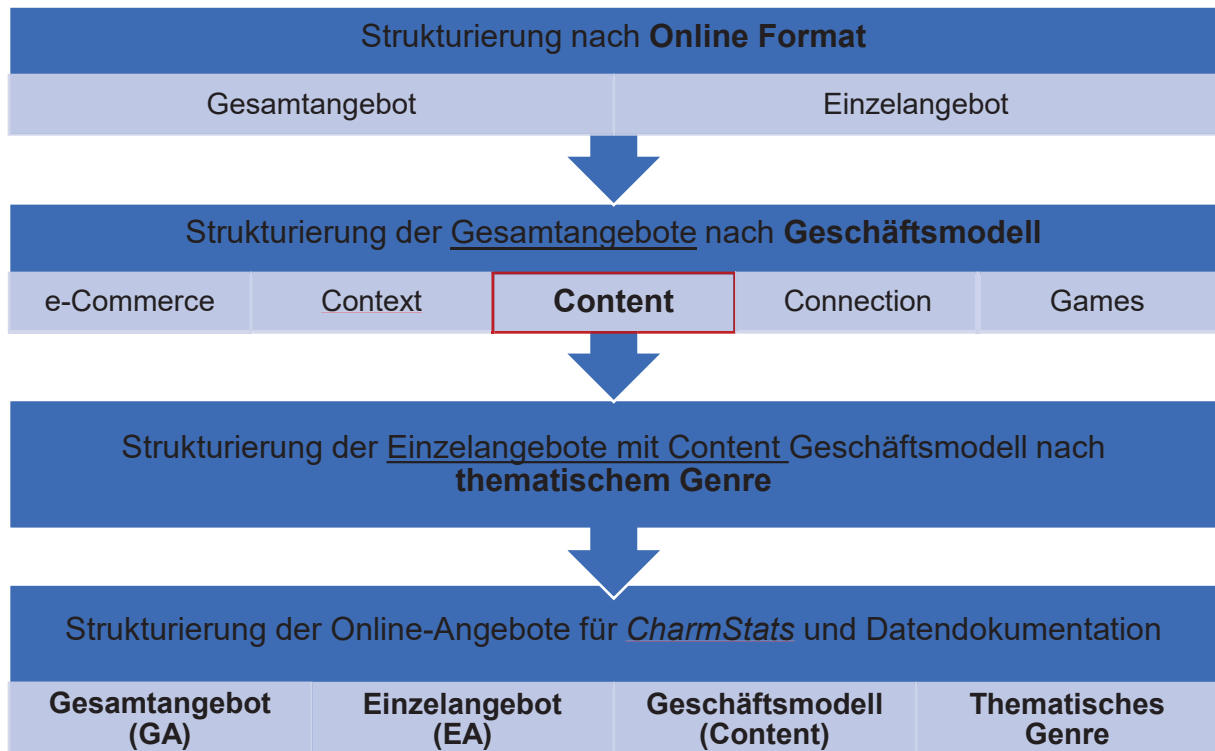
Die Datenquelle der Media Analyse *IntermediaPlus*, in erster Linie die Daten der Online-Tranche, hat sich als eine geeignete Datenbasis für Deutschland zur Untersuchung von Publikumsfragmentierung online gezeigt. Sie resultiert aus umfassenden Tracking-Daten (ca. 300.000 Unique User auf Tagesbasis) des Medienverhaltens und deckt die Nutzung auf Ebene der Unterseiten für eine große Bandbreite kommerzieller Medienangebote in Deutschland ab. Die Daten der Media Analyse gelten als die „größte repräsentative Erhebung des Medienkonsumverhaltens in Deutschland“ (Stark & Kist 2020, 1146). Es handelt sich dabei um Fusionsdaten, welche im Rahmen der Marktforschung erhoben werden (siehe Hofsäss et al. 2020; agma-mmc.de)³¹. Die Online-Tranche der MA *IntermediaPlus* wird durch die agof (Arbeitsgemeinschaft Online Forschung e.V.) zugeliefert und anhand eines Multimethodenansatzes erhoben, bei welcher die technische Messung (Online-Tracking) mit Telefon-, On-Site und In-App-Befragungen kombiniert wird. So werden Unique User identifiziert und Merkmale der getrackten Nutzenden erfasst (vgl. Stark & Kist 2020, 1146; agof.de).

Da die Rohdaten zum Zweck der Marktforschung erhoben wurden, sind sie eher unstrukturiert und weisen einen Silo-Charakter auf. Daher mussten die Rohdaten für eine wissenschaftliche Nutzung aufwendig aufbereitet werden. Zum einen sollten die Rohdaten für den Längsschnitt von 2014 bis 2016 harmonisiert werden. Zum anderen mussten die Rohdaten grundlegend strukturiert, die Silo-Struktur aufgebrochen und durch offene Datenspeicher-Strukturen ersetzt werden. Die Strukturierung der Daten erfolgte anhand der Markteigenschaften (siehe Abbildung 2), wozu die Daten um Marktdaten angereichert wurden. Eine Unterscheidung der erfassten Onlineangebote nach Gesamt- und Einzelangeboten sowie nach dem jeweiligen Geschäftsmodell für die Gesamtangebote (siehe Text 1 sowie Brentel et al. 2021, insbesondere „Konzeptionelle Überlegungen zur Datenaufbereitung MA IntermediaPlus Online“ für weitere Information) bildete die Basis. Das Augenmerk für die Harmonisierung und Aufbereitung der Datenquelle lag auf den Einzelangeboten der Gesamtangebote mit dem Geschäftsmodell „Content“ (vgl. Wirtz 2018), da die Nutzung von Einzelangeboten, welche den Zweck beziehungsweise das Geschäftsmodell verfolgen Inhalte anzubieten, die relevanten für die Betrachtung von Publikumsfragmentierung sind. Durch eine halb-automatisierte Textanalyse der Bezeichnungen der Einzelangebote – sprich der Unterseiten und Unter-Unterseiten – wurde eine induktive Kategorienbildung für die inhaltlichen Genres

³¹ Mittels einer HUB-Survey werden Merkmale der Nutzenden sowie die TV-, Audio- und Printnutzung mit den Trackingdaten anhand mehrerer Verfahren und In-App und On-Site Befragungen fusioniert (vgl. Hofsäss et al. 2020; agof.de).

vorgenommen (siehe Kampes 2020b und Brentel et al. 2021) und als Struktureinheit hinzugefügt.

Abbildung 2: Strukturierungseinheiten und -logik für die Aufbereitung der Rohdaten



Quelle: Eigene Darstellung

Resultat der Aufbereitung der Rohdaten der Media-Analyse *IntermediaPlus* für die Online-Tranche ist der bei GESIS publizierte Datensatz des harmonisierten *Längsschnitts der IntermediaPlus (2014-2016)*, bereitgestellt und offen zugänglich für wissenschaftliche Forschung. Begleitet wird die Publikation von einer ausführlichen Datendokumentation, welche wie bereits die Daten selbst den FAIR-Prinzipien (Wilkinson et al. 2016; Mons et al. 2017) sowie den Leitlinien der DGPuK für Forschungsdaten (Peter et al. 2020) entspricht. Ebenso ist ein vier-schrittiger (konzeptioneller) Prozess zur Aufbereitung mit Hilfe der Harmonisierungssoftware CharmStats³² entwickelt worden, der auf andere unstrukturierten big-data Daten übertragen werden kann und den FAIR-Prinzipien gerecht wird (siehe Text 1 und 2).

4.1.3 Limitation

Auf Grund der Tatsache, dass es sich bei der Media-Analyse um Marktforschungsdaten handelt, ist die Erfassung und der Detailgrad dieser – sprich für wie viele Unterebenen einer

³² Eine Software entwickelt von CESSDA und GESIS zur Digitalisierung von Harmonisierungsprozessen.

Webseite (Gesamtangebot vs. Einzelangebot; Unterseite vs. Unter-Unterseite) die Nutzung einzeln ausgewiesen wird – von der Mitgliedschaft bei agof, IVW und / oder agma abhängig. Daher werden lediglich privatwirtschaftliche Onlineangebote erfasst, welche sich durch Werbung finanzieren. Bei kleineren Onlineangeboten kann es sein, dass die Erfassung nicht jährlich, sondern lediglich alle zwei Jahre erfolgt (z.B. 4trips.de, bike-magazin.de, klassikradio.de, computerwoche.de) oder in Kombination mit anderen Onlineangeboten eines Anbieters (z.B. gmx.de und web.de oder regionale Zeitungen wie Express und Rheinische Post). Da jedoch viel genutzte, kommerzielle Angebote auch eine hohe Relevanz auf dem Werbemarkt als Werbeträger-Medien haben, sind gerade diese in der Regel sehr detailliert mit allen vorhanden Unterebenen einer Webseite für die Online-Tranche der Media-Analyse *IntermediaPlus* erhoben (z.B. bild.de, rtl.de, t-online.de). Ein vollständiges Abbild des Internets kann aber auch die für Deutschland umfassendste Studie zu Medienkonsum nicht liefern, es bleibt bei einem Ausschnitt der tatsächlichen Medien-(nutzungs-)Realität des Internets. Zusätzlich handelt es sich bei der *IntermediaPlus* um Fusionsdaten, weshalb Studien zu Kausalitäten (z.B. Regressionsanalysen) nur eingeschränkt möglich sind. Denn eine Zeile in den Daten entspricht nicht einer natürlichen Person, sondern einer konstruierten Person: Das Verhalten aus den Einzelstudien für TV, Audio, Print und Online wird anhand von Markern, in der Regel basierten auf soziodemographischen Merkmalen, fusioniert. Im Fall der Media-Analyse handelt es sich um ein komplexes Multimethoden Design zur Fusionierung, das auf einer HUB-Survey von 18.000 Personen basiert (siehe Hofsäss et al. 2020; agma-mmc.de).

Trotzdem handelt es sich um die umfassendste und aktuellste Datenquelle zur Online-Mediennutzung. Somit werden neue empirische Studien erst ermöglicht, um wissenschaftliche Erkenntnisse neu zu gewinnen oder weiter zu entwickeln. Auch wenn es sich um Marktforschungsdaten handelt, sind sie nicht von geringerer Qualität. In der Wissenschaft renommierte Datenerhebungsinstitute sind an der Erhebung beteiligt (siehe agma-mmc.de, Codepläne). Durch das Multi-Methoden Design und die HUB-Survey, ist die Fusionierung sehr präzise und bietet damit umfassende und detaillierte Daten, welche datenschutzkonform die Anonymität der Nutzenden und Befragten gewährleistet. Ein Joint Venture (Joint-Industry-Committees), bestehend aus den verschiedenen Akteuren des Medienmarktes mit unterschiedlichen Interessen, wird die Qualität der Daten gewährleistet (vgl. Brentel & Winters 2021 [Text 1]; siehe Hofsäss et al. 2020), um einem Interessengeleitete Verzerrung in der Messung von Mediennutzung u.a. entgegenzuwirken.

Eine weitere Limitation der *IntermediaPlus* (2014-2016) ist das Alter der Daten: 2016 als aktuellstes Jahr der verfügbar gemachten Zeitreihe ist im schnelllebigen Internetzeitalter mit seinem volatilen Medienmarkt vergleichsweise alt. Die Daten werden durch agma mit einem zwei jährigen Embargo versehen, das heißt die Datenaufbereitung kann immer im Versatz von

zwei Jahren erfolgen. Demnach wären inzwischen die Rohdaten der *IntermediaPlus* Online Tranche bis einschließlich 2019 verfügbar. Eine Aktualisierung des Längsschnittes um die Jahre 2017, 2018 und 2019 war im Rahmen dieser Dissertationsleistung nicht mehr möglich, da die Entwicklung der digitalen Prozesskette zur Aufbereitung der Media-Analyse *IntermediaPlus* (siehe Brentel & Winters, *forthcoming* [Text 2]) sowie die Bereitstellung dieser zur Nachnutzung (siehe Brentel et al. 2021; Brentel 2021a; Brentel 2021b; Brentel & Fortmann 2021 [Begleitpublikationen Daten 1, 2, 3 und 4]) bereits viel Zeit in Anspruch genommen hat. Dennoch wird mit der Zeitreihe von 2014 bis 2016 ein wichtiger Zeitraum der beginnenden Marktkonsolidierung abgebildet, wie beispielsweise der Kauf von t-online durch Ströer im Jahr 2014 und die darauffolgenden Veränderungen im Onlinemedienmarkt.

4.1.4 Ertrag

Eine nachhaltig aufbereitete, archivierte und für Wissenschaft zugängliche, umfassende Datenquelle zur Onlinemediennutzung in Deutschland, der Längsschnitt *IntermediaPlus* (2014-2016), ist ein beständiger Ertrag für das kommunikationswissenschaftliche Fach (vgl. Strippel 2021; Peter et al. 2020, 2 ff.). So kann durch eine Sekundärnutzung der *IntermediaPlus* (2014-2016) beispielweise die Methodenentwicklung sowie eine längst überfällige (Weiter-)Entwicklung von Theorien im Kontext digitaler Gesellschaft erfolgen (Strippel et al. 2018, 19ff.; vgl. Brosius 2013; Hepp 2016, 237-240; van Aelst et al. 2017, 20)³³. Durch eine gemeinsame Datenbasis kann eine Entwicklung standardisierter Operationalisierungskonzepte für Mediennutzung online vorangetrieben werden (vgl. van Aelst et al. 2017, 20; ein Vorschlag befindet sich in Text 4). Denn aktuell fehlt eine Diskussion darüber, wer als Nutzende gilt, wie sich regelhafte Nutzung konstituiert, ob und inwiefern aus dem Offlinekontext bekannte Strukturierungen wie „Weitester Nutzerkreis“, „regelmäßige Nutzung“, etc. für den Onlinekontext übertragen werden können (ein Vorschlag für die Einteilung und Operationalisierung in Nutzende, regelmäßige Nutzende und tägliche Nutzende findet sich in Text 3). Zudem können Nachfolgestudien auf die Datenarbeit aufbauen, um Publikumsfragmentierung online weiter zu ergründen. Anhand der gleichen Datenbasis kann die aktuell fehlende Vergleichbarkeit von Studien und ihren Ergebnissen erreicht werden.

³³ Siehe außerdem zur Bedeutung von Theorien im Kontext von Computational Communication Science sowie Makrophänomenen Waldherr et al. 2021. Zum Potenzial der Sekundäranalysen für die Theoriebildung sowie Methodenentwicklung siehe z.B. Medjedovic 2014, u.a. 25, 34, 215-224 und Friedrichs 1983, 355. Zur Bedeutung von Transparenz und Verfügbarmachen von Forschungsdaten siehe zudem Peter et al. 2020, 2ff.

4.2 Text 2: Big-Data Management Case Study: The Process-Cycle for the Longitudinal IntermediaPlus (2014-2016) Dataset (Brentel & Winters, forthcoming)

Ergänzt wird diese Einzelschrift des Kumulus durch vier weitere Publikationen, welche bei SoWiDataNet, einem kuratierten Repositorium unter Verwendung von DOIs als persistenten und einzigartigen Identifier für Forschungssoftware, Skripte, Daten u.ä. des GESIS Leibniz-Instituts in Köln:

- Brentel et al. 2021 „Datenaufbereitung der MA-IntermediaPlus: Teil 1 – Ein Überblick und Prozess-Schritt 1 "Strukturieren & Anpassen"“ <https://doi.org/10.7802/2339>.
- Brentel 2021a „Datenaufbereitung der MA-IntermediaPlus: Teil 2 – Prozess-Schritt 2 "Import & Harmonisieren" in und mit CharmStats“ <https://doi.org/10.7802/2340>.
- Brentel 2021b „Datenaufbereitung der MA-IntermediaPlus: Teil 3 – Prozess-Schritt 3 "Harmonisierung Anwenden" mit charmstana“ <https://doi.org/10.7802/2341>.
- Brentel & Fortmann 2021 „Datenaufbereitung der MA-IntermediaPlus: Teil 4 – Prozess-Schritt 4 "Dokumentation erstellen"“ <https://doi.org/10.7802/2342>.

Diese dokumentieren die Datenaufbereitung, Harmonisierung und das Datenmanagement für die Media-Analyse *IntermediaPlus*, vor allem für die Online-Tranche. Enthalten ist ein Projektplan, konzeptionelle Überlegungen für die Strukturierung und Aufbereitung der *IntermediaPlus* Online-Tranche für die Harmonisierung der einzelnen Querschnitte zu cross-sectional Längsschnittdaten, Schritt-für-Schritt-Anleitungen zu den einzelnen Prozessabläufen, Dateien und Skripte zur Teil-Automatisierung sowie Umsetzung der Aufbereitung für die Media-Analyse *IntermediaPlus*-Daten. Dies gewährleistet in Verbund mit Text 1 und 2 die Erfüllung der FAIR-Prinzipien und stellt die Antwort auf die Forschungsfrage, wie die Daten der *IntermediaPlus* (2014-2016) nachhaltig (FAIR) aufbereitet und für akademische Forschung zugänglich gemacht werden und ein Datenmanagementsystem für diese big-data Datenquelle aussehen kann, dar.

4.2.1 Forschungslücke

In der Wissenschaft im Allgemeinen und der Kommunikationswissenschaft im Speziellen ist seit Jahren von einer Replikationskrise die Rede. Der Ruf nach Open Science, also dem Zugänglichmachen von Forschung, den Analysematerialien und Daten, u.a. zur Gewährleistung von Transparenz, wird lauter (Dienlin et al. 2020; van Atteveldt et al. 2019b). Denn „um die Robustheit, Reproduzierbarkeit, Replizierbarkeit und Generalisierbarkeit langfristig zu gewährleisten, so gilt auch für deskriptive Analysen und explorative Forschung,

dass erst die Verfügbarkeit der Daten das vollständige nachvollziehen, wie auf Basis der Daten die Ergebnisse entstanden sind, ermöglicht.“ (Peter et al. 2020, 4). Dabei ist „die Einhaltung methodischer (Qualitäts-)Standards in der Erhebung und Auswertung von Daten die notwendige Voraussetzung für einen Erkenntniszuwachs“ der Wissenschaft (Peter et al. 2020, 2). Erste Leitlinien und Qualitätskriterien wurden in den letzten Jahren, vor allem im Kontext von Open Science Initiativen, entwickelt und zur Diskussion gestellt (siehe Wilkinson et al. 2016). So hat auch die DGPuK 2019 bzw. 2020 Leitlinien für Forschungsdaten veröffentlicht (Peter et al. 2020). Diese bleiben jedoch wie die FAIR-Prinzipien (Wilkinson et al. 2016, Mons et al 2017) als Gütekriterien relativ allgemein. Ein Best-Practice für die Umsetzung bzw. Anwendung dieser Leitlinien und Qualitätskriterien auf big-data Datenquellen sowie digitalen Verhaltensdaten sind sehr rar (siehe Jungherr, Jürgens & Schoen 2018 sowie Jürgens & Jungherr 2016 für eine Anwendung auf Twitter-Textdaten). Die Diskussion hinsichtlich Datendokumentation, Bereitstellung, Aufbereitung, Datenqualitätsstandards, geeigneten (Datei-)Formate, etc. steht im Kontext von big-data und digitalen Verhaltensdaten noch am Anfang (vgl. Text 2). Doch wäre dies angesichts des Potenzials dieser Art der Daten für die Forschung und ihre wachsende Beliebtheit, insbesondere im Kontext von Computational Social Science, notwendig. Denn aktuell „stehen Forscher (...) vor einer Reihe von praktischen und bisher ungelösten Problemen“ (Jungherr et al. 2018, 255), wenn sie mit digitalen Verhaltensdaten und big-data arbeiten wollen. So entstehen „grundlegende[n] Fragen im praktischen Umgang mit diesem neuen Datentyp“ (Jungherr et al. 2018, 256) hinsichtlich der Sammlung, Speicherung, Verarbeitung und Analyse (vgl. auch van Atteveldt et al. 2019b). Ein Best-Practice zur Bereitstellung von Daten und dem Umgang, also Datenmanagement, mit einer big-data Datenquelle wie der Media-Analyse *IntermediaPlus*, welche digitale Verhaltensdaten beinhaltet, bei dem die bislang diskutierten Gütekriterien (siehe Wilkinson et al. 2016; van Atteveldt et al. 2019a, 2019b) sowie die Leitlinien der DGPuK berücksichtigt werden, fehlt bislang.

4.2.2 Ergebnisse

Im Rahmen der Aufbereitung der Media-Analyse *IntermediaPlus* Online-Tranche zum *Längsschnitt der IntermediaPlus (2014-2016)* wurde ein methodologisches Vorgehen für den FAIRen Umgang mit den eher unstrukturierten, big-data Rohdatensätzen entwickelt, welches den Leitlinien für Forschungsdaten der DGPuK (Peter et al. 2020) entspricht (siehe Text 2; Brentel et al. 2021; Brentel 2021a; Brentel 2021b; Brentel & Fortmann 2021 [Begleitpublikationen Daten 1, 2, 3 und 4]). Ergebnis ist eine digitale Prozesskette zur Harmonisierung der Media-Analyse Daten und dem Datenmanagement für den big-data Datensatz (> 50 GB). Diese digitale Prozesskette umfasst vier Schritte: (1) Strukturieren und

(Dateiformate) Anpassen, (2) Import in und Harmonisierung der Metadaten mit der Harmonisierungssoftware *CharmStats* von CESSDA und GESIS, (3) Harmonisierung anwenden, (4) Datendokumentation erstellen. Beginnend mit dem ersten Prozess-Schritt werden die nötigen Meta-Informationen der Rohdaten digitalisiert und fortlaufend über eigens entwickelte oder weiterentwickelte technische Schnittstellen verbunden. Dies ermöglicht nun eine Teil-automatisierte Aufbereitung und Harmonisierung der Media-Analyse *IntermediaPlus*-Daten. Anhand der Begleitpublikationen Daten 1, 2, 3 und 4 wird der gesamte Prozess der Datenaufbereitung und -harmonisierung für die *IntermediaPlus (2014-2016)* transparent und zugänglich gemacht. Die digitale Prozesskette kann vollständig nachvollzogen, reproduziert und für die Aufbereitung weiterer Media-Analyse Daten der *IntermediaPlus* nachgenutzt und repliziert werden (vgl. Leitlinien der DGPK: Peter et al. 2020). Sowohl die Datenbank mit dem aufbereiteten *Längsschnitt der IntermediaPlus für 2014 bis 2016* ist bei GESIS im kuratierten Datenbestandskatalog (DBK) als auch die Begleitpublikationen der Daten mit den Dateien zur digitalen Prozesskette im SoWiDataNet von GESIS publiziert. Bei beiden handelt es sich um „unabhängige öffentliche Repositorien“, welche laut DGPK Leitlinien für die Publikation von Forschungsdaten „im Hinblick auf die Sichtbarkeit bzw. Auffindbarkeit der Daten die zu bevorzugende Variante“ sind (Peter et al. 2020, 11). Der DBK und SoWiDataNet werden als Repositorien durch die DGPK empfohlen und folgen allen sieben Gütekriterien der DGPK für Repositorien (Peter et al. 2020, 11; siehe Klein et al. 2018), wie beispielsweise die Garantie einer langfristigen Archivierung oder eine Veröffentlichung der Daten und Dateien mit eindeutigen und persistenten Identifikationsnummern wie DOIs zur Auffindbarkeit und Zitierfähigkeit.

Im ersten Schritt der digitalen Prozesskette werden die Informationen, die aus den eher unstrukturierten Rohdaten sowie den dazugehörigen Codeplänen entnommen werden können sortiert und für die konzeptionelle Strukturierung der Daten aufgearbeitet – dies folgt u.a. der Logik der „information extraction“ (siehe Warin & Sanger 2014), einem methodologischen Ansatz aus der Computerwissenschaft. Die in Abbildung 2 und bereits für Text 1 beschriebenen Strukturierungsebenen für die Onlinemedienangebote resultieren aus diesem Verfahren (siehe auch Text 2 und Brentel et al. 2021, „Konzeptionelle Überlegungen zur Datenaufbereitung MA *IntermediaPlus* Online“ [Begleitpublikation Daten 1]). Die verschiedenen Dateiformate der Rohdaten, Codepläne und Fragebögen werden angepasst. Das heißt die Rohdaten werden zu *.sav-Dateien (SPSS-Datensätze) konvertiert sowie relevante Information aus Codeplänen und Fragebögen in ein System aus Excel-Dateien übertragen (siehe Text 2 und Brentel et al. 2021 [Begleitpublikation Daten 1]). Hier werden Forscher nun anhand von Excel-Makros durch den Datenaufbereitungs- und Harmonisierungsprozess geführt, die Informationen für bereits vorhandene

Onlinemedienangebote in den Jahren 2014 bis 2016 für nachfolgende Jahre automatisch übernommen und Harmonisierungspläne entstehen jetzt Großteils automatisiert.

Im zweiten Schritt werden die Meta-Informationen der Rohdaten (der Quellvariablen) sowie die gewünschte Zielvariable in die Harmonisierungssoftware *CharmStats* importiert. Die Meta-Informationen umfassen Variablenname, -label, Werte, Wertelabel, Datenquelle, Studieninformation und ggf. Angaben zum Fragebogen, sprich der konkreten Fragen mit Interviewer-Anweisungen (Frageinformation), auf die eine Variable zurückzuführen ist. *CharmStats* wurde von CESSDA und GESIS zur Vereinfachung von Harmonisierungsprozessen sowie der transparenten Bereitstellung von Harmonisierungskonzepten entwickelt. Die Anwendung bietet eine klare Struktur, die beim Harmonisierungsprozess eingehalten werden muss, mehrere Forscher können zeitgleich an der Datenaufbereitung arbeiten und Hinweise in Form von Kommentaren zu einzelnen Variablen oder Harmonisierungskonzepten für nachnutzende Forschende hinterlegt werden (vgl. Text 1). Im Hintergrund wird durch Nutzung von *CharmStats* eine MySQL-Datenbank angelegt. Nach dem Import der benötigten Information können beliebig viele Quellvariablen einer Zielvariable zugeordnet werden; die einzelnen Werte-Kategorien einer Zielvariable werden durch einfache Drop-Down-Menüs den Werten der Quellvariablen zugewiesen. So entsteht die jeweilige Re-Kodierungsinformation für die Quellvariablen aus den Rohdatensätzen – die Harmonisierungsinformation wird in der *CharmStats*-Datenbank gespeichert. Die Entwickler der Software *CharmStats* hatten vor allem nominale oder kategoriale Daten aus qualitativen Erhebungen sowie Befragungen im Sinn. Für die digitale Prozesskette zur Aufbereitung und Harmonisierung der Media-Analyse Daten wurde die Anwendung *CharmStats* für die Verwendung von big-data und metrischen Variablen weiterentwickelt (siehe Text 1 und 2): Das Feature „virtuelle Variablen“ ermöglicht es bis zu 300 Variablen auf einmal zu importieren und zu harmonisieren. Voraussetzung für dieses Feature ist, dass die Meta-Information bezüglich der Werte-Kategorien, Studien- sowie der Frageinformation für diese Variablen gleich sind. Dies ist beispielweise bei Onlineangeboten eines Genres auf Ebene der Einzelangebote (vgl. Abbildung 2) oder den Gesamtangeboten eines Geschäftsmodells online der Fall. Für die metrischen Variablen wird lediglich eine Werte-Kategorie, ein Wert für Missings, angelegt und zur Harmonisierung zugeordnet (siehe Brentel 2021a [Begleitpublikation Daten 2]).

Im dritten Schritt der digitalen Prozesskette wird die Harmonisierung anhand eines Python-Skripts mit Rückgriff auf die *CharmStats* Datenbank angewendet: die zu *.sav-Format konvertierten Roh-Datensätze werden eingelesen und harmonisiert in eine neue MySQL-Datenbank, *charmstana*, eingepflegt. Die Datenbank *charmstana* wurde so konzipiert und aufgebaut, dass die Daten möglichst sparsam gespeichert und für Analysen abrufbar sind

(siehe Text 2; Brentel 2021b [Begleitpublikation Daten 3]). So kann den big-data Herausforderungen von *Volume* und *Velocity* (Gudivada, Apon & Ding 2017, 4) entgegnet werden. Die Harmonisierungsinformation aus *CharmStats* wird in gesonderten Datenbanktabellen abgespeichert und lassen sich u.a. wie die „Variablenansicht“ in SPSS lesen. Die harmonisierten Daten der Mediennutzung und mit den Merkmalen zu den Befragten sind für Analysen je Jahr in eigene Datenbanktabellen gespeichert. Anders als bei SPSS oder STATA Dateien, bei welchen eine Zeile einem Fall / einer Person entspricht, sind die Daten zu einer Person in den *charmstana* Analyse-Tabellen untereinander angeordnet: jede Variable zu einer Person entspricht einer Zeile in der Datentabelle. So kann für Computational Social Science Methoden, wie für Kantenlisten bei sozialen Netzwerkanalysen, in der Regel direkt auf die MySQL-Datenbank zugegriffen werden ohne große Anpassungen vornehmen zu müssen. Doch auch für klassische Methoden können Teildatensätze aus der *charmstana*-Datenbank in das gewünschte Dateiformat exportiert werden.

Im letzten und vierten Schritt wird die Datendokumentation mittels eines PHP-Skripts erstellt (siehe Text 2; Brentel & Fortmann 2021 [Begleitpublikation Daten 4]). Mit ausführen des Skripts wird unter Rückgriff auf die MySQL-Datenbanken *charmstana* und von *CharmStats* eine umfassende Datendokumentation im *.html-Format ausgegeben. Hier werden zum einen die Informationen der harmonisierten Variablen zu Variablenname, -label, Werte und Wertelabel dargestellt. Zum anderen werden die entsprechenden Informationen der Rohdaten, also der Variablen aus jedem Jahr, in einer Übersichtstabelle bereitgestellt. So können Frageformulierungen, Interviewer-Anweisungen oder Veränderungen in Fragebatterien (z.B. der Liste abgefragter Items) auf einen Blick für alle harmonisierten Jahre von Nach-Nutzenden erfasst werden. Zudem werden hier für die Harmonisierung hinterlegte Kommentare in *CharmStats* mitaufgenommen, um ggf. notwendige Entscheidungen in der Datensammlung und -aufarbeitung bewusst zu machen und offen zu legen. Durch die Datendokumentation führen Verlinkungsstrukturen direkt zu einzelnen Variablenkomplexen, damit Forschende sich möglichst schnell und einfach zurechtfinden können und die für sie relevante Information herausfiltern können. So werden beispielsweise die Informationen nach thematischen Variablenblöcken sortiert: soziodemographische Merkmale der Befragten, Freizeitverhalten der Befragten, Besitz der Befragten, Besitz des Haushalts der Befragten, Online-Mediennutzung der Befragten und interviewspezifische Variablen. Informationen zu virtuellen Variablen, also Variablen mit gleichen Meta- und Harmonisierungsinformation, werden dabei zusammengefasst und die Datendokumentation somit übersichtlicher.

Mit dieser Prozesskette kann die Einhaltung der FAIR-Prinzipien und DGPuK-Leitlinien für Daten sowie deren Management und Bereitstellung sichergestellt werden. Da bislang keine Best-Practice Studien für big-data und dem Umgang damit einhergehender

Herausforderungen gab, mussten im Rahmen der vorliegenden Dissertation entsprechende Lösungen entwickelt werden. In der Einzelschrift werden daher die konzeptionellen und daraus resultierenden technischen Überlegungen und Entscheidungen zur Einhaltung jedes der vier FAIR-Prinzipien geschildert sowie transparent gemacht. Denn so können die entwickelten Mechanismen für andere Forschungsprojekte übersetzt und angepasst werden.

4.2.3 Limitation

Inzwischen wurde die aktive, technische Betreuung von *CharmStats* durch GESIS eingestellt. Lediglich ein (kostenpflichtiger) Harmonisierungs-Service wird noch auf Nachfrage angeboten, bei dem der Import und die Harmonisierung in und mit *CharmStats* übernommen wird. Das heißt, eine Weiterentwicklung und Aktualisierung der Software sind nicht zu erwarten. Zum Zeitpunkt der Anfertigung dieser Arbeit ist eine Beeinträchtigung der Funktionsweise der digitalen Prozesskette ausgeschlossen. Ob jedoch künftige, Updates der Programm-Versionen, auf denen die verschiedenen technischen Schnittstellen der Prozesskette (u.a. Python, MySQL, PHP) aufgebaut sind, mit der Java-Script und MySQL-Version, auf welcher *CharmStats* basiert, interoperabel sind, kann zum Zeitpunkt der Anfertigung dieser Arbeit nicht abschließend beurteilt werden. Da *CharmStats* auf einer sehr komplexen und verschachtelten MySQL-Datenbank basiert und nur ein Teil der Funktionen von *CharmStats* für die digitale Prozesskette zur Aufbereitung der Media-Analyse Daten repliziert und kopiert werden muss, wäre daher ohnehin eine Vereinfachung und Reduktion auf die, für die digitale Prozesskette relevanten Aspekte, sinnvoll. Die Programmierung eines eigenen Interfaces für die Befüllung der MySQL-Datenbank mit den Metainformationen der Quell- und Zielvariablen sowie der zugehörigen Harmonisierungsinformation als Ersatz der *CharmStats*-Anwendung sollte für geübte Programmierer kein großes Problem darstellen. Denn die benötigten Funktionen sind nicht sehr komplex und *CharmStats* ist eine Open-Source Software, sprich der Code ist über das bei Programmierern beliebte Repositorium Github beziehbar oder auf Nachfrage bei GESIS zu erhalten.

Eine weitere Limitation ist die Spezifikation der digitale Prozesskette für den Fall der Media Analyse Daten. So sind einige Teil-Automatisierungen auf die Media Analyse Daten, insbesondere die Online-Tranche der *IntermediaPlus* Daten, abgestimmt. Dennoch kann die digitale Prozesskette für andere Anwendungsfälle angepasst oder nur Teil der Prozesskette verwendet werden. Die benötigten Unterlagen, Skripte und das Know-how in Form von Konzeptionellen Überlegungen, Anleitungen, Erläuterungen u.a. für den Aufbau einer solchen digitalen Prozesskette zum Management von big-data und digitalen Verhaltensdaten liegen vor.

4.2.4 Ertrag

Die digitale Prozesskette zur Harmonisierung der Media-Analyse *IntermediaPlus* Daten kann als Best-Practice Beispiel für die Aufbereitung, Bereitstellung und den Umgang mit big-data Forschungsdaten herangezogen werden, da es die Gütekriterien (Wilkinson et al. 2016; Dienlin et al. 2020) und Leitlinien für Forschungsdaten (Peter et al. 2020) erfüllt und in jedem Schritt berücksichtigt. Zudem werden die FAIR-Prinzipien als grundlegende Gütekriterien in Text 2 erläutert und genau erörtert, was dies im Kontext von big-data und digitalen Verhaltensdaten bedeutet und welche Implikationen diese für das Datenmanagement sowie die Aufbereitung bedeutet. Darüber hinaus sind alle nötigen Dateien für die einzelnen Prozessschritte zugänglich und können anhand der ausführlichen Beschreibungen sowie der Schritt-für-Schritt-Anleitung (siehe Brentel et al. 2021; Brentel 2021a; Brentel 2021b; Brentel & Fortmann 2021 [Begleitpublikationen Daten 1, 2, 3 und 4]) nachvollzogen und reproduziert werden. Die konzeptionelle Strukturierung und Aufbereitung berücksichtigt eine Interoperabilität, die über den Schwerpunkt des Forschungsinteresses dieser Dissertation hinausgeht und die Sekundärnutzung der Daten für andere Forschungsinteressen ermöglicht³⁴.

Durch die transparente Bereitstellung der digitale Prozesskette können zum einen Forschende auch ohne Vorkenntnisse an das Arbeiten mit digitalen Verhaltensdaten und big-data herangeführt werden und „notwendige Entscheidungen in der Datensammlung und -aufbereitung bewusst“ (Jungherr et al. 2018, 257) gemacht werden, um vor einem „naiven Empirismus“ (Jungherr et al. 2018, 256) im Umgang mit big-data und digitalen Verhaltensdaten wie denen der *IntermediaPlus (2014-2016)* zu schützen. Die „Daten fall so nicht mehr „analysefertig“ aus einer gleichsam magischen Black-Box, sondern sind das Ergebnis eines auf für andere transparenten Entscheidungs- und Abwägungsprozesses, der Forschern die Bedingtheit und Begrenztheit ihrer Messung bewusst macht“ (Jungherr et al. 2018, 257). Zum anderen kann die Datenaufbereitung der Media-Analyse *IntermediaPlus* als repräsentativste und größte Erhebung von Medienkonsum in Deutschland (vgl. Stark & Kist 2020) mit dieser digitalen Prozesskette sehr Zeit- und Ressourcen-effizient fortgeführt bzw. ergänzt werden. Lediglich nach 2016 neu erfasste Onlinemedienangebote der Media-Analyse müssen der konzeptionellen Struktur entsprechend aktiv in die Harmonisierungspläne eingepflegt werden. Ein weiterer Arbeitsprozess für ein Fortschreiben der Zeitreihe liegt im Import der Daten zu *CharmStats*, was auf Grund der Schritt-für-Schritt-Anleitungen und die Handhabung der

³⁴ So sind beispielsweise stärker soziologisch ausgerichtete Fragestellungen denkbar, welche den Besitz von Befragten, der Haushalte sowie das Freizeitverhalten in den Blick nehmen. Geographische Kenngrößen zu Gemeindegrößen u.ä. sind ebenso in den Daten enthalten.

Software gut mit studentischen Hilfskräften zu bewältigen ist. Demnach kann die digitale Prozesskette für den Aufbau einer zuverlässigen Datenquelle für das Monitoring von Medienkonsumverhalten (z.B. die Entwicklung von Publikumsfragmentierung) herangezogen werden. Sie kann aber auch für den Aufbau von einer digitalen Dateninfrastruktur für das Fach der Kommunikationswissenschaft weiterentwickelt und ausgebaut werden. Denn auch andere Datenquellen können mit dieser (technischen) Systematik aufbereitet und in eine Erweiterung des Datenbanksystems für *charmsana* eingepflegt werden. Ein (kuratiertes) open-source Datenbanksystem mit einheitlicher technischer Systematik würde das Teilen von Forschungsdaten und deren Nachnutzung fördern (vgl. Strippel 2021). So steigert „das Teilen von Forschungsdaten auch die Effizienz der Forschung, da mit denselben Daten idealerweise unterschiedliche Fragen beantwortet werden können (durch zusätzliche bzw. weitergehende Analysen). Dies ist nicht nur forschungsökonomisch, sondern hat auch eine ethische Komponente“ (Peter et al. 2020, 4). Gerade im Kontext von digitalen Verhaltensdaten, big-data und Computational Social Science ist das Teilen von Forschungsdaten und die Bereitstellung von Untersuchungsmaterialien relevant, um Qualität zu sichern und das Potenzial digitaler Daten und Computational Methoden ausschöpfen zu können (vgl. Jungherr et al. 2018).

4.3 Text 3: The German online media market: Online-born information offerings and their audiences – A shift towards digital inequalities? (Kampes & Brentel 2020)

Diese erste empirische Studie des Kumulus (Text 3) widmet sich der Forschungsfrage, wie und von wem politische Information online genutzt wird. Zudem erörtert diese Einzelstudie, ob eine ungleiche Nutzung online eine Reproduktion aus dem offline-Kontext ist oder sich online neue Klüfte zeigen. Folglich nähert sich diese erste empirische Studie dem Fragmentierungsphänomen auf der Mikroebene, wenngleich die Makroebene durch den medienökonomischen Teil der Einzelschrift der Co-Autorin berücksichtigt wird.

4.3.1 Forschungslücke

Bislang gibt es keine Studien für Deutschland, welche die politische Informationsnutzung im Kontext der Onlinenutzung mit dem Detailgrad von Einzelangeboten und Umfang mehrerer tausend Onlineseiten, deren Auswahl unabhängig ihrer Reichweite ist, untersucht. Der Detailgrad für die Untersuchung politischer Informationsnutzung online ist bisher auf Grund fehlender Daten lediglich bis zur Ebene der Gesamtangebote online möglich und oftmals auf

die online-Pendants der Tageszeitungen sowie politischer Prizmazine begrenzt. Eine Differenzierung nach der Nutzung der thematischen Unterseiten, wie beispielsweise dem Politik-Ressort und Sport-Ressort, also dem Einzelangebot online mit politischem Inhalt (z.B. faz.net / Politik) und den Unterseiten zum Thema Sport (z.B. faz.net / Sport), erfolgte daher nicht. Thematisch orientierte, ungleiche Nutzung konnte entsprechend nicht abgebildet werden. Oft wurde daher bislang angenommen, wer das Gesamtangebot eines Online-Pendants von politischen Zeitschriften oder Tageszeitungen nutzt, würde politische Information konsumieren und die dadurch entstehende Unschärfe hingenommen. Außerdem wurden im Kontext der Digital Divide Forschung vorrangig nutzerzentrierte Analysen durchgeführt und die Angebotsseite im Sinne einer zuschauerorientierten Perspektive auf die Online-Mediennutzung vernachlässigt. Obwohl angebotsseitige Effekte auf die Online-Mediennutzung im Sinne des Digital Divide vermutet werden (vgl. Bonfadelli 2002): So wird bei online-born Medienangeboten eine stärkere Ausrichtung an der Maximierung von Klicks und ein „commercially-dominated product portfolio“ (Kampes & Brentel 2020, 9 [Text 3]) gesehen, welches stark unterhaltungsorientierte statt informationsorientierte Inhalte bereitstellt und Wissensklüfte verstärken kann. Darüber hinaus ist unklar, ob es sich bei digitalen Klüften um die Herausbildung neuer Klüfte, der Reproduktion oder einer Verstärkung bereits aus dem Offlinekontext der Mediennutzung bekannter Klüfte, handelt (van Deursen & van Dijk 2014, 521f.; Wei & Hindman 2011, 217f.). Dies ist jedoch zur Evaluation, Einordnung und Folgenabschätzung von Klüften im Onlinekontext elementar

4.3.2 Ergebnisse

Die Einzelstudie (Text 3) umfasst einen Vergleich der Mediennutzung sowie die deskriptive Beschreibung des Publikums von zwei typischen Fällen für online-born Online-Einzelangebote und solche Online-Einzelangeboten mit Ursprung im Offlinemedienmarkt, die aus der Logik des Medienmarktes hergeleitet wurden. Das Genre „Digitales“ steht stellvertretend für ein online-born Produktportfolio mit großer Reichweite und eher unterhaltungs-orientiertem Fokus. Im Gegensatz dazu steht das Genre „Politik“ mit ausschließlich aus dem Offlinekontext stammende Online-Einzelangeboten und für Inhalte mit stark informations-orientiertem Fokus steht, welche oftmals journalistischen Maximen folgen (vgl. Cunningham 2018, 38, 43; Egan & Tsao 2015, 4; Hallin & Mancini 2004, 277; Kiefer & Steininger 2014, 24, 26f; Sjøvaag & Kvalheim 2019, 302; Dahlberg 2007, 838; siehe Text 3). Dadurch wird evaluiert, inwiefern digitale Ungleichen singular digital sind oder sich eher aus dem offline Kontext reproduzieren. Dies erfolgt zum einen unter Rückgriff auf die Literatur zu Klüften in der Offlinemediennutzung, zum anderen durch den Vergleich der beiden Genres „Digitales“ als Stellvertreter für online-born Onlineangebote und „Politik“ als typischer Fall der Einzelangebote mit offline-Ursprung.

Das Ergebnis der Analysen, welche für den Vergleich der Onlinemediennutzung im Sinne des Digital Divides einem innovativen methodischem Ansatz Bonfadelli's (2002) folgt, zeigt überwiegend eine Reproduktion aus dem Offlinekontext bezüglich der unterschiedlichen Onlinemediennutzung (dem second digital divide), insbesondere für das Geschlecht: Männer nutzen Onlinemedien mehr und das auch mit einer höheren Nutzungshäufigkeit. Dabei ist die Nutzung des Genres „Politik“ für männliche Nutzende durchweg auf einem höheren Niveau als für „Digitales“. Hinsichtlich der Bildung konnte eher eine Verstärkung der bekannten, ungleichen Nutzungsmuster für den Onlinekontext festgestellt werden. Menschen mit höherer Bildung, und unter diesem Gesichtspunkt einem höheren sozio-ökonomischen Status, nutzen Einzelangebote des Genres „Politik“ eher und gehörten vermehrt zum Kreis der regelmäßigen Nutzenden (definiert als durchschnittlich wöchentliche Nutzung). Eine „neue“, digitale Kluft bzw. Ungleichheit der Mediennutzung online wurde hinsichtlich der Altersstrukturen der Nutzenden festgestellt. So nutzen auffällige viele junge Menschen politische Information, auch als regelmäßige Nutzende, was im Kontext von Print und TV ungewöhnlich ist (siehe Text 3). Denn für junge Menschen wurde bislang bei der Nutzung von Offline-Medien eher eine unterhaltungsorientierte Mediennutzung festgestellt, während ältere Menschen eher zu einer Mediennutzung zur politischen Information neigen und politische Inhalte häufiger nutzen (vgl. Bonfadelli 2002, 80; Peiser & Jandura 2015).

4.3.3 Limitation

Die Analysen basieren auf den Daten der *IntermediaPlus (2014-2016)* für das Jahr 2016 und wurden hinsichtlich der Onlinenutzung exemplarisch auf die beiden Genre-Kategorien „Digitales“ und „Politik“ begrenzt (siehe Text 3). Zwar konnte so die spezifisch politische Informationsnutzung online deskriptiv beschrieben und analysiert werden, doch für eine Einordnung welche Rolle politische Information in der Onlinemediennutzung spielen, reicht der Vergleich mit nur einem weiteren Genre nicht aus. Auch ist die Analyse auf die Momentaufnahme von 2016, was bereits ein paar Jahre zurück liegt, begrenzt. Entsprechend kann die politische Onlinenutzung nach 2016 anders aussehen, neue Ungleichheiten in der Onlinemediennutzung auftreten oder sich weiter verstärkt haben. Folglich: “more exploration is needed (...) thematically for the media use (including more genres) and time-related (including recent market data) to analyse digital inequalities in information supply and use” (Kampes & Brentel 2020, 29 [Text 3]). Zudem wäre für die Evaluation, ob neue, digitale Ungleichheiten in der Mediennutzung auftreten, eine ergänzende, crossmediale Betrachtung notwendig. Da der Aufbau von Wissensbeständen und gemeinsamen Themen nicht auf eine Mediengattung begrenzt ist, können vermeintliche Unterschiede in der Onlinemediennutzung durch gemeinsam genutzte Offlinemedien überbrückt werden oder bei crossmedialer

Betrachtung andere Klüfte in der Mediennutzung zu Tage treten. Desweiteren handelt es sich hier um die Betrachtung der Onlinemediennutzung, während in der Digital Divide Forschung ein großer Fokus auf den möglichen Folgen ungleicher bzw. fragmentierter Nutzung, den resultierenden Klüften im Wissen, liegt. Diese Dimension der Folgen von Fragmentierung wird in der vorliegenden Einzelschrift außenvor gelassen, da Fokus des hier vorliegenden Dissertationsprojekts die Nutzung von Onlinemedien ist. Die Analyse möglicher Folgen von Fragmentierung ist jedoch zur Evaluation der „consequences of those digital inequalities identified for possibly resulting social exclusion and social as well as democratic divides“ (Kampes & Brentel 2020, 29 [Text3]) notwendig.

4.3.4 Ertrag

Die Einzelschrift verfolgt einen innovativen und interdisziplinären Ansatz, um dem komplexen Untersuchungsgegenstand der (fragmentierten) Online-Mediennutzung gerecht zu werden (vgl. Strippel et al. 2018, 16f., 21f.). So wird die Onlinemediennutzung zwar im Sinne des methodologischen Individualismus auf der Mikroebene betrachtet, der Kontext der Medienstruktur jedoch mitgedacht und im medienökonomisch-orientierten Teil des Textes analysiert. Zudem gibt die Studie erste Aufschlüsse zur politischen Informationsnutzung, welche besonders relevant für die Konstitution von Öffentlichkeit ist und am meisten unter Verdacht steht fragmentiert zu sein. So zeigen Onlinemediennutzungsmuster für politische Information ein Bild, welches aus dem Kontext der TV und Zeitungsnutzung bekannt ist: informations-orientierte Medien werden eher von gut gebildeten Menschen genutzt, während Personen mit niedriger Formalbildung eher auf unterhaltungs-orientierte Medienangebote zurückgreifen (vgl. Wei & Hindman 2011, 222; Bonfadelli 2002, 69; siehe Text 3). Im Onlinekontext scheinen sich diese Nutzungsmuster verstärkt zu haben, was eine immer weiter voranschreitende Vertiefung von Wissensklüften bedeuten kann und durch weitere Studien untersucht werden sollte, auch hinsichtlich Intersektionalität von Bildung und anderen individuellen Merkmalen von Nutzenden. Ansätze wie der des „digital capital“ von Ragnedda und Kollegen (2020), welche Bildung und Medienkompetenz stark mit einbeziehen, scheinen vielversprechend zur Erklärung der Onlinenutzung in Deutschland.

Die Ergebnisse dieser Einzelschrift zeigen auch, ein Monitoring für Publikumsfragmentierung online im Kontext politischer Informationsnutzung, gerade für die Gruppe der Nutzenden mit geringer Formalbildung, wäre wichtig. Vor allem wenn es um mögliche Folgen dieser Form von ungleicher bzw. fragmentierter Nutzung wie der im Digital Divide angenommenen Vergrößerung der Wissensklüfte geht. Dennoch handelt es sich um kein neues Phänomen von Onlinemedien, sodass geprüft werden sollte, inwiefern Maßnahmen des Offline-Medienmarktes auch für den Online-Medienmarkt angewendet werden können, um einer

Vertiefung von Wissensklüften und einem daraus möglicherweise resultierenden „democratic divide“ (Norris 2001) entgegenzuwirken. Im Kontext demokratischer Gesellschaften und der Informationsfunktion von Massenmedien ist der hohe Anteil junger Menschen im Publikum von Onlineeinzelangeboten des Genres „Politik“ positiv zu bewerten, da deren Inhalte einen informations-orientierten Fokus haben und eher journalistischen Maximen folgen. Vertiefende Studien zur politischen Informationsnutzung junger Menschen, nicht nur für Social Media, wären darauf aufbauend wünschenswert.

4.4 Text 4: Online media consumption in Germany: An analysis of German mass communication online (Brentel, forthcoming)

In dieser empirischen Einzelschrift wird der Kernfrage, inwiefern Onlinemedien gemeinsam genutzt werden, nachgegangen. Es wird anhand einer sozialen Netzwerkanalyse sowie einer innovativen Operationalisierung von Mediennutzungsmustern in einem explorativen Untersuchungsdesign analysiert, ob es Teilpublika im Gesamtpublikum gibt und wenn ja, wie beziehungsweise was diese konstituiert. Sprich, welche Faktoren sich für voneinander abweichende Mediennutzungsmuster identifizieren lassen und Publikumsfragmentierung online in Deutschland determinieren.

4.4.1 Forschungslücke

Wie bereits in den Limitationen zu Text 3 dargelegt, ist zur Bewertung der Rolle von politischer Informationsnutzung online eine Ausweitung der Analyse auf weitere Genre notwendig (vgl. auch Neuberger & Lobigs 2010, 132). Zudem fehlt allgemein eine umfassende publikumszentrierte Betrachtung von Mediennutzungsverhalten online in Deutschland für eine Bestandsaufnahme der Publikumsfragmentierung online. Zwar wird mit Aufkommen des Internets auf Grund der Besonderheiten im Onlinemedienmarkt sowie neuen Möglichkeiten in der Mediennutzung eine Zersplitterung des Gesamtpublikums angenommen, konnte bisher jedoch nicht empirisch belegt werden (Stark 2013; Stark et al. 2021a; Rau & Stier 2019; Bruns 2019b; Mahrt 2020; vgl. Webster & Ksiazek 2012; Dubois & Blank 2018; Taneja et al. 2012; Mukerjee et al. 2018a). In Deutschland bleiben, auch auf Grund fehlender Daten, umfassende Studien zu überlappenden Mediennutzungsmustern von Massenmedien online aus, welche die Vielzahl an Onlinemedienangeboten, deren thematische Ausdifferenziertheit sowie den hohen Grad an Individualisierung in der Mediennutzung modellieren können und entsprechend analysieren. Demnach mangelt es an einer Studie, welche den ersten Schritt der Fragmentierungskette (vgl. Stark 2013, 199ff.; Handel 2000, 12), die gemeinsame,

überlappende Mediennutzung vs. nicht-überlappende Mediennutzung auf Makroebene mit einer publikumszentrierten Perspektive, im Onlinekontext untersucht, um evaluieren zu können, inwiefern die vermutete Fragmentierung auf dieser Dimension bereits vorliegt oder nicht. Folglich ist auch unklar, welche Faktoren der Onlinemediennutzung tatsächlich fragmentierte Mediennutzungsmuster erklären können.

4.4.2 Ergebnisse

Das Konzept der „audience duplication“, definiert als “the degree to which two media outlets share audience members” (Ksiazek 2011, 237), wird hier als ein Gradmesser von Publikumsfragmentierung herangezogen. Keine Überlappung von Publika, also in individuellen Mediennutzungsmustern online, entspräche einer totalen Publikumsfragmentierung. Während eine große Überlappung in den Publika von Einzelangeboten online eher gegen bzw. zumindest eine geringe Publikumsfragmentierung online steht. Basierend auf Webster und Ksiazek (2012) wird anhand der Methode der sozialen Netzwerkanalysen mit einem publikumszentrierten Verständnis von Fragmentierung die Publikumsüberschneidung der Einzelangebote online für 2016 in Deutschland mit den Daten der *IntermediaPlus (2014-2016)* untersucht. Dabei wird ein explorativer Ansatz gewählt. Dem Vorschlag von Mukerjee (2021) folgend werden keine vorformulierten Hypothesen getestet, sondern analysiert, „if the structure of audience behavior that emerges, is in line with what we would expect if these theoretical considerations were valid” (ebd. 2021a, 18). Das heißt die Analysen fußen auf dem theoretischen und methodologischen Ansatz der News Reading Publics. Zusätzlich wird ein neuer, innovativer methodischer Ansatz zur Aggregation individueller Mediennutzungsmuster im Rahmen der sozialen Netzwerkanalyse online verwendet (vgl. Lietz 2020; siehe Text 4). Bei diesem wird die Anzahl der genutzten Onlinemedien des jeweiligen Onlinerepertoires einzelner Nutzenden sowie die Nutzungshäufigkeit berücksichtigt. Überlappende Publikumsstrukturen werden durch die Community Detection, einer der induktiven Clusteranalyse ähnlichen Methode im Rahmen der sozialen Netzwerkanalyse, identifiziert und auf Gemeinsamkeiten sowie Unterschiede in den soziodemographischen Merkmalen der Nutzenden sowie dem thematischen Genre der genutzten Onlineangebote (Einzelangebote) untersucht.

Auch mit der größeren Bandbreite an Onlinemedienangeboten und dem deutlich größeren Detailgrad der gemessenen Mediennutzung online auf Ebene der Einzelangebote konnte eine große Überlappung in den Publika beobachtet werden (vgl. Webster & Ksiazek 2012). Es konnten keine vom Gesamtpublikum vollständig abgetrennten Teilpublika identifiziert werden. Jedoch ließen sich zwei Teilpublika unterscheiden, welche zwar Überschneidungen bei einigen Nutzenden aufweisen, aber durch voneinander abweichende Nutzungsmuster

gekennzeichnet sind. So wurden insgesamt drei Publikumsgruppen identifiziert: das exklusiv blaue³⁵ Teilpublikum (6 Prozent der Nutzenden, 54 Prozent der genutzten Einzelangebote), das exklusiv gelbe Teilpublikum (51 Prozent der Nutzenden, 46 Prozent der genutzten Einzelangebote) sowie die cross-over Nutzergruppe (43 Prozent der Nutzenden), bei welcher Nutzende sowohl zum blauen als auch dem gelben Teilpublikum gehören.³⁶ Trennend für die beiden identifizierten Teilpublika scheint dabei die Nutzung spezifischer Inhalte: die theoretische Erklärung von issue publics, also eine stark interessen geleitete Mediennutzung, welche zu eher themenspezifischen Mediennutzungsmustern und daraus resultierenden Teilpublika führt, scheint daher am vielversprechendsten zu sein, um Tendenzen einer Publikumsfragmentierung online zu erklären (vgl. Mukerjee 2021a; siehe Text 4). So schreiben auch Waldherr und Kollegen (2021, 164), „[i]n the process of digitalization, the public sphere has transformed into many smaller, though loosely coupled, public spheres – often evolving around specific issues.“ Doch auch soziodemographische Merkmale, wie das Geschlecht, Bildung, Alter und Einkommen lassen sich nach wie vor als Determinanten von Mediennutzungsmustern erkennen (siehe Text 4).

Das *exklusiv gelbe Teilpublikum* zeichnet sich durch einen starken thematischen Schwerpunkt der genutzten Onlinemedienangebote aus; lediglich Einzelangebote des Genres „Politik“, „Digitales“, „Homepage“, „Kultur“ und „Lifestyle“ werden genutzt. Wobei alle Einzelangebote des Genres „Politik“ ausschließlich im gelben Teilpublikum vertreten sind und gemeinsam mit Onlineangeboten des Genre „Digitales“ die genutzten Medienangebote dominieren. Zudem sind eher junge Nutzende, gerade 14- bis 29-Jährige, im exklusiv gelben Teilpublikum vertreten³⁷.

Das *exklusiv blaue Teilpublikum* zeichnet sich dadurch aus, dass kein klarer inhaltlicher Schwerpunkt erkennbar ist; es sind Einzelangebote nahezu aller 27 Genre-Kategorien (außer „Politik“) vertreten, vor allem das Genre „Homepage“ charakterisiert die genutzten Einzelangebote dieses Teilpublikums (siehe Text 4). Auffällig ist zudem, dass Nutzende des exklusiv blauen Teilpublikums überwiegend weiblich sind. Doch auch mittlere Bildung und ein Einkommen bis 1.500 Euro scheinen die exklusiv blaue Publikumsgruppe auszuzeichnen.

Die *cross-over Nutzergruppe* ist vor allem durch Nutzende mit hoher Formalbildung sowie einem Alter von 25 bis 29 Jahren gekennzeichnet. Zudem gehören Nutzende mit einem

³⁵ Die Farben beziehen sich auf die Farben der Knotenpunkte in der Visualisierung des Netzwerkes der Medienprojektion und zeigen die Zugehörigkeit zu den unterschiedlichen, durch die Community Detection identifizierten Publikumscluster an (siehe Text 4, Abbildung 2). Die Farben haben dabei keine inhaltliche Bedeutung.

³⁶ Einzelmedienangebote sind auf Grund der Methodik exklusiv dem blauen oder gelben Teilpublikum zugeordnet.

³⁷ Bis zu einem Alter von 54 sind Nutzende im Vergleich zur Verteilung in den Daten überproportional häufiger im exklusiv gelben Teilpublikum vertreten, der Unterschied sinkt jedoch mit steigendem Alter (siehe Text 3, Abbildung 6).

Einkommen über 3.000 Euro der cross-over Nutzergruppe überproportional häufig an und sind unter 50 Jahre alt. Insgesamt spielt politische Information aber eine große Rolle: So wird das „Politik“-Genre auch intensiv von den cross-over Nutzenden genutzt; selbst im exklusiv blauen Teilpublikum sind die Homepage-Einzelangebote der Tageszeitungen, politischer Zeitschriften und online-born Angebote, mit dem Selbstanspruch eines Überblicksmediums mit Informationen von allgemeinen Belangen, wie t-online, vertreten und gehören zu den meist genutzten Angeboten.

4.4.3 Limitation

Die Vergleichbarkeit der Ergebnisse mit anderen Studien im Forschungsfeld ist auf Grund der Anwendung einer neuen Operationalisierung von Mediennutzungsmustern und daraus resultierender Publikumsüberschneidung online eingeschränkt. Anhand der Fortführung der methodischen Debatte, weiterer Erprobung sowie Replikation muss diese Operationalisierung im Kontext der Mediennutzung immer mehr validiert werden. Zudem handelt es sich bei der untersuchten Onlinemediennutzung lediglich um einen zeitlichen Ausschnitt von 2016. Aussagen zur Publikumsfragmentierung online sind entsprechend auf eine Momentaufnahme begrenzt und können keine Prozesshaftigkeit abbilden. Zudem handelt es sich bei 2016 um eine weiter zurückliegende Momentaufnahme und Mediennutzungsmuster können sich im dynamischen Umfeld des Internets bereits stark gewandelt haben, was die Aussagekraft der Ergebnisse für den Zeitpunkt der Veröffentlichung der Studie limitiert. Wenngleich es sich bei der *IntermediaPlus (2014-2016)* um die umfassendste, verfügbare Datengrundlage des Medienkonsumverhaltens in Deutschland handelt (vgl. Strippel 2021; Stark & Kist 2020, 1146; Brentel & Winters 2021 [Text 1]), ist zudem nur ein Ausschnitt des Medienmarktes (online) abgedeckt: öffentlich-rechtliche Medienangebote, für welche eine große integrative Leistung angenommen wird (vgl. Rau & Stier 2019, 406f.; Fletcher & Nielsen 2017, 483; Handel 2000, 53), fehlen in der Betrachtung der Einzelstudie gänzlich; ebenso Nutzung von Blogs oder Nutzung von publizistischen Inhalten über Social Media. Auch handelt es sich um eine Auswahl möglicher Determinanten von Fragmentierung und Mediennutzung (online), welche untersucht wurden; eine Ausweitung auf weitere möglicher Erklärungsfaktoren, wie das politische Interesse (vgl. Mahrt 2019), verfügbare Zeit für Mediennutzung (z.B. „Attitudes“ des Zuschauermodells nach Webster 2018 oder „Viewer Availability“ des „Program Choice“-Models von Webster & Wakshlag 1983) steht noch aus. „Nevertheless, this is a first explanatory study on online news consumption patterns in Germany with this granularity and more research is needed to draw sound conclusions about audience fragmentation online in Germany.“ (Brentel, *forthcoming*, 26, [Text 4]). So sind zum einen Studien, welche weitere Teilaspekte der Mediennutzungsmuster behandelt und dabei weiter ins Detail gehen nötig:

beispielsweise mit einem Fokus auf die Nutzungsintensität, um zu sehen, wie sich das Bild der Publikumsüberschneidung verändert und ob sich bei Betrachtung einzelner Publikumscluster noch weitere Gruppierungen von Nutzungsmustern identifizieren lassen, welche sich zu Nutzungstypen verdichten lassen. Zum anderen ist eine Ausweitung auf weitere Dimensionen von Publikumsfragmentierung nötig, wie Analysen der rezipierten Inhalte, Wahrnehmung und Wirkung dieser.

4.4.4 Ertrag

Neben den konkreten Erkenntnissen über den Stand der Publikumsfragmentierung online in Deutschland für 2016 sowie der Konstitution der thematisch orientierten Teilpublika („Digitales“ und „Politik“ vs. eher unspezifische Informationsnutzung mit „Homepage“ und einem breiten Mix aus thematischen Genres) und den Nutzungsgruppen, liegt der Ertrag dieser Einzelschrift in dem hohen Detailgrad der Messung von Mediennutzung, was die Reliabilität dieser Ergebnisse verbessert. Es handelt sich bei der Einzelschrift (Text 4) um eine umfassende Studie für das privatwirtschaftliche Marktsegment der Onlinemedien mit einem bislang international einzigartig hohem Detailgrad in der Messung der Onlinenutzung auf Ebene der Einzelangebote. Dadurch wird ein möglichst umfangreiches Bild der Onlinemediennutzung dieses Marktsegmentes realisiert. So können nun weiterführende Studien auf Mikroebene anschließen, welche den Fokus auf Ausschnitte des Gesamtpublikums legen (vgl. Handel 2000, 30); Mittels der Ergebnisse der Einzelstudie (Text 4) können interessante Gruppen identifiziert werden und Ergebnisse von Teilpublika, wie beispielsweise von Text 3, sowie von weiteren spezifischen Fallstudien eingeordnet werden und in Beziehung zueinander gesetzt werden.

Darüber hinaus leistet die Einzelschrift (Text 4) einen Beitrag zur methodischen Diskussion bezüglich der Aggregation von individuellen Mediennutzungsmustern beziehungsweise der Operationalisierung von Publikumsüberschneidung (siehe Mukerjee et al. 2018a, 2018b und Webster & Taneja 2018), welcher über die vorliegende Einzelschrift (Text 4) und die hier vorliegende Dissertation hinausreicht. Ein weiterer Ertrag für das Fach, insbesondere die Onlinemediennutzungsforschung, ist die Anwendung und damit Testung des neuen theoretischen Konzepts von Mukerjee (2021a) zur Analyse von (Online)Mediennutzung der News Reading Publics. Dieser Ansatz wurde dazu entwickelt, den Vergleich von Einzelstudien zu Mediennutzung online zuzulassen, auch länderübergreifend (vgl. Mukerjee 2021a). Über eine Steigerung der Vergleichbarkeit von Einzelstudien, auch im internationalen Kontext, kann die Theoriebildung vorangetrieben und ein erhöhter Erkenntniswert solcher Fallstudien zu (Online-)Mediennutzung für das Fach gefördert werden.

5 Ertrag des Kumulus, Diskussion und Ausblick

„Ob Einzelpersonen oder Gruppen im Mittelpunkt der Betrachtung stehen, es handelt sich lediglich um zwei unterschiedliche Betrachtungsweisen von Fragmentierung, die sich nicht ausschließen.“ (Handel 2000, 30)

Im Rahmen der vorliegenden Dissertation wurde sowohl der Individual- und Gruppenperspektive als auch Strukturen des Medienangebotes zur Erörterung der Publikumsfragmentierung online Rechnung getragen, um der Komplexität der Mediennutzung online (vgl. Waldherr et al. 2021; Strippel et al. 2018, 16f., 21f.) und dem Fragmentierungsphänomens gerecht zu werden. Durch die Verbindung der nutzerzentrierten Forschung zum Digital Divide durch Text 3 und den publikumszentrierten Ansätzen des Forschungsstrangs der ‚audience duplication‘ in Text 4 im Dissertationsprojekt wurden beide Perspektiven unter dem Dach des Mikro-Makro-Modells in die Analysen miteinbezogen und in Beziehung zueinander gesetzt statt „in der Diskussion und der Literatur unverbunden nebeneinander“ (Handel 2000, 30) zu stehen. Zudem wurden in jeder der empirischen Einzelstudien Elemente der jeweils anderen Perspektive aufgenommen: So wurde in Text 3, der eher nutzerzentrierten Studie, ein interdisziplinärer Ansatz verfolgt, bei welchem Makrostrukturen des Medienangebots berücksichtigt wurden und die individuelle Mediennutzung auf Gruppenebene für den Vergleich der Klüfte in der Nutzung nach soziodemographischen Merkmalen auszuwerten. Der Fokus lag vor allem auf der Nutzung politischer Information. In Text 4 wurde hingegen eher eine Makro-Perspektive und ein publikumszentrierter Ansatz gewählt, um (fragmentierte) Teilpublika online über verschiedene thematische Genre hinweg identifizieren und untersuchen zu können. Auch hier wurden die Strukturen des Medienangebotes berücksichtigt, dieses Mal als Teil der Analyse (siehe auch Anhang I). Die individuelle Mediennutzung stellte die Grundlage für die Identifikation von überlappenden Mediennutzungsmustern massenmedialer, kommerzieller Onlineangebote sowie der daraus resultierenden Teilpublika online dar – und das auf einem bisher nicht da gewesenen Detailgrad.

Die Anwendung der Sozialtheorie als Rahmung für die Untersuchung des Fragmentierungsphänomens online, lässt eine gewisse Offenheit für verschiedene theoretische Ansätze zur Medienselektion. So können verschiedene theoretische Erkenntnisse, Konzepte, Perspektiven und Erklärungen für die Analysen in Beziehung zueinander gesetzt werden und in den eher explorativen Charakter der empirischen Einzelstudien (Text 3 und Text 4) einfließen. Gleichzeitig ist die Anschlussfähigkeit, die Interoperabilität, sowohl theoretisch als auch empirisch für (erweiternde) Folgestudien, beispielsweise zu weiteren Dimensionen der Publikumsfragmentierung (online) gewährleistet.

Durch die Datenarbeit im ersten Teil der Dissertation (Text 1 und Text 2, die Datenpublikation sowie deren Begleitpublikationen) wurden die „neuen Möglichkeiten, die Komplexität digitaler Mediennutzung adäquat zu erfassen“ (Strippel et al. 2018, 17) ausgeschöpft und der *Längsschnitt der IntermediaPlus (2014-2016)* zur wissenschaftlichen Nachnutzung bereitgestellt. Mit dem entwickelten Datenmanagementsystem wurde die Aufbereitung der Rohdaten der Media-Analyse *IntermediaPlus* deutlich vereinfacht und beschleunigt. Die Daten können beliebig erweitert werden: Eine Aktualisierung der Datenreihe um neuere Jahre ist ebenso möglich, wie die Aufbereitung der Rohdaten zur Mediennutzung im Offline-Bereich, um das Fragmentierungsphänomen im crossmedialen Kontext zu analysieren. Damit steht nun für wissenschaftliche Forschung anhand des Umfangs der erhobenen Online-Medien sowie dem hohen Detailgrad der erfassten Nutzung dieser eine deutlich besser geeignete Datenquelle, für die Analyse des Fragmentierungsphänomens online in Deutschland zur Verfügung. Die Replikation sowie Fortführung der im Rahmen des Dissertationsprojekts begonnen Analysen wurde dadurch genauso ermöglicht wie die Nutzung der Daten für Forschung mit alternativen Konzepten, theoretischen Ansätzen und Perspektiven auf sowie zu weitere Betrachtungsebenen und Dimensionen des Fragmentierungsphänomens, um dieses besser ergründen zu können.

5.1 Erkenntnismehrwert des Kumulus und Diskussion

Ziel des hier vorliegenden Dissertationsprojektes war es den ersten Teil des Fragmentierungsprozesses, gemeint ist hiermit die (gemeinsame) Mediennutzung, für kommerzielle, massenmediale Onlinemedien in Deutschland zu untersuchen. Nun liegt anhand des Kumulus für Deutschland eine Analyse massenmedialer Onlinenutzung vor. Sie fehlte in diesem Umfang und der thematischen Breite von Medienangeboten sowie der Detailtiefe auf Ebene von Einzelangeboten bisher, war aber dringend nötig (vgl. Mahrt 2019, 172ff., 178). Denn nur so kann die vielfach angenommenen Publikumsfragmentierung im Online-Kontext exakter erfasst, dadurch besser analysiert und gesellschaftliche Implikationen evaluiert werden (vgl. Webster & Ksiazek 2012, 51; Handel 2000, 35). Dies war nur durch die Erschließung der Datenquelle der MA *IntermediaPlus* möglich, welche die Basis für die empirischen Studien der vorliegenden Doktorarbeit ist, aber auch für weitere Forschung sein kann. Zusätzlich konnte ein Beitrag zur Überwindung der beiden übergreifenden Limitationen, welche in der Forschungsliteratur zum Zeitpunkt des Dissertationsprojekts ausgemacht werden konnten, geleistet werden:

1) Mithilfe der Identifikation sowie Aufbereitung der *IntermediaPlus (2014-2016)* wurde eine geeigneten Datenbasis zur Untersuchung von fragmentierter Onlinemediennutzung für die wissenschaftliche Forschung bereitgestellt. Denn sie hat den nötigen Umfang der erfassten

Onlinemedien, Detailgrad sowie Zuverlässigkeit in der Erfassung der Onlinemediennutzung – sichergestellt durch den Methodenmix von Verhaltens- und Befragungsdaten. Die Möglichkeit zur Nachnutzung für weitere Studien dieser einmaligen Datenquelle ist wegen der Art und Weise der Aufbereitung sowie deren umfassende Dokumentation unter Einhaltung der FAIR-Prinzipien gewährleistet. Darüber hinaus wurde mit dem Datenmanagementsystem und der dafür entwickelten digitalen Prozesskette ein Best-Practice für die Bereitstellung von Forschungsdaten unter Einhaltung der FAIR-Prinzipien (Wilkinson et al. 2016; Mons et al. 2017) sowie der DGPK-Leitlinien (Peter et al. 2020) entwickelt. Gerade für Verhaltensdaten und big-data fehlte dies bisher.

2) Auf Grund der Verbindung von Mikro- und Makro-Perspektive in diesem Kumulus konnten neue Erkenntnisse erlangt werden, die unterstreichen, wie wichtig die Verbindung beider Perspektiven für die Analysen von Publikumsfragmentierung online ist. Denn für sich genommen, können die Ergebnisse der eher nutzerzentrierten empirischen Einzelstudie (Text 3) fehlgedeutet werden oder im Fall der Studie der Publikumsüberschneidung mit einer eher Makro-Perspektive Fragen offenbleiben. Erst durch die Verbindung der Erkenntnisse der Analysen, die auf denselben Daten basieren, kann mit diesem Kumulus ein ganzheitlicheres Bild der Publikumsfragmentierung online in Deutschland für das Jahr 2016 gezeichnet werden:

So weiß man durch Text 4, dass die Onlinemedienangebote der thematischen Genre Digitales und Politik ein enges Teilpublikum sind. Ausgehend von einer Marktperspektive ist es überraschend, dass gerade die zwei sehr unterschiedlichen Genre (siehe Text 3) so stark überlappend genutzt werden und ein Teilpublikum bilden (Text 4). Auch aus der eher nutzerzentrierten Analyse bezüglich der Digital Divides in der Nutzung von Politik und Genre war nicht erkenntlich, dass es sich bei diesen Nutzenden oft um die gleichen handelt (Text 3). Durch die Makro-Perspektive ist eine Einschätzung zu Onlinenutzung politischer Information im gesamten Themenspektrum möglich sowie die Identifikation von Teilpublika und daraus abgeleitete Nutzergruppen. Anhand der Mikro-Perspektive auf die Nutzung der Genre Politik und Digitales in Text 3 klar, dass innerhalb des (gelben) Teilpublikums Unterschiede in der Nutzung, vor allem hinsichtlich der Regelmäßigkeit, sind. Demnach werden politische Informationsangebote des Genres Politik zwar von vielen genutzt – zumindest in dem erfassten drei Monatszeitraum und unregelmäßig –, bei regelmäßiger, also intensiverer Nutzung dann aber doch Klüfte anhand der Mediennutzung und entlang der soziodemographischen Merkmale innerhalb des Teilpublikums bestehen. Das sind alles Hinweise darauf, dass sich verschiedene Nutzertypen in den identifizierten Teilpublika bewegen und sich ein immer feinkörnigerer Blick in die Nutzergruppen lohnt, um eine Antwort auf die Frage zu finden „WER eigentlich fragmentiert ist“ (Handel 2000, 31). Dies zeigt, wie wichtig ein publikumszentrierter Ansatz auch im analytischen Vorgehen ist. Es macht auch

deutlich, welchen Mehrwert die Betrachtung der Mediennutzung des Gesamtpublikums für die Publikumsfragmentierung online ist, der für das Marktsegment der kommerziellen Onlinemedien in Text 4 erfolgte.

Der vorliegende Kumulus leistet einen wichtigen Beitrag zur Erforschung der Publikumsfragmentierung online, zu einer Bestimmung des aktuellen Stands bezüglich des ersten Schritts eines Fragmentierungsprozesses im Online-Kontext – der (überlappenden) Mediennutzung. Gleichzeitig liegt auf der Hand, dass weiterführende Forschung dringend nötig ist, um eine Publikumsfragmentierung online weiter und detaillierter zu ergründen. Eine Datengrundlage dafür wurde mit der *IntermediaPlus (2014-2016)* geschaffen, eine Erweiterung ist jedoch nötig, um die Limitationen bezüglich des Ausschnitts der in diesem Dissertationsprojekt betrachteten Mediennutzung zu überwinden. Eine Betrachtung des sich komplexen, sich entwickelnden Fragmentierungsprozesses im Zeitverlauf genauso nötig wie die Untersuchung weiterer Schritte, Ebenen und Dimensionen eines Fragmentierungsprozesses, um mit einem umfassenderen Bild daraus erwachsene gesellschaftliche Implikationen evaluieren zu können.

5.2 Ausblick

Fortführende Forschung zum Fragmentierungsprozess ist also unabdingbar und angesichts der andauernden wissenschaftlichen und gesellschaftlichen Debatte in einer immer pluraler werdenden Mediengesellschaft und den multiplen Optionen der Informationsgewinnung. Mit dem aus Text 4 generierten Wissen, sähe die Untersuchung potenzieller Digital Divides in der Nutzung anders aus – eine Replikation der Studie von Text 3, welche die Nutzergruppen (exklusiv blaues und gelbes Teilpublikum sowie die cross-over Nutzergruppe) unter der Berücksichtigung ihrer Markteigenschaften mit einander vergleicht und nach Unterschieden in Ihrer Nutzungsintensität sowie dem jeweiligen Anteil von politischer Information analysiert, sinnvoll. Die politische Information sollte dabei nicht nur auf das Politik-Genre begrenzt werden, sondern auch die Nutzung von oberflächlicheren Onlineangeboten mit politischen Informationen wie die Homepage der Tageszeitungen, E-Mail-Providern u.ä. (vgl. Kleinen-von Königslow 2016, 257f.). Zu vermuten wäre anhand der These des Digital Divides, welches sich auch in Einzelstudien widerspiegelte (siehe z.B. Zillien & Hargittai 2009), dass ein Divide von hochgebildeten vs. niedrig gebildeten besteht: Menschen mit höherer Formalbildung stehen im Verdacht (politische) Information online effizienter zu nutzen, eher Qualitätsmedien zu nutzen und einen größeren Umfang im Medienrepertoire zu haben sowie eine höhere Nutzungshäufigkeit zu zeigen. Hinsichtlich der intensiveren Nutzung politischer Informationsangebote sprechen die Ergebnisse von Text 3 bezüglich der Regelmäßigkeit der Nutzung des Politik-Genre auch im Gesamtkontext der Onlinenutzung noch dieser These. Für

die Untersuchung von Qualitätsunterschieden in den genutzten Onlinemedien sowie der konkreten Zusammensetzung der Medienrepertoires – sprich auch dem Anteil an eher oberflächlichen und sehr spezifischen politischen Informationsnutzung online – fehlen grundlegende Studien; Eine umfassende Untersuchung der Qualität der Berichterstattung von Onlinemedien in Deutschland steht noch aus. Erste Studien (siehe Kösters 2020) legen nahe, dass die Qualitätsverständnisse des offline Zeitungs- und Zeitschriftenmarktes nicht ohne weitere auf den Onlinekontext bezogen werden können. Studien zum politischen Informationsgehalt von online-borne Medienangebote fehlen gänzlich. Dies sollte von der kommunikationswissenschaftlichen Forschung schnellstens aufgearbeitet werden. Für eine ganzheitliche Erforschung ist die Marktperspektive unabdingbar, vor allem wenn es um Implikationen von Publikumsfragmentierung für Meinungsmacht (online) geht und damit verbunden medienpolitische Schlussfolgerungen. Dazu ist eine Fortführung der Analysen des Kumulus nötig, welche weitere Determinanten des Medienmarkts von Fragmentierung erforschen und beispielsweise die Richtung, in die das ergänzende Analysematerial zu Text 4 in Anhang I geht, weiterverfolgen.

Noch scheint Publikumsfragmentierung online und daraus ableitbare potenzielle Meinungsmacht einzelner Marktakteure unbedenklich. Ein Monitoring für aktuellere Jahre jedoch nötig, um Entwicklung der Publikumsfragmentierung online ebenso wie die Marktkonzentration und (Aus-)Differenzierung zu begleiten. Gerade vor dem Hintergrund, dass seit 2016 neben dem Kauf einiger reichweitenstarker Onlineangebote wie t-online durch Stroer-Gruppe 2014 weitere erhebliche Zusammenschlüsse im Werbemarkt zu beobachten sind: So bildete sich die „Ad Alliance“, unter deren Dach sich ansonsten konkurrierende Medienanbieter zusammengeschlossen haben, um ein Gegengewicht zur Stroer-Gruppe auf Vermarkter-Ebene zu bilden. Eine weitere Veränderung des Medienmarktes hat sich durch den Vorstoß des RTL-Konzerns in das Marktsegment der politischen Information ergeben. Das heißt zum einen, dass der Prozess der Publikumsfragmentierung online seit 2016 beeinflusst von den veränderten Kontextbedingungen bereits deutlich vorangeschritten sein könnte. Zum anderen könnten sich die Marktkonzentrationstendenzen (vgl. Kampes et al. 2021) bereits deutlicher in den Mediennutzungsmustern online niedergeschlagen haben und Meinungsmacht indizieren. Mit einem Monitoring kann dies sichtbar gemacht werden, gesellschaftliche Implikationen abgeleitet und medienpolitische Konsequenzen zur Sicherung von Vielfalt in der Online-Medienlandschaft diskutiert werden (vgl. Lobigs & Neuberger 2018).

Außerdem ist unklar, welche Auswirkungen die skizzierten Prozesse auf den Erhalt journalistischer Qualität haben und welche Rolle öffentlich-rechtliche Onlineangebote dabei spielen können. Dafür sind Projekte wie das D-A-CH-Projekt „Media Performance and Democracy“ (Projektnummer 323375807) nötig, um die Inhaltsebene (nicht) genutzter

Medienangebote online evaluieren zu können. Insbesondere Nutzergruppen mit einem geringen Repertoireumfang online sollten dabei im Blick behalten werden. Sind diese von eher geringer Qualität und zeigen insgesamt eine geringe Vielfalt auf, können diese Nutzende stark gefährdet sein, in Echokammern o.ä. abzugleiten.

Zudem bleibt offen, welche Rolle (neue) Intermediäre bei der Verteilung der Aufmerksamkeit und damit dem wirtschaftlichen Erfolg und Überleben bestimmter Onlinemedien spielen. So kann die Auswahl von Medienangeboten für Google News, Pocket und dergleichen eine große Rolle für die Nutzungszahlen und Mediennutzungsmuster spielen und eine wichtige Determinante von Publikumsfragmentierung online sein. Unklar ist zusätzlich, wie Medienangebote dafür ausgewählt werden und inwiefern Medienangebote ihre Aufnahme bei solchen potenziellen Intermediären, z.B. durch Teilnahme an Programmen wie der ‚Google News initiative‘, steuern können.

Des Weiteren ist, wie bereits diskutiert, eine Ausweitung auf Untersuchungen zu den anderen Dimensionen und Untersuchungsebenen von Fragmentierung wichtig. Allen voran sollten die Analysen neben der crossmedialen Nutzung und aktuellerer Erhebungszeitpunkte auf die Inhalte, die Wahrnehmung rezipierter Medien, das Vertrauen in und Medienbewertung von einzelnen Medienangeboten ausgeweitet werden. Denn nur so kann ein Gesamtbild des Fragmentierungsprozesses erreicht werden, eine Evaluation im Gesamtkontext der Mediennutzung, also im Online- und Offline-Bereich, vorgenommen und gesellschaftliche Implikationen wissenschaftsbasiert diskutiert werden.

Literatur

- Anderson, C. (2006). *The long tail: Why the future of business is selling less of more*. Hachette Books.
- Anthony Giddens. (1984). *The constitution of society: Outline of the theory of structuration*. University of California Press.
- Babić, D., & Jandura, O. (2017). Ein Blick in die Vergangenheit. Die Fragmentierung des Publikums von Zeitschriften. In O. Jandura, M. Wendelin, M. Adolf, & J. Wimmer (Eds.), *Zwischen Integration und Diversifikation: Medien und gesellschaftlicher Zusammenhalt im digitalen Zeitalter* (pp. 105–119). Springer Fachmedien Wiesbaden. https://doi.org/10.1007/978-3-658-15031-0_7
- Bächler, J. (2017). Medienrepertoires und Fragmentierung: Impulse zur Deutung der Medienkrise. In W. A. Meier (Ed.), *Medienstrukturen: Vol. 11. Abbruch - Umbruch - Aufbruch: Globaler Medienwandel und lokale Medienkrisen* (1st ed., pp. 319–342). Nomos Verlagsgesellschaft mbH & Co. KG. <https://doi.org/10.5771/9783845276663-319>
- Bähr, M., & Feurer, M. (2018). Influencer-Marketing. In C. Zerres & D. Drechsler (Eds.), *Hamburger Schriften zur Marketingforschung: Band 100. Social Media Marketing und Data Analytics: Aktuelle Entwicklungen und Herausforderungen* (1st ed., pp. 49–74). Rainer Hampp Verlag. <https://doi.org/10.5771/9783957103284-49>
- Barnidge, M., Diehl, T., Sherrill, L. A., & Zhang, J. (2021). Attention Centrality and Audience Fragmentation: An Approach for Bridging the Gap Between Selective Exposure and Audience Overlap. *Journal of Communication*, 71(6), 898–921. <https://doi.org/10.1093/joc/jqab023>
- Beaufort, M. (2018). Digital media, political polarization and challenges to democracy. *Information, Communication & Society*, 21(7), 915–920. <https://doi.org/10.1080/1369118X.2018.1451909>
- Beaufort, M. (2021). Die Demokratie? Die Öffentlichkeit? Die Rolle der Medien? In M. Magin, U. Rußmann, & B. Stark (Eds.), *Demokratie braucht Medien* (pp. 69–89). Springer VS. https://doi.org/10.1007/978-3-658-34633-1_5
- Beck, K. (2018). *Das Mediensystem Deutschlands: Strukturen, Märkte, Regulierung* (2nd ed.). Springer Fachmedien. <https://doi.org/10.1007/978-3-658-11779-5>
- Benoit, W. L., & Billings, A. C. (2020). *The rise and fall of mass communication. Mass communication and journalism: vol. 27*. Peter Lang. <https://www.peterlang.com/view/product/92866?format=HC>
- Blank, G. (2017). The Digital Divide Among Twitter Users and Its Implications for Social Research. *Social Science Computer Review*, 35(6), 679–697. <https://doi.org/10.1177/0894439316671698>

- Bonfadelli, H. (2002). The Internet and Knowledge Gaps. *European Journal of Communication*, 17(1), 65–84. <https://doi.org/10.1177/0267323102017001607>
- Bourdieu, P. (1984). *Distinction: A social critique of the judgement of taste*. Harvard University Press.
- Brentel, I., Kampes, C. F., & Jandura, O. (2022). *Media-Analyse Daten: IntermediaPlus Daten von 2014 bis 2016 (MA IntermediaPlus)*. GESIS, Cologne. ZA5769 Data file Version 1.0.0, <https://doi.org/10.4232/1.13530>
- Brentel, I., Jandura, O., Kampes, F. C., & Khabyuk, O. (2021). *Datenaufbereitung der MA-IntermediaPlus: Teil 1*. GESIS, Cologne. SoWiDataNet Datenfile Version 1.0.0, <https://doi.org/10.7802/2339>
- Brentel, I. (2021a). *Datenaufbereitung der MA-IntermediaPlus: Teil 2. Prozess-Schritt 2 „Import & Harmonisieren“ in und mit CharmStats*, GESIS, Cologne. SoWiDataNet Datenfile Version 1.0.0, <https://doi.org/10.7802/2340>
- Brentel, I. (2021b). *Datenaufbereitung der MA-IntermediaPlus: Teil 3. Prozess-Schritt 3 „Harmonisierung Anwenden“ mit charmstana*. GESIS, Cologne. SoWiDataNet Datenfile Version 1.0.0, <https://doi.org/10.7802/2341>
- Brentel, I., & Fortmann, N. (2021). *Datenaufbereitung der MA-IntermediaPlus: Teil 4. Prozess-Schritt 4 „Dokumentation erstellen“*. GESIS, Cologne. SoWiDataNet Datenfile Version 1.0.0, <https://doi.org/10.7802/2342>
- Brentel, I., & Winters, K. (2021). The Longitudinal IntermediaPlus (2014–2016): A Case Study in Structuring Unstructured Big Data: Other Humanities. *Research Data Journal for the Humanities and Social Sciences*, 1–16. <https://doi.org/10.1163/24523666-06010001>
- Brentel, I. (forthcoming). Online media consumption in Germany: The role of political information. An analysis of German mass communication online. *Journal of Communication*. [proceeding].
- Brentel, I. & Winters, K. (forthcoming). Processing and Documenting Big Data According to the FAIR Principles: A Communications Research Use Case. *Publizistik*, (Special Issue “Data, Archives and Tools”). [proceeding].
- Breyer-Mayländer, T. (2018). Social Media und Data Analytics im gesellschaftlichen Kontext. In C. Zerres & D. Drechsler (Eds.), *Hamburger Schriften zur Marketingforschung: Band 100. Social Media Marketing und Data Analytics: Aktuelle Entwicklungen und Herausforderungen* (1st ed., pp. 15–35). Rainer Hampp Verlag. <https://doi.org/10.5771/9783957103284-13>
- Brosius, H.-B. (2013). Neue Medienumgebungen. Theoretische und methodische Herausforderungen. In O. Jandura, A. Fahr, & H.-B. Brosius (Eds.), *Reihe Rezeptionsforschung: Vol. 25. Theorieanpassungen in der digitalen Medienwelt* (1st

- ed., pp. 13–30). Nomos; Nomos Verlagsgesellschaft mbH & Co. KG.
<https://doi.org/10.5771/9783845242682-13>
- Bruns, A. (2019a). *Are filter bubbles real?* Polity Press.
<https://ebookcentral.proquest.com/lib/ulbd/reader.action?docID=5887595>
- Bruns, A. (2019b). Filter bubble. *Internet Policy Review*, 8(4).
<https://doi.org/10.14763/2019.4.1426>
- Bruns, A., Burgess, J., Highfield, T., Kirchhoff, L., & Nicolai, T. (2011). Mapping the Australian Networked Public Sphere. *Social Science Computer Review*, 29(3), 277–287.
<https://doi.org/10.1177/0894439310382507>
- Bruns, A., & Enli, G. (2018). The Norwegian Twittersphere: Structure and Dynamics. *Nordicom Review*, 39(1), 129–148. <https://doi.org/10.2478/nor-2018-0006>
- Bruns, A., Moon, B., Münch, F., & Sadkowsky, T. (2017). The Australian Twittersphere in 2016: Mapping the Follower/Followee Network. *Social Media + Society*, 3(4), 205630511774816. <https://doi.org/10.1177/2056305117748162>
- Coleman, J. S. (1990). *Foundations of social theory*. Harvard University Press.
- Coleman, S. (2017). *Can the internet strengthen democracy?* Polity.
- Converse, P. E. (1964). The Nature of Belief Systems in Mass Publics. In D. E. Apter (Ed.), *Ideology and Discontent* (pp. 206–261). The Free Press of Glencoe.
- Costera Meijer, I., & Groot Kormelink, T. (2015). Checking, Sharing, Clicking and Linking: Changing patterns of news use between 2004 and 2014. *Digital Journalism*, 3(5), 664–679. <https://doi.org/10.1080/21670811.2014.937149>
- Cunningham, B. M. (2018). Evolving Research and Theories in Media Economics. In A. B. Albarran, B. I. Mierzejewska, & J. Jung (Eds.), *Handbook of Media Management and Economics* (2nd ed., pp. 36–51). Routledge.
- Dahlberg, L. (2007). Rethinking the fragmentation of the cyberpublic: from consensus to contestation. *New Media & Society*, 9(5), 827–847.
<https://doi.org/10.1177/1461444807081228>
- Dienlin, T., Johannes, N., Bowman, N. D., Masur, P. K., Engesser, S., Kümpel, A. S., Lukito, J., Bier, L. M., Zhang, R., Johnson, B. K., Huskey, R., Schneider, F. M., Breuer, J., Parry, D. A., Vermeulen, I., Fisher, J. T., Banks, J., Weber, R., Ellis, D. A., . . . Vreese, C. de (2020). An Agenda for Open Science in Communication. *Journal of Communication*, 17(5), 1. <https://doi.org/10.1093/joc/jqz052>
- Dohle, M., Jandura, O., & Vowe, G. (2014). Politische Kommunikation in der Online-Welt. Dimensionen des strukturellen Wandels politischer Kommunikation. *Zeitschrift Für Politik*, 61(4), 414–436. <https://doi.org/10.5771/0044-3360-2014-4-414>
- Downey, J., & Fenton, N. (2003). New Media, Counter Publicity and the Public Sphere. *New Media Society*(5), 185–202.

- Dubois, E., & Blank, G. (2018). The echo chamber is overstated: the moderating effect of political interest and diverse media. *Information, Communication & Society*, 21(5), 729–745. <https://doi.org/10.1080/1369118X.2018.1428656>
- Dvir-Gvirsman, S. (2017). Media audience homophily: Partisan websites, audience identity and polarization processes. *New Media & Society*, 19(7), 1072–1091. <https://doi.org/10.1177/1461444815625945>
- Egan, B., & Tsao, J. (2015). The Rise of native Advertising: an exploration of its impact on Advertising. Journalism and the Consumer. In American Academy of Advertising (Chair), *American Academy of Advertising. Conference. Proceedings*, Chicago.
- Enke, N., Bigl, B., Schubert, M., & Welker, M. (2021). *Studie zu Werbepraktiken und direkten Kaufappellen an Kinder in sozialen Medien: Ein Forschungsprojekt im Auftrag der KJM*. https://www.kjm-online.de/fileadmin/user_upload/KJM/Publikationen/Studien_Gutachten/210811_Abschlussbericht_KJM.pdf
- Esser, H. (1993). *Soziologie. Allgemeine Grundlagen*. Campus.
- Festinger, L. (1957). *A theory of cognitive dissonance* (1. ed). Stanford University Press.
- Fletcher, R., & Nielsen, R. K. (2017). Are News Audiences Increasingly Fragmented? A Cross-National Comparative Analysis of Cross-Platform News Audience Fragmentation and Duplication. *Journal of Communication*, 67(4), 476–498. <https://doi.org/10.1111/jcom.12315>
- Friedrichs, J. (1983). *Methoden empirischer Sozialforschung*. Westdeutscher Verlag.
- Friemel, T. N. (2013). *Sozialpsychologie der Mediennutzung: Motive, Charakteristiken und Wirkungen interpersonalen Kommunikation über massenmediale Inhalte. Kommunikationswissenschaft*. UVK-Verlagsges.
- Gehrau, V., & Goertz, L. (2010). Gespräche über Medien unter veränderten medialen Bedingungen. *Publizistik*, 55(2), 153–172. <https://doi.org/10.1007/s11616-010-0082-5>
- Georgiou, M. (2006). *Diaspora, identity, and the media: Diasporic transnationalism and mediated spatialities*. Hampton Press.
- Görland, S. O., Reiter, G., Gonser, N., Grammel, M., Gründl, J., Nikolaus, U., Werner, N., Zimmer, F., & Rademacher, U. (2017). Ansätze und Methoden zur Messung veränderter Mediennutzung. In G. Hooffacker & C. Wolf (Eds.), *Technische Innovationen - Medieninnovationen? Herausforderungen für Kommunikatoren, Konzepte und Nutzerforschung* (pp. 247–305). Springer Fachmedien. https://doi.org/10.1007/978-3-658-14953-6_5
- Greve, J., Schnabel, A., & Schützeichel, R. (2008). Das Makro-Mikro-Makro-Modell der soziologischen Erklärung - zur Einleitung. In J. Greve, A. Schnabel, & R. Schützeichel

- (Eds.), *Das Mikro-Makro-Modell der soziologischen Erklärung* (pp. 7–17). VS Verlag für Sozialwissenschaften. https://doi.org/10.1007/978-3-531-91774-0_1
- Grotz, F., & Schroeder, W. (2021). *Das politische System der Bundesrepublik Deutschland*. Springer Fachmedien Wiesbaden. <https://doi.org/10.1007/978-3-658-08638-1>
- Gudivada, V., Apon, A., & Ding, J. (2017). Data quality considerations for big data and machine learning: Going beyond data cleaning and transformations. *International Journal on Advances in Software*, 10(1), 1–20.
- Haas, A., & Brosius, H.-B. (2013). Fragmentierung der Publikumsagenda im Zeitverlauf? In K. Imhof, R. Blum, H. Bonfadelli, & O. Jarren (Eds.), *Mediensymposium. Stratifizierte und segmentierte Öffentlichkeit* (pp. 187–202). Springer VS, Wiesbaden; Springer VS. https://doi.org/10.1007/978-3-658-00348-7_11
- Habermas, J. (2021). Überlegungen und Hypothesen zu einem erneuten Strukturwandel der politischen Öffentlichkeit. In M. Seeliger & S. Seignani (Eds.), *Leviathan Sonderband: Vol. 37. Ein neuer Strukturwandel der Öffentlichkeit?* (1st ed., pp. 470–498). Nomos.
- Haim, M., Graefe, A., & Brosius, H.-B. (2018). Burst of the Filter Bubble? *Digital Journalism*, 6(3), 330–343. <https://doi.org/10.1080/21670811.2017.1338145>
- Hallin, D. C., & Mancini, P. (2004). *Comparing media systems: Three models of media and politics* (1. publ). *Communication, society, and politics*. Cambridge Univ. Press.
- Handel, U. (2000). *Die Fragmentierung des Medienpublikums: Bestandsaufnahme und Empirische Untersuchung Eines Phänomens der Mediennutzung und Seiner Determinanten*. VS Verlag für Sozialwissenschaften GmbH.
- Hasebrink, U., & Popp, J. (2006). Media repertoires as a result of selective media use. A conceptual approach to the analysis of patterns of exposure. *Communications*, 31(3), 422. <https://doi.org/10.1515/COMMUN.2006.023>
- Hepp, A. (2016). Kommunikations- und Medienwissenschaft in datengetriebenen Zeiten. *Publizistik*, 61(3), 225–246. <https://doi.org/10.1007/s11616-016-0263-y>
- Hofsäss, M., Engel, D., & Kellerwessel, L. (2020). Kommerzielle Publikumsmarktforschung. In *Handbuch Medienökonomie* (pp. 445–515). Springer VS, Wiesbaden. https://doi.org/10.1007/978-3-658-09560-4_27
- Hölig, S., Hasebrink, U., & Behre, J. (2021). *Reuters Institute Digital News Report 2021: Ergebnisse für Deutschland*. <https://doi.org/10.21241/SSOAR.73637>
- Holtz-Bacha, C. (1997). Das fragmentierte Medien-Publikum. Folgen für das politische System. *Aus Politik Und Zeitgeschichte*, 42(97), 13–21.
- Holtz-Bacha, C., & Peiser, W. (1999). Verlieren die Massenmedien ihre Integrationsfunktion? Eine empirische Analyse zu den Folgen der Fragmentierung des Medienpublikums. In U. Hasebrink & P. Rössler (Eds.), *Publikumsbindungen. Medienrezeption zwischen Individualisierung und Integration* (pp. 41–53). R. Fischer.

- Hopkin, J. (2002). Comparative methods. In D. Marsh & G. Stoker (Eds.), *Political analysis. Theory and methods in political science* (2nd ed., pp. 249–270). Palgrave Macmillan.
- Jandura, O., & Friedrich, K. (2014). 18. The quality of political media coverage. In C. Reinemann (Ed.), *Political communication* (pp. 351–374). De Gruyter Mouton; De Gruyter. <https://doi.org/10.1515/9783110238174>
- Jarren, O. (2000). Gesellschaftliche Integration durch Medien? Zur Begründung normativer Anforderungen an Medien. *M&K Medien & Kommunikationswissenschaft*, 48(1), 22–41. <https://doi.org/10.5771/1615-634x-2000-1-22>
- Jarren, O. (2014). Erfüllen die Medien heute einen demokratischen Auftrag? Wie sich der Medienwandel auf Gesellschaft und Demokratie auswirkt. *ZPol Zeitschrift für Politikwissenschaft*, 24(3), 317–327. <https://doi.org/10.5771/1430-6387-2014-3-317>
- Jungherr, A., Jürgens, P., & Schoen, H. (2018). 12 Twitter-Daten in der Wahlkampfforschung: Datensammlung, Aufarbeitung und Analysebeispiele. In A. Blätte, J. Behnke, K.-U. Schnapp, & C. Wagemann (Eds.), *Schriftenreihe der Sektion Methoden der Politikwissenschaft der Deutschen Vereinigung für Politische Wissenschaft. Computational Social Science: Die Analyse von Big Data* (1st ed., pp. 255–294). Nomos Verlagsgesellschaft.
- Jürgens, P., & Jungherr, A. (2016). A Tutorial for Using Twitter Data in the Social Sciences: Data Collection, Preparation, and Analysis. *SSRN Electronic Journal*. Advance online publication. <https://doi.org/10.2139/ssrn.2710146> (Original work published 05.01.2016)
- Jürgens, P., Jungherr, A., & Schoen, H. (2011). Small worlds with a difference. *WebSci '11: Proceedings of the 3rd International Web Science Conference*, Article 21, 1–5. <https://doi.org/10.1145/2527031.2527034>
- Jürgens, P., & Stark, B. (2017). The Power of Default on Reddit: A General Model to Measure the Influence of Information Intermediaries. *Policy & Internet*, 9(4), 395–419. <https://doi.org/10.1002/poi3.166>
- Kampes, C. F. (2020a). Tail or no Tail? Applicability of the Long Tail Theory to the German Online Media Market. *Central European Journal of Communication*, 13(3), 371–389.
- Kampes, C. F. (2020b). Welche Genres existieren für Online-Medienangebote? Eine Analyse der Themenstruktur aus Anbietersicht [What genres exist for online media offerings? An analysis of the thematic structure from the perspective of media providers]. In W. Deiters, S. Geisler, F. Hörner, & A. K. Knaup (Eds.), *Die Kommunikation und ihre Technologien. Interdisziplinäre Perspektiven auf Digitalisierung* (pp. 13–44).
- Kampes, C. F., & Brentel, I. (2020). The German online media market: Online-born information offerings and their audiences – A shift towards digital inequalities? *World of Media. Journal of Russian Media and Journalism Studies*, 1(4), 5–34. <https://doi.org/10.30547/worldofmedia.4.2020.1>

- Kampes, C. F., Jandura, O., & Khabyuk, O. (2021). *Angebotsfragmentierung online. Empirische Analysen struktureller Differenzierung von Medienangeboten und Medienanbietern im Online-Medienmarkt* [Doctoral dissertation], Düsseldorf. https://docserv.uni-duesseldorf.de/servlets/derivateservlet/derivate-62684/schriftliche%20dissertationsleistung%20kampes%20august%202021_angebot_sfragmentierung%20online_publicationsversion.pdf
- Karnowski, V. (2019). Die theoretische Modellierung der Nutzung mobiler Medien – vom Innovationscluster Mobiltelefon zum Metamedium Smartphone. In H. Schramm, J. Matthes, & C. Schemer (Eds.), *Research. Emotions meet cognitions: Zum Zusammenspiel von emotionalen und kognitiven Prozessen in der Medienrezeptions- und Medienwirkungsforschung : Werner Wirth gewidmet anlässlich seines 60. Geburtstages* (pp. 83–91). Springer VS. https://doi.org/10.1007/978-3-658-25963-1_7
- Katz, E., Blumler, J. G., & Gurevitch, M. (1973). Uses and Gratifications Research. *The Public Opinion Quarterly*, 37(4), 509–523. <http://www.jstor.org/stable/2747854>
- Kiefer, M. L., & Steininger, C. (2014). *Medienökonomie* (3rd).
- Klein, O., Hardwicke, T. E., Aust, F., Breuer, J., Danielsson, H., Mohr, A. H., IJerman, H., Nilsonne, G., Vanpaemel, W., & Frank, M. C. (2018). A Practical Guide for Transparency in Psychological Science. *Collabra: Psychology*, 4(20). <https://doi.org/10.1525/collabra.158>
- Kleinen-von Königslöw, K. (2016). Publikumsfragmentierung in der Online-Nachrichtenumgebung. In P. Henn & D. Frieß (Eds.), *Politische Online-Kommunikation: Voraussetzungen und Folgen des strukturellen Wandels der politischen Kommunikation* (pp. 253–278). ifpuk - Institute for Media and Communication Studies at FU Berlin. <https://doi.org/10.17174/dcr.v3.11>
- Knobloch-Westerwick, S. (2014). *Choice and preference in media use: Advances in selective exposure theory and research*. Routledge Taylor & Francis Group. <https://doi.org/Silvia>
- Kösters, R. (2020). *Medien als Mittler im Konflikt? Der Streit um die Migration im Spiegel der Berichterstattung* [Doctoral dissertation]. Universitäts- und Landesbibliothek der Heinrich-Heine-Universität Düsseldorf. <https://docserv.uni-duesseldorf.de/servlets/DerivateServlet/Derivate-55445/K%C3%B6sters%202020%20-%20Medien%20als%20Mittler%20im%20Konflikt.pdf>
- Kösters, R., Jandura, O., Weiß, R., & Schreiber, J. (2021). Diskursallianzen in der Migrationsdebatte? Politischer Parallelismus zwischen Medien und Parteien im Framing der Flucht- und Asylmigration im Jahr 2018 [Discursive Alliances in the Debate on Migration?]. *Politische Vierteljahresschrift*, 62(3), 461–487. <https://doi.org/10.1007/s11615-021-00324-z>

- Ksiazek, T. B. (2011). A Network Analytic Approach to Understanding Cross-Platform Audience Behavior. *Journal of Media Economics*, 24(4), 237–251. <https://doi.org/10.1080/08997764.2011.626985>
- Kümpel, A. S. (2020a). The Matthew Effect in social media news use: Assessing inequalities in news exposure and news engagement on social network sites (SNS). *Journalism*, 21(8), 1083–1098. <https://doi.org/10.1177/1464884920915374>
- Kümpel, A. S. (2020b). Nebenbei, mobil und ohne Ziel? Eine Mehrmethodenstudie zu Nachrichtennutzung und -verständnis von jungen Erwachsenen. *M&K Medien & Kommunikationswissenschaft*, 68(1-2), 11–31. <https://doi.org/10.5771/1615-634X-2020-1-2-11>
- Lee, J. K. (2007). The Effect of the Internet on Homogeneity of the Media Agenda: A Test of the Fragmentation Thesis. *Journalism & Mass Communication Quarterly*, 84(4), 745–760. <https://doi.org/10.1177/107769900708400406>
- Libor, C. (2021). Studie der KJM zu Influencer-Werbung mit Einfluss auf Kinder. *AfP*, 52(5), 396. <https://doi.org/10.9785/afp-2021-520508>
- Lietz, H. (2020). Drawing impossible boundaries: Field delineation of Social Network Science. *Scientometrics*, 125(3), 2841–2876. <https://doi.org/10.1007/s11192-020-03527-0>
- Lobigs, F., & Neuberger, C. (2018). *Meinungsmacht im Internet und die Digitalstrategien von Medienunternehmen: Neue Machtverhältnisse trotz expandierender Internet-Geschäfte der traditionellen Massenmedien-Konzerne: Gutachten für die Kommission zur Ermittlung der Konzentration im Medienbereich (KEK)*. Schriftenreihe der Landesmedienanstalten: Vol. 51. VISTAS Verlag.
- Luhmann, N. (2005). *Konstruktivistische Perspektiven* (3. Aufl. [Sondered.]). *Soziologische Aufklärung / Niklas Luhmann: Vol. 5*. VS Verl. für Sozialwiss. <https://doi.org/10.1007/978-3-663-11449-9>
- Machill, M., Beiler, M., & Zenker, M. (2007). Suchmaschinenforschung. Überblick und Systematisierung eines interdisziplinären Forschungsfeldes. In M. Machill & M. Beiler (Eds.), *Die Macht der Suchmaschinen - The power of search engines* (pp. 7–43). von Halem.
- Magin, M., Geiß, S., Stark, B., & Jürgens, P. (2022). Common Core in Danger? Personalized Information and the Fragmentation of the Public Agenda. *The International Journal of Press/Politics*, 27(4), 887–909. <https://doi.org/10.1177/19401612211026595>
- Mahrt, M. (2019). *Beyond Filter Bubbles and Echo Chambers: The Integrative Potential of the Internet*. ifpuk - Institute for Media and Communication Studies at FU Berlin. <https://doi.org/10.17174/dcr.v5.0>
- Mahrt, M. (2020). Why we find little evidence of digital fragmentation, but should not stop researching it. In V. Gehrau, A. Waldherr, & A. Scholl (Eds.), *Jahrbuch der Publizistik-*

- und Kommunikationswissenschaft. *Integration durch Kommunikation (in einer digitalen Gesellschaft): Jahrbuch der Deutschen Gesellschaft für Publizistik- und Kommunikationswissenschaft 2019*. SSOAR - GESIS Leibniz Institute for the Social Sciences. <https://doi.org/10.21241/SSOAR.66412>
- Maletzke, G. (1980). Integration — eine gesellschaftliche Funktion der Massenkommunikation. *Publizistik*, 25(2-3), 199–206.
- Mangold, F., Vogelgesang, J., & Scharnow, M. (2017). Nachrichtennutzung in Deutschland. Eine nutzerzentrierte Repertoireanalyse. *M&K Medien & Kommunikationswissenschaft*, 65(4), 704–723. <https://doi.org/10.5771/1615-634X-2017-4-704>
- McNair, B. (2003). *An introduction to political communication* (3rd ed.). *Communication and society*. Routledge.
- McNair, B. (2011). *An introduction to political communication* (5th ed.). *Communication and society*. Routledge.
- McQuail, D. (1997). *Audience analysis*. SAGE Publications.
- McQuail, D. (2010). *Mcquail's mass communication theory* (6. ed.). SAGE.
- McQuail, D., & Deuze, M. (2020). *Mcquail's media & mass communication theory* (Seventh edition). SAGE. <https://doi.org/Denis>
- Medjedović, I. (2014). *Qualitative Sekundäranalyse: Zum Potenzial einer neuen Forschungsstrategie in der empirischen Sozialforschung*. Zugl.: Bremen, Univ., Diss., 2013. *Research*. Springer VS. <https://doi.org/10.1007/978-3-658-05488-5>
- Meier, W. A. (Ed.). (2017). *Medienstrukturen: Vol. 11. Abbruch - Umbruch - Aufbruch: Globaler Medienwandel und lokale Medienkrisen* (1. Auflage). Nomos Verlagsgesellschaft mbH & Co. KG. <https://doi.org/10.5771/9783845276663>
- Meier, W. A., & Jarren, O. (2001). Ökonomisierung und Kommerzialisierung von Medien und Mediensystem. Einleitende Bemerkungen zu einer (notwendigen) Debatte. *M&K Medien & Kommunikationswissenschaft*, 49(2), 145–158. <https://doi.org/10.5771/1615-634x-2001-2-145>
- Merten, L. (2020). Contextualized Repertoire Maps: Exploring the Role of Social Media in News-related Media Repertoires. Advance online publication. <https://doi.org/10.17169/FQS-21.2.3235> (Forum Qualitative Sozialforschung / Forum: Qualitative Social Research, Vol 21, No 2 (2020): Challenging Times— Qualitative Methods and Methodological Approaches to Research on Time).
- Merten, L. (2021). Block, Hide or Follow—Personal News Curation Practices on Social Media. *Digital Journalism*, 9(8), 1018–1039. <https://doi.org/10.1080/21670811.2020.1829978>
- Mons, B., Neylon, C., Velterop, J., Dumontier, M., da Silva Santos, L. O. B., & Wilkinson, M. D. (2017). Cloudy, increasingly FAIR; revisiting the FAIR Data guiding

- principles for the European Open Science Cloud. *Information Services & Use*, 37(1), 49–56. <https://doi.org/10.3233/ISU-170824>
- Mukerjee, S. (2021a). Rethinking Audience Fragmentation Using a Theory of News Reading Publics: Online India as a Case Study. *Pre-Print*. Advance online publication. <https://doi.org/10.31219/osf.io/3f47z>
- Mukerjee, S. (2021b). A systematic comparison of community detection algorithms for measuring selective exposure in co-exposure networks. *Scientific Reports*, 11(1), 15218. <https://doi.org/10.1038/s41598-021-94724-1>
- Mukerjee, S., Majó-Vázquez, S., & González-Bailón, S. (2018a). Networks of Audience Overlap in the Consumption of Digital News. *Journal of Communication*, 68(1), 26–50. <https://doi.org/10.1093/joc/jqx007>
- Mukerjee, S., Majó-Vázquez, S., & González-Bailón, S. (2018b). Response to Webster and Taneja's Response to "Networks of Audience Overlap in the Consumption of Digital News". *Journal of Communication*, 68(3), E15-E18. <https://doi.org/10.1093/joc/jqy022>
- Nechushtai, E., & Lewis, S. C. (2019). What kind of news gatekeepers do we want machines to be? Filter bubbles, fragmentation, and the normative dimensions of algorithmic recommendations. *Computers in Human Behavior*, 90(8), 298–307. <https://doi.org/10.1016/j.chb.2018.07.043>
- Nelson, J. L., & Webster, J. G. (2017). The Myth of Partisan Selective Exposure: A Portrait of the Online Political News Audience. *Social Media + Society*, 3(3). <https://doi.org/10.1177/2056305117729314>
- Neuberger, C. (2009). Internet, Journalismus und Öffentlichkeit. In C. Neuberger, C. Nuernbergk, & M. Rischke (Eds.), *Journalismus im Internet: Profession, Partizipation, Technisierung* (1st ed., pp. 19–105). VS Verl. für Sozialwiss. https://doi.org/10.1007/978-3-531-91562-3_2
- Neuberger, C. (2017). Die Rückkehr der Masse. Kollektivphänomene im Internet aus Sicht der Massen- und Komplexitätstheorie. *M&K Medien & Kommunikationswissenschaft*, 65(3), 550–572. <https://doi.org/10.5771/1615-634X-2017-3-550>
- Neuberger, C., Lobigs, F., Herbers, M. R., Karthaus, A., & Nuernbergk, C. (2010). *Die Bedeutung des Internets im Rahmen der Vielfaltssicherung: Gutachten im Auftrag der Kommission zur Ermittlung der Konzentration im Medienbereich (KEK)*. Schriftenreihe der Landesmedienanstalten: Vol. 43. Vistas.
- Ng, Y. M. M., & Taneja, H. (2019). Mapping User-Centric Internet Geographies: How Similar are Countries in Their Web Use Patterns? *Journal of Communication*, 69(5), 467–489. <https://doi.org/10.1093/joc/jqz030>

- Norris, P. (2001). *Digital divide: Civic engagement, information poverty, and the Internet worldwide. Communication, society and politics*. Cambridge University Press. <https://doi.org/10.1017/CBO9781139164887>
- Olsen, R. K. (2020). Connecting People? Understanding Media's Role as Democratic Resources for People in Digitally Advanced Local Environments. *Digital Journalism*, 8(4), 506–525. <https://doi.org/10.1080/21670811.2019.1679029>
- Ørmen, J. (2019a). Testing the Myth of Enclaves: A discussion of research designs for assessing algorithmic curation. In S. A. Eldridge & B. Franklin (Eds.), *Taylor & Francis eBooks. The Routledge handbook of developments in digital journalism studies* (pp. 132–142). Routledge Taylor & Francis Group.
- Ørmen, J. (2019b). Not So Distinct After All: Assessing Social Stratification of News Users on the Web. *Journalism Studies*, 20(11), 1653–1670. <https://doi.org/10.1080/1461670X.2018.1539342>
- Palmgreen, P. (1984). Uses and Gratifications: A Theoretical Perspective. *Annals of the International Communication Association*, 8(1), 20–55. <https://doi.org/10.1080/23808985.1984.11678570>
- Pariser, E. (2011). *The filter bubble: How the new personalized web is changing what we read and how we think*. Penguin.
- Peiser, W., & Jandura, O. Political Media Use. In *Donsbach (Hg.) 2015 – The international encyclopedia of communication* (pp. 1–6).
- Peng, Y., & Yang, T. (2022). Anatomy of audience duplication networks: How individual characteristics differentially contribute to fragmentation in news consumption and trust. *New Media & Society*, 24(10), 2270–2290. <https://doi.org/10.1177/1461444821991559>
- Peter, C., Breuer, J., Masur, P. K., Scharnow, M., & Schwarzenegger, C. (2020, December 11). *Empfehlungen zum Umgang mit Forschungsdaten in der Kommunikationswissenschaft*. https://www.dgpuk.de/sites/default/files/AG_Forschungsdaten%20Empfehlungen%20DGPuK_0.pdf
- Puschmann, C. (2019). Beyond the Bubble: Assessing the Diversity of Political Search Results. *Digital Journalism*, 7(6), 824–843. <https://doi.org/10.1080/21670811.2018.1539626>
- Ragnedda, M., Ruiu, M. L., & Addeo, F. (2020). Measuring Digital Capital: An empirical investigation. *New Media & Society*, 22(5), 793–816. <https://doi.org/10.1177/1461444819869604>
- Range, S., & Schweins, R. (2007). *Klicks, Quoten, Reizwörter: Nachrichten-Sites im Internet : wie das Web den Journalismus verändert ; Gutachten im Auftrag der Friedrich-Ebert-*

- Stiftung. *Medien- und Technologiepolitik*. Stabsabt. der Friedrich-Ebert-Stiftung. <https://library.fes.de/pdf-files/stabsabteilung/04417.pdf>
- Rau, J. P., & Stier, S. (2019). Die Echokammer-Hypothese: Fragmentierung der Öffentlichkeit und politische Polarisierung durch digitale Medien? *Zeitschrift Für Vergleichende Politikwissenschaft*, 13(3), 399–417. <https://doi.org/10.1007/s12286-019-00429-1>
- Scharkow, M., & Bachl, M. (2017). How Measurement Error in Content Analysis and Self-Reported Media Use Leads to Minimal Media Effect Findings in Linkage Analyses: A Simulation Study. *Political Communication*, 34(3), 323–343. <https://doi.org/10.1080/10584609.2016.1235640>
- Schröder, M. (2021). Medienkompetenz als Schlüssel für Demokratiekompetenz. In M. S. Hubacher & M. Waldis (Eds.), *Research. Politische Bildung für die digitale Öffentlichkeit: Umgang mit politischer Information und Kommunikation in digitalen Räumen* (pp. 89–107). Springer VS. https://doi.org/10.1007/978-3-658-33255-6_5
- Schulz, W. Fernsehen und sozialer Wandel: Untersuchungen zur Integrations- und Fragmentierungsthese. In J. Wilke (Ed.), *Schriftenreihe der DGPK: Vol. 26. Massenmedien und Zeitgeschichte* (pp. 90–105). UVK Medien.
- Schweiger, W. (2007). *Theorien der Mediennutzung: Eine Einführung*. Springer eBook Collection Humanities, Social Science. VS Verlag für Sozialwissenschaften. <https://doi.org/10.1007/978-3-531-90408-5>
- Schweiger, W. (2017). *Der (des)informierte Bürger im Netz: Wie soziale Medien die Meinungsbildung verändern*. Springer. <https://doi.org/10.1007/978-3-658-16058-6>
- Sjøvaag, H., & Kvalheim, N. (2019). Eventless news: Blindspots in journalism and the “long tail” of news content. *Journal of Applied Journalism & Media Studies*, 8(3), 291–310. https://doi.org/10.1386/ajms_00003_1
- Stark, B. (2013). Fragmentierung Revisited: eine theoretische und methodische Evaluation im Internetzeitalter. In F. Sattelberger & W. Seufert (Eds.), *Langfristiger Wandel von Medienstrukturen: Theorie, Methoden, Befunde* (1st ed., pp. 199–220). Nomos Verlagsgesellschaft mbH & Co. KG. <https://doi.org/10.5771/9783845249278-199>
- Stark, B. (2014). Informationsverhalten im 21. Jahrhundert – eine repertoireorientierte Analyse veränderter Nutzungsmuster. In K. Kleinen-von Königslöw (Ed.), *Reihe Rezeptionsforschung: Vol. 31. Medienkonvergenz und Medienkomplementarität aus Rezeptions- und Wirkungsperspektive* (1st ed., pp. 37–58). Nomos.
- Stark, B. (2019). Mythos „Filterblase“ – Fiktion oder Realität? Der Stand der Forschung aus kommunikationswissenschaftlicher Perspektive. *MedienWirtschaft*, 16(3), 6–10. <https://doi.org/10.15358/1613-0669-2019-3-6>

- Stark, B., & Kist, E. L. (2020). Mediennutzung. In J. Krone & T. Pellegrini (Eds.), *Handbuch Medienökonomie* (pp. 1137–1163). Springer Fachmedien Wiesbaden. https://doi.org/10.1007/978-3-658-09560-4_57
- Stark, B., Magin, M., & Jürgens, P. (2021a). Maßlos überschätzt. Ein Überblick über theoretische Annahmen und empirische Befunde zu Filterblasen und Echokammern. In M. Eisenegger, M. Prinzing, P. Ettinger, & R. Blum (Eds.), *Research. Digitaler Strukturwandel der Öffentlichkeit: Historische Verortung, Modelle und Konsequenzen* (pp. 303–321). Springer VS. https://doi.org/10.1007/978-3-658-32133-8_17
- Stark, B., Rußmann, U., & Magin, M. (2021b). Warum die Demokratie die Medien braucht. In M. Magin, U. Rußmann, & B. Stark (Eds.), *Demokratie braucht Medien* (pp. 3–16). Springer VS. https://doi.org/10.1007/978-3-658-34633-1_1
- Steiner, M., Magin, M., Stark, B., & Jürgens, P. (2019). Aus Versehen informiert? In P. Weber, F. Mangold, M. Hofer, & T. Koch (Eds.), *Reihe Rezeptionsforschung: Band 40. Meinungsbildung in der Netzöffentlichkeit: Aktuelle Studien zu Nachrichtennutzung, Meinungs Austausch und Meinungsbeeinflussung in Social Media* (1st ed., pp. 17–36). Nomos. <https://doi.org/10.5771/9783845293837-17>
- Steppat, D., Castro Herrero, L., & Esser, F. (2022). Selective exposure in different political information environments - How media fragmentation and polarization shape congruent news use. *European Journal of Communication*, 37(1), 82–102. <https://doi.org/10.1177/02673231211012141>
- Storey, J. (1999). *Cultural consumption and everyday life* (1. ed). *Cultural studies in practice*. Arnold.
- Strippel, C. (2021). Forschungsinfrastrukturen für die Kommunikations- und Medienforschung im deutschsprachigen Raum. Initiativen, Bedarfe und Perspektiven. *M&K Medien & Kommunikationswissenschaft*, 69(1), 136–157. <https://doi.org/10.5771/1615-634X-2021-1-136>
- Strippel, C., Bock, A., Katzenbach, C., Mahrt, M., Merten, L., Nuernbergk, C., Pentzold, C., Puschmann, C., & Waldherr, A. (2018). Die Zukunft der Kommunikationswissenschaft ist schon da, sie ist nur ungleich verteilt. *Publizistik*, 63(1), 11–27. <https://doi.org/10.1007/s11616-017-0398-5>
- Stroud, N. J. (2010). Polarization and partisan selective exposure. *Journal of Communication*, 60(3), 556–576. <https://doi.org/10.1111/j.1460-2466.2010.01497.x>
- Struckmann, S., & Karnowski, V. (2016). News consumption in a changing media ecology: An MESM-study on mobile news. *Telematics and Informatics*, 33(2), 309–319. <https://doi.org/10.1016/j.tele.2015.08.012>
- Sunstein, C. R. (2001). *Republic.com*. Princeton University Press.
- Sunstein, C. R. (2007). *Republic.com 2.0*.

- Sunstein, C. R. (2017). *#Republic*. Princeton University Press.
<https://doi.org/10.1515/9781400884711>
- Taneja, H. (2020). The Myth of Targeting Small, But Loyal Niche Audiences: Double-Jeopardy Effects In Digital-Media Consumption. *Journal of Advertising Research*, 60(3), 239–250. <https://doi.org/10.2501/JAR-2019-037>
- Taneja, H., Webster, J. G., Malthouse, E. C., & Ksiazek, T. B. (2012). Media consumption across platforms: Identifying user-defined repertoires. *New Media & Society*, 14(6), 951–968. <https://doi.org/10.1177/1461444811436146>
- Taneja, H., & Wu, A. X. (2018). Pathways to Fragmentation. In H. Akkermans, K. Fontaine, I. Vermeulen, G.-J. Houben, & M. S. Weber (Eds.), *Proceedings of the 10th ACM Conference on Web Science* (pp. 255–259). ACM.
<https://doi.org/10.1145/3201064.3201107>
- Tewksbury, D. (2005). The Seeds of Audience Fragmentation: Specialization in the Use of Online News Sites. *Journal of Broadcasting & Electronic Media*, 49(3), 332–348.
https://doi.org/10.1207/s15506878jobem4903_5
- Tewksbury, D., & Riles, J. M. (2015). Polarization as a Function of Citizen Predispositions and Exposure to News on the Internet. *Journal of Broadcasting & Electronic Media*, 59(3), 381–398. <https://doi.org/10.1080/08838151.2015.1054996>
- Tewksbury, D., & Rittenberg, J. (2009). Online news creation and consumption: Implications for modern democracies. In A. Chadwick & P. N. Howard (Eds.), *Routledge international handbooks. Routledge handbook of internet politics* (1st ed., pp. 186–200). Routledge.
- Thimm, C., & Anastasiadis, M. (2017). Kernkonzepte der Kommunikationswissenschaft im Wandel – Digitale Öffentlichkeit zwischen Fragmentierung, Polymedia und „Mini-Publics“. In M. Beiler & B. Bigl (Eds.), *Schriftenreihe der Deutschen Gesellschaft für Publizistik- und Kommunikationswissenschaft: Band 44. 100 Jahre Kommunikationswissenschaft in Deutschland: Von einem Spezialfach zur Integrationsdisziplin* (pp. 231–246). Herbert von Halem Verlag.
- Tichenor, P. J., Donohue, G. A., & Olien, C. N. (1970). Mass Media Flow and Differential Growth in Knowledge. *The Public Opinion Quarterly*, 34(2), 159–170.
<http://www.jstor.org/stable/2747414>
- Trilling, D., & Schoenbach, K. (2013). Patterns of News Consumption in Austria: How Fragmented Are They? *International Journal of Communication*, 7(0), 25.
<https://ijoc.org/index.php/ijoc/article/view/1769>
- Turow, J. (1998). *Breaking Up America: Advertisers and the New Media World* (New edition). University of Chicago Press.

- Turow, J. (2011). *The daily you: How the new advertising industry is defining your identity and your worth*. Yale University Press.
<https://yalebooks.yale.edu/book/9780300188011/daily-you>
- van Aelst, P., Strömbäck, J., Aalberg, T., Esser, F., Vreese, C. de, Matthes, J., Hopmann, D., Salgado, S., Hubé, N., Stępińska, A., Papathanassopoulos, S., Berganza, R., Legnante, G., Reinemann, C., Sheaffer, T., & Stanyer, J. (2017). Political communication in a high-choice media environment: a challenge for democracy? *Annals of the International Communication Association*, 41(1), 3–27.
<https://doi.org/10.1080/23808985.2017.1288551>
- van Atteveldt, W., Margolin, D., Shen, C., Trilling, D., & Weber, R. (2019a). A Roadmap for Computational Communication Research. *Computational Communication Research*, 1(1), 1–11. <https://doi.org/10.5117/CCR2019.1.001.VANA>
- van Atteveldt, W., Strycharz, J., Trilling, D., & Welbers, K. (2019b). Toward Open Computational Communication Science: A Practical Road Map for Reusable Data and Code. *International Journal of Communication*, 13, 3935–3954.
- van Atteveldt, W., & Peng, T.-Q. (2018). When Communication Meets Computation: Opportunities, Challenges, and Pitfalls in Computational Communication Science. *Communication Methods and Measures*, 12(2-3), 81–92.
<https://doi.org/10.1080/19312458.2018.1458084>
- van Cuilenburg, J. (1998). Diversity revisited: Towards a critical rational model of media diversity. In K. Brants, J. Hermes, & L. van Zoonen (Eds.), *The media in question: Popular cultures and public interests* (pp. 38–49). SAGE Publications.
- van Deursen, A. J., & van Dijk, J. A. (2014). The digital divide shifts to differences in usage. *New Media & Society*, 16(3), 507–526. <https://doi.org/10.1177/1461444813487959>
- van Dijck, J. (2012). Facebook as a Tool for Producing Sociality and Connectivity. *Television & New Media*, 13(2), 160–176. <https://doi.org/10.1177/1527476411415291>
- Vlašić, A. (2004). *Die Integrationsfunktion der Massenmedien.: Begriffsgeschichte, Modelle, Operationalisierung*. VS Verlag für Sozialwissenschaften.
- Waldherr, A., Geise, S., Mahrt, M., Katzenbach, C., Nuernbergk, C., Waldherr, A., Geise, S., & Katzenbach, C. (2021). Toward a Stronger Theoretical Grounding of Computational Communication Science. *Computational Communication Research*, 3(2), 1–28.
<https://doi.org/10.5117/CCr2021.02.002.WALd>
- Warin, T., & Sanger, W. (2014). Structuring big data: How financial models may help. *Journal of Computer Science and Technology*, 2(1), 1.
- Weber, P., Mangold, F., Hofer, M., & Koch, T. (Eds.). (2019). *Reihe Rezeptionsforschung: Band 40. Meinungsbildung in der Netzöffentlichkeit: Aktuelle Studien zu*

- Nachrichtennutzung, Meinungsaustausch und Meinungsbeeinflussung in Social Media* (1. Auflage). Nomos.
- Webster, J. G. (2005). Beneath the veneer of fragmentation: Television audience polarization in a multichannel world. *Journal of Communication*, 55(2), 366–382.
- Webster, J. G. (2014). *The Marketplace of Attention: How Audiences Take Shape in a Digital Age*. MIT Press.
- Webster, J. G. (2018). 6. Audience Behavior. In P. M. Napoli (Ed.), *Mediated Communication* (pp. 91–104). De Gruyter. <https://doi.org/10.1515/9783110481129-007>
- Webster, J. G., & Ksiazek, T. B. (2012). The dynamics of audience fragmentation: Public attention in an age of digital media. *Journal of Communication*, 62(1), 39–56. <https://doi.org/10.1111/j.1460-2466.2011.01616.x>
- Webster, J. G., & Taneja, H. (2018). Building and Interpreting Audience Networks: A Response to Mukerjee, Majo-Vazquez & Gonzalez-Bailon. *Journal of Communication*, 68(3), E11-E14. <https://doi.org/10.1093/joc/jqy024>
- Webster, J. G., & Wakshlag, J. J. (1983). A theory of television program choice. *Communication Research*, 10(4), 430–446.
- Wei, L., & Hindman, D. B. (2011). Does the Digital Divide Matter More? Comparing the Effects of New Media and Old Media Use on the Education-Based Knowledge Gap. *Mass Communication and Society*, 14(2), 216–235. <https://doi.org/10.1080/15205431003642707>
- Weiß, R. (2013). Segmentierung politischer Kommunikation in Milieus. In K. Imhof, R. Blum, H. Bonfadelli, & O. Jarren (Eds.), *Mediensymposium. Stratifizierte und segmentierte Öffentlichkeit* (pp. 205–217). Springer VS, Wiesbaden; Springer VS. https://doi.org/10.1007/978-3-658-00348-7_12
- Weiß, R., & Jandura, O. (2017). Medien und gesellschaftlicher Zusammenhalt. In O. Jandura, M. Wendelin, M. Adolf, & J. Wimmer (Eds.), *Zwischen Integration und Diversifikation: Medien und gesellschaftlicher Zusammenhalt im digitalen Zeitalter* (pp. 11–31). Springer Fachmedien Wiesbaden. https://doi.org/10.1007/978-3-658-15031-0_2
- Weiß, R., Kösters, R., & Mahrt, M. (2020). Democratic Theory and the Potential of Value Frames in Assessing Media Performance. *Media and Communication*, 8(3), 281–292. <https://doi.org/10.17645/mac.v8i3.3058>
- Wendelin, M. (2011). *Medialisierung der Öffentlichkeit: Kontinuität und Wandel einer normativen Kategorie der Moderne*. Teilw. zugl.: München, Univ., Diss., 2010. *Theorie und Geschichte der Kommunikationswissenschaft: Vol. 10*. von Halem.
- Wessler, H., & Rinke, E. M. (2016). Journalismus und Politik. In M. Löffelholz & L. Rothenberger (Eds.), *Handbuch Journalismustheorien* (pp. 639–653). Springer Fachmedien Wiesbaden. https://doi.org/10.1007/978-3-531-18966-6_39

- Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., . . . Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, 3, 160018. <https://doi.org/10.1038/sdata.2016.18>
- Williams, B. A., & Delli Carpini, M. X. (2011). *After broadcast news: Media regimes, democracy, and the new information environment. Communication, society and politics*. Cambridge University Press.
- Wirtz, B. W. (2018). *Electronic Business* (6., aktualisierte und erweiterte Auflage). Springer Gabler.
- Wöste, M. (1999). Öffentlich-rechtliches Fernsehen: für Jugendliche nicht jung genug? Nutzungsverhalten und Präferenzen junger Leute beim Fernsehen. *Media Perspektiven*(11), 583–590.
- Wu, A. X., & Taneja, H. (2016). Reimagining Internet Geographies: A User-Centric Ethnological Mapping of the World Wide Web. *Journal of Computer-Mediated Communication*, 21(3), 230–246. <https://doi.org/10.1111/jcc4.12157>
- Wu, A. X., & Taneja, H. (2021). Platform enclosure of human behavior and its measurement: Using behavioral trace data against platform episteme. *New Media & Society*, 23(9), 2650–2667. <https://doi.org/10.1177/1461444820933547>
- Wu, A. X., Taneja, H., & Webster, J. G. (2021). Going with the flow: Nudging attention online. *New Media & Society*, 23(10), 2979–2998. <https://doi.org/10.1177/1461444820941183>
- Yuan, E. (2011). News consumption across multiple media platforms. A repertoire approach. *Information, Communication & Society*, 14(7), 998–1016. <https://doi.org/10.1080/1369118X.2010.549235>
- Zillien, N., & Hargittai, E. (2009). Digital Distinction: Status-Specific Types of Internet Usage. *Social Science Quarterly*, 90(2), 274–291. <https://doi.org/10.1111/j.1540-6237.2009.00617.x>
- Zillmann, D. (1985). The experimental exploration of gratifications from media entertainment. In K. E. Rosengren, L. A. Wenner, & P. Palmgreen (Eds.), *Media gratifications research: Current perspectives* (1st ed., pp. 225–239). SAGE.

Anhang I

Inhalt Anhang I

Tabelle 1: Die fünf meist genutzten Onlinemedienangebote des gelben Teilpublikums **Fehler!**

Textmarke nicht definiert.

Tabelle 2: Die fünf meist genutzten Onlinemedienangebote des blauen Teilpublikums **Fehler!**

Textmarke nicht definiert.

Bei den beiden nachfolgenden Tabellen handelt es sich um ergänzendes Analysematerial der sozialen Netzwerkanalyse, das im Kontext der Analysen zur vierten Einzelschrift des Kumulus (Brentel *forthcoming* [Text4]) entstanden ist. Sie zeigen jeweils die fünf meist genutzten und damit – im Sinne der Netzwerkanalyse – zentralsten Onlinemedienangebote der beiden identifizierten Teilpublika (blau vs. gelb). Anders als bei den Nutzenden können die Medienangebote ausschließlich Teil eines Teilpublikums sein.

Die Information, die in den Tabellen abgetragen ist, können Hinweise auf die Nutzungsintensität und Rolle einzelner Medienangebote in einem Teilpublikum geben. Darüber hinaus ist sie Anhaltspunkt für die Entstehung und Verteilung möglicher Meinungsmacht von Anbietern (und Vermarktern), welche gerade für die (medienpolitische) Praxis von großem Interesse ist (vgl. die KEK-Studie von Lobigs & Neuberger 2018).

Tabelle 1: Die fünf am meisten genutzten Onlinemedienangebote des gelben Teilpublikums

| Onlineangebote des gelben Teilpublikums | Degree (aggr. Anzahl der Nutzernde) | Anzahl der Nutzungen | Genre | Gattungsherkunft | Medienanbieter | Vermarkter |
|---|-------------------------------------|----------------------|-----------|-----------------------------|--------------------------------|----------------|
| FOCUS ONLINE - Politik #Politik | 68.596 | 1.424.900 | Politik | Offline Print (Zeitschrift) | FOCUS Online Group GmbH | BurdaForward |
| t-online.de - E-Mail Center #Digitales | 45.830 | 754.520 | Digitales | Online-born | Ströer Digital Publishing GmbH | Ströer Digital |
| SPIEGEL ONLINE - Politik #Politik | 42.452 | 900.534 | Politik | Offline Print (Zeitschrift) | SPIEGEL ONLINE GmbH & Co. KG | SPIEGEL QC |
| gutefrage.net - Telekommunikation #Digitales | 41.338 | 71.858 | Digitales | Online-born | gutefrage.net GmbH | Ströer Digital |
| gutefrage.net - Computer #Digitales | 34.348 | 52.304 | Digitales | Online-born | gutefrage.net GmbH | Ströer Digital |

Notiz: Die Tabelle umfasst die fünf am meisten genutzten Onlinemedienangebot sowohl in der Anzahl der Befragten, welche im Erhebungszeitraum 2016 das jeweilige Angebot nutzen und damit durch ein Medienangebot erreicht wurden, als auch in der Anzahl der Zugriffe der Nutzenden, sprich der Nutzungshäufigkeit der einzelnen Nutzenden. Dadurch kann rückgeschlossen werden, dass die Nutzung der in der Tabelle abgebildeten Angebote für die Formation des gelben Teilpublikums eine zentrale Rolle spielen.

Tabelle 2: Die fünf am meisten genutzten Onlinemedienangebote des blauen Teilpublikums

| Onlineangebote des gelben Teilpublikums | Degree (aggr. Anzahl Nutzende) | Anzahl der Nutzungen | Genre | Gattungsherkunft | Medianbieter | Vermarkter |
|--|--------------------------------|----------------------|------------|-------------------------|----------------------------|--------------|
| WELT - Politik #Politik | 31.549 | 244.261 | Politik | Offline Print (Zeitung) | Axel Springer SE | Media Impact |
| CHEFKOCH - Rezepte #Gesundheit | 20.119 | 74.339 | Gesundheit | Online-born | Gruener + Jahr Medien GmbH | G+J eIMS |
| Sueddeutsche Zeitung - Politik #Politik | 18.488 | 126.033 | Politik | Offline Print (Zeitung) | Süddeutsche Zeitung GmbH | iq digital |
| ntv - Politik #Politik | 17.379 | 181.139 | Politik | Offline TV | Nachrichtenfernsehen GmbH | G+J eIMS |
| BILD - Homepage #Homepage | 15.483 | 77.456 | Politik | Offline Print (Zeitung) | Axel Springer SE | Media Impact |

Notiz: Die Tabelle umfasst die fünf am meisten genutzten Onlinemedienangebot sowohl in der Anzahl der Befragten, welche im Erhebungszeitraum 2016 das jeweilige Angebot nutzen und damit durch ein Medienangebot erreicht wurden, als auch in der Anzahl der Zugriffe der Nutzenden, sprich der Nutzungshäufigkeit der einzelnen Nutzenden. Dadurch kann rückgeschlossen werden, dass die Nutzung der in der Tabelle abgebildeten Angebote für die Formation des blauen Teilpublikums eine zentrale Rolle spielen.

Anhang II

Nachfolgend sind die zwei Einzelschriften der kumulativen Dissertation in voller Länge aufgeführt, welche momentan im Review-Verfahren oder Veröffentlichungsprozess sind. Der Anhang ist in folgender Reihenfolge angeordnet:

Text 2 in voller Länge

Text 4 in voller Länge

TEXT 2

Processing and Documenting Big Data According to the FAIR Principles

A Communications Research Use Case

Status of the Article:

The article is currently under review for the special issue “Data, Archives and Tools”
Publizistik.

Authors and Affiliation:

First author: Inga Brentel is a PhD student at the Department of Social Sciences at the
Heinrich Heine University Düsseldorf, Germany.

Second author: Dr. Kristi Winters, GESIS, Cologne, Germany

Contact Corresponding Author:

Inga Brentel, Department of Social Sciences, Heinrich Heine University Düsseldorf,
Universitätsstr. 1, 40225 Düsseldorf, Germany. E-Mail: inga.brentel@hhu.de

Funding:

This research was supported by the Digital Society research program funded by the
Ministry of Culture and Science of the German State of North Rhine-Westphalia.

Declaration:

The authors have no potential conflict of interests regarding this article. All authors
have agreed to the submission.

Processing and Documenting Big Data According to the FAIR Principles: A Communications Research Use Case

Abstract

This article redresses a gap in the communications literature by providing a FAIR principles based, best-practice use case for harmonizing and managing big data sources, including digital trace data. We review our conceptual and technical solutions, how we solved big, semi-structured data handling and management by applying the FAIR principles and used metadata to generate digital documentation. We include examples of our digitized outputs, including the individualized, per-variable documentation. The Longitudinal IntermediaPlus (2014-2016) dataset (50GB) is currently Germany's biggest data source on media consumption with online and offline media use, plus socio-demographic metrics, containing over 18,000 harmonized variables for more than 1 million people. We conclude by reviewing the important lessons we learned and recommend ways to use FAIR principles to guide big data management best practices in communication science.

Keywords: big-data data management, FAIR-principles, practical use case, Media-Analyse IntermediaPlus (2014-2016)

Acknowledgement

The original data as well as information on data collection was kindly provided by ag.ma and agof.

1. Introduction: Big Data, FAIR Data and Communications Research

When first confronting the challenge of big data management, we could not find published, peer reviewed use cases detailing communications big data generation, with supporting documentation, according to FAIR principles. We ended up innovating, rather than replicating, due to this gap in the extant literature. Below, we document our FAIR-based, best-practice use case using metadata to harmonize and manage big data, filling a lacuna in communication research data management (Peter et al., 2020; Dienlin et al., 2020) by documenting the creation of big data guided by FAIR principles (Wilkinson et al., 2016; Mons et al., 2017). We first define the FAIR principles. Next, we outline our big data handling challenges in processing the Media-Analysis data source from the *IntermediaPlus*. We describe our conceptual and technical solutions, our four steps to success, the use of metadata to manage our information, and review the practical issues that arose while trying to produce FAIR big data. We conclude by reviewing the important lessons we learned how we used FAIR principles to guide our big data management best practices for communication science. Our use case is intended to be a guide for those confronting unstructured big data.

2. The FAIR Principles

'FAIR' data are increasingly required by funders because the principles ensure data reusability and replicability (Peter et al., 2020; Dienlin et al., 2020; van Atteveldt et al., 2019b), and sustainable data. FAIR stands for *F*indability, *A*ccessibility, *I*nteroperability and *R*eusability (Wilkinson et al., 2016; Mons et al., 2017). Wilkinson et al. (2016) proposed the FAIR principles based on the academic maxims of transparency, interoperability, and reproducibility to increase data reusability and adding value for the research community. The FAIR principles are cited across academia, especially in the context of open science. Mons et al. (2017) explicated what was meant by FAIR, as they felt their principles had been misunderstood as fixed standards, e.g., mandating open access. The principles are not fixed standards to manage research data or other materials, nor is it necessary that the resulting data be open access (Mons et al., 2017, 50), but that is welcomed where possible.

To produce FAIR data, one must “describe a process – mechanised or manual – for accessing discovered data...” “openly and richly...” “to enable evaluation of its utility” (Mons et al., 2017, p. 52) and ensure reusability. Existing guidelines (Peter et

al., 2020) and data management recommendations (Dienlin et al., 2020) view FAIR principles as key, not only in overcoming the replication crisis, but also for fostering data quality and sharing in communication research (van Atteveldt et al., 2019b). FAIR as well as “[o]pen access of data implies that data can be used by anyone, not just by the experts who develop or run advanced codes. If data were openly available (and well described), many more people would work with the data” (Draxl & Scheffler, 2019, p. 6). Also, FAIR data should be easily used in computational methods (Draxl & Scheffler, 2019, p. 3). We briefly review the principles of each letter.

2.1 F- Findable

Data and related research materials must be findable. A “clear, preferably machine readable, license” (Mons et al., 2017, p. 52) is required and should be “published in a trusted repository that can be searched and indexed” (van Atteveldt et al., 2019b, 3938), e.g., GESIS in Germany (Peter et al., 2020). Therefore, “a proper data infrastructure, including documentation, metadata, search engines, and hardware” (Draxl & Scheffler, 2019, p. 7) is required. Using persistent identifiers such as DOIs, metadata, and describing the data via keywords make data findable by humans and machines (Wilkinson et al., 2016). The advantage of findable data is “to avoid doubling of work and thus save human and computational resources and energy” (Draxl & Scheffler, 2019, p. 7). Making research data findable and available as a public good has an ethical component and increases the efficiency of research (Peter et al., 2020; Mons et al., 2017).

2.2 A- Accessible

Important to accessibility is “the formal description of the data, i.e., its metadata that also consider the metadata interrelations” (Draxl & Scheffler, 2019, p. 7). Data must be accessible, meaning retrievable, with its connected persistent identifier (Wilkinson et al., 2016). Access can be restricted by registration were needed, e.g., privacy issues or proprietary data (Mons et al., 2017; see van Atteveldt et al., 2019b). Publishing and archiving data in an institutional, public repository with the support of long-term archiving ensures data accessibility including conforming to readable data file formats ensure use far into the future (Peter et al., 2020)³⁸.

³⁸ see Peter et al.’s recommended social science research repositories for the German association of communication research (DGPuK). The Oxford Academic’s “Research Data Policy” is at: https://academic.oup.com/journals/pages/authors/preparing_your_manuscript/research-data-policy.

2.3 I - Interoperable

“Interoperability and reusability require that data are stored in an open (non-proprietary) format” (van Atteveldt et al., 2019b, p. 3938). Terms, keywords and metadata description must be in useable forms for the wider research community (cf. Wilkinson et al., 2016). The challenge is balancing common formats, tools, programs, etc., against the freedom to create individualized research designs, technical solutions, and workarounds for specific projects on data (cf. Draxl & Scheffler, 2019, p. 7; Haim, 2021, p.72). To ensure interoperability, Draxl and Scheffler (2019, p.7) recommend publishing the programming code, input files, main output files and information on statistical packages and software versions used.

In computational communication research R and Python are commonly used (Haim, 2021; van Atteveldt et al., 2019b). MySQL is a free, open-source software that is compatible, thus interoperable, and easily accessed with most statistical programs. Often MySQL, or NoSQL, databases are used as they are easily accessed with R and Python coding, are non-proprietary and “size-efficient formats aimed at high performance analytics applications” (Mons et al., 2017, p. 51). Both databases facilitate exporting data to other open file formats such as JSON or CSV (van Atteveldt et al., 2019b, p. 3942).

2.4 R- Re-usable

Draxl and Scheffler (2019, p. 3) describe re-usability as data that is “re-purposeable” and allows “another researcher to use the same data for focussing on a different aspect” (p. 10). Thus, “reusability goes beyond strict reproducibility or replicability” (van Atteveldt et al., 2019b, p. 3937). van Atteveldt and colleagues (2019b, p. 3937) emphasise “that data and code are structured, stored, and documented in such a way as to allow and encourage other scholars to adapt them to their specific needs”. Maximal data reusability requires adherence to each of the above principles and “implies transparency and reproducibility” (van Atteveldt et al., 2019b, p. 3937). Mons and colleagues (2017, p. 53) underline the connectivity of all FAIR principles for that of reusability when writing what reusability is not:

“Reuseless data are, for instance, those published as obscure and unstable links to supplemental data in narrative articles, not even (as a set) having a proper, machine-resolvable, Persistent, Unique Identifier (PID) which renders

both the data elements themselves, as well as their metadata, non-machine-readable.”

3. Unstructured Big data: the ag.ma’s Media-Analysis data lake

The analytic power and research potential that free and rich big data offer attracts many quantitative social scientists (Jungherr, Jürgens, & Schoen, 2018, p. 255). Big data, also known as *data lakes*, are a) large repositories of data that are b) stored in a raw format, and c) are usually stored as large-scale and unstructured big data (Maroto, 2019). These researchers will face data handling obstacles quite challenging to those who only worked with structured data, e.g., secondary analysis with published survey data. (cf. Jürgens & Jungherr, 2016, p. 7; Jungherr, Jürgens, & Schoen, 2018, p. 256), and even for experienced users of pre-structured secondary data (Jungherr, Jürgens, & Schoen, 2018, p. 255f.). These unstructured big data require a structure first be imposed on them by the researcher before they can be meaningfully analyzed. No communications research guidelines exist to guide the structuring of unstructured big data to make it useable, nor are there documentation standards for data structuring and processing to ensure they meet the scientific standards of transparency and facilitate re-use and replication.

Our data sources for the pooled *Longitudinal IntermediaPlus (2014-2016)* dataset came from the data lake of ag.ma’s Media-Analysis (MA) cross-sectional datasets in the IntermediaPlus bundle. Published in 2022 this ag.ma bundle was the most recent MA data sources, combining digital trace data for online media use with survey data.³⁹ Prior to our work, the MA datasets were stored in a *data silo*. The original, raw MA datasets were not linked to other systems and therefore inaccessible to the outside world (see Tekiner & Keane, 2013; van Atteveldt & Peng, 2018). The raw data exist in a binary data format (*.dat) not readable by common statistical programs, and thus, not interoperable. Further, the MA raw data were divided into several files per year: one file for each socio-demographic metric, for radio use, press-

³⁹ *Intermedia* and *IntermediaPlus* data bundle are a product of a joint venture between the three main Media-Analysis Agencies: ag.ma, AGOF and AGF/GfK. The joint venture ensures quality by a checks and balance system on data services like measures and interpretation of media penetration. It uses multi-method data collection, including technical measures, of some 100.000 cases for about 3.000 webpages, with on-site and in-app questionnaires and a CATI-questionnaire to around 30.000 respondents biannually. The online bundle excludes online media outlines under public law like ZDF and ARD.

media use, tv use, online use, and others with data with geographic and life-style information on respondents. We had to collect all this information and then structure the different files and formats to be able to grasp the data content. Another hurdle was that the data were available by request only and one needed to deep dive into the organisational and technical structures to be able to convert it into a readable data format let alone understand what the stored variables meant and how they could be interpreted.

The online media use section, made up of digital trace data in the Media-Analysis *IntermediaPlus*, can be characterized by the five Vs of big data (Gudivada et al., 2017, p. 4): volume, velocity, variety, value and veracity.⁴⁰ Volume refers to the number of variables and amount of data. Velocity describes the speed needed for data processing and analysis.⁴¹ Variety refers to the heterogeneity of data, either different formats or types of data (Gudivada et al., 2017, p. 4; van Atteveldt & Peng, 2018; Warin & Sanger, 2014, p. 2). Value in Media-Analysis (MA) relates to overcoming the pitfalls of self-reported data, especially online media consumption, by relying on digital trace data (cf. van Atteveldt & Peng, 2018). Veracity refers to the uncertainty of quality, reliability, and variability at question in big-data sources coming from digital trace data, behavioural data, etc. as it is often unclear which parts of the population are covered by the data, since standards for measures are not yet established.⁴²

4. Applying FAIR to Communications Research Big Data

Social science fields that routinely use cross-national, longitudinal and panel data have discussed, in-depth, survey data collection issues and documentation quality standards (see e.g., Corti et al., 2014; CESSDA Training Working Group, 2017 – 2018).⁴³ This is not yet the case for big data (cf. Jungherr, Jürgens & Schoen, 2018,

⁴⁰ Gudivada et al. (2017, p. 4) uses the five ‘Vs’ referenced here, while van Atteveldt and Peng (2018, p. 85) use four ‘Vs’ leaving out ‘value’. Warin and Sanger (2014, p. 2) reference volume, variety and velocity only. See also Team CESSDA Training (2020, p. 13).

⁴¹ Especially when working with real-time data or doing web scraping (Gudivada et al., 2017); here it only refers to the speed and engine needed to process and handle data volume.

⁴² Veracity is assured via the joint venture that ensures the media consumption measures quality and reliability. This was less of an issue for the data processing where the data representativeness is ensured by survey data accompanying that of digital traces. Also, the criteria of selection of media offerings covered in digital trace data is made transparent, which is often an issue working with Twitter data or other scraped data.

⁴³ CESSDA’s data management guidelines note that saving all versions of big data might not be a suitable way. Still versioning is an important topic for big data.

pp. 255f.). Social science researchers are only starting to discuss big data quality standards (see e.g., Jürgens & Jungherr, 2016; Wilkinson et al., 2016; van Atteveldt et al., 2019b; van Atteveldt & Peng, 2018) ensured by high-quality data management. For communications research, data sharing and open science discussions are becoming more common (e.g., van Atteveldt et al., 2019b; Dienlin et al., 2020; Strippel, 2021). However, communications research has so far failed to systematize and routinize collaborations, especially regarding data sharing and data infrastructures (Strippel, 2021; see also Jarren, 2002; Lauf, 2002). Fostering an open science culture includes supporting replication efforts and normalizing data sharing within the field (Mons et al., 2017, p. 52). To facilitate big data analysis with computational methods, data quality must be assured and sufficient data knowledge is required for researchers to find and understand “structure or patterns in the data” (Draxl & Scheffler, 2019, p. 11) to separate them from the noise of data structure artefacts, collection mode effects and so on. Normalizing quality data management and a data-sharing culture requires a research community with good data management skills that improve the value of the scientific return (van Atteveldt, 2019b, p. 3938) such that data can be repurposed (cf. Draxl & Scheffler, 2019, p. 3).

The FAIR principles are cited in non-big data communications research, in existing guidelines on data management, (Peter et al., 2020) and calls to promote open science (Dienlin et al., 2020; van Atteveldt et al., 2019b). Data quality (management) is extremely relevant to big data, as data quality strongly determines the validity of the (computational) analysis: “It is often said that the biggest challenge for big data is the quality of big data itself” (Gudivada et al., 2017, 2). Thus, “[d]ata quality plays a critical role in computing applications in general, and data-intensive applications in particular” (Gudivada et al., 2017, 1).

We adopted the FAIR principles and the communication research community’s data management guidelines as our quality guidelines (see Peter et al., 2021; Dienlin et al., 2020; Wilkinson et al., 2016; Mons et al., 2017), to ensure “that research objects are reusable, and actually will be reused, and so become as valuable as is possible” (Mons et al., 2017, p. 50). Our research took place at the nexus of discussions around big-data management and its documentation challenges. We wanted to tailor the data to meet the anticipated secondary analysis needs of communications researchers. However, we could not find published, peer reviewed use cases where communications big data and their supporting documentation were generated

according to FAIR principles. We ended up innovating, rather than replicating, due to this gap in the extant literature.

4.1 Our Four Main Challenges

Ensuring that our data and documentation were FAIR required conceptual, technical, and digital solutions. We identified four big data management challenges:

- 1) Plan, design, generate and publish the data management and documentation process according to the FAIR principles (cf. van Atteveldt et al., 2019b, 3927; Hopp et al., 2019, p. 14).
- 2) Solve the accessibility and Velocity technical challenges in processing more than 50 GB of data (Volume). The cases and/or variables totalled to the hundreds of thousands and were in different file formats for relational data that “requires massive storage capacities, high computing power, a lot of energy, etc.” (Warin & Sanger, 2014, p. 3.)
- 3) Manage the Variety of unstructured and semi-unstructured data formats (van Atteveldt et al., 2019a, p. 2) to make our new dataset highly Reusable.
- 4) Use data storage formats that allow primary and secondary data users to conduct complex statistical analysis, maximize computational methods options and increase efficiency in data handling volume and velocity to ensure Interoperability.

5. Our Solution: Four Steps to Success

A well-considered, conceptual structuring plan was required for effective and efficient big data management, variable transformation, and documentation creation. The four-step, FAIR principles guided process (see Figure 2) is summarized below:

- Structure and Adapt: Use metadata structures to conceptually manage data source heterogeneity and data volume technical challenges.
- Import and Harmonize: create a harmonization metadata model (operationalize our concepts with real world metrics) that realizes the conceptual model's goals, import the metadata into CharmStats Pro (more on this software below), and use it to document and digitize data processing.
- Apply Harmonization: use the ‘charmstana’ script to process the data to meet velocity in processing big data.
- Produce Documentation: document the data in ways readable for humans and machines and generate intuitively understandable variable documentation.

5.1 Structure and Adapt

Conceptual organization was key for handling and organizing unstructured big data, as was reviewing and organizing the different file formats. Our goals were conceptual clarity and coherent data structures, reflected by their resulting metadata structures. Another consideration was planning for repurposing by structuring the data in ways and formats useable by others.

5.1.1 Conceptual Structuring

A multi-level structure was determined to best fit our, and the research community's, interests. Structuring was guided by:

1. Identifying extant latent variable concept structures in existing datasets, this is done in information science with *information extraction* (Warin & Sanger, 2014) to manage data heterogeneity (Gudivada et al., 2017, p. 4).
2. Structuring the data to facilitate the lead author's research interests, analysing audiences and market fragmentation online, i.e., news consumption patterns among different thematically genres, as well as socio-demographic metrics and market structures.
3. Structuring the data for others' research utility in analysing, *inter alia*, media consumption and media economics.
4. Generating a final, filtered dataset to answer the lead author's research question with flexible and interoperable data structures to facilitate future research and answer other research questions (Hopp et al., 2019)

After determining a conceptual variables structure (see Brentel, 2021) and outlining the planned structures for documentation, we generated a draft metadata structure with all the information needed for our final codebook. This metadata structure was used to reverse-engineer our workflow strategy, thereby covering every step. We then organized the variables, according to their conceptual structures, into six variable sections and structural subgroups. Each variable section and subgroups were processed separately, building a stand-alone project module. This reduced our technical challenges: it was more efficient to process data sliced into smaller pieces, and the raw data in smaller data files, reducing the data volume and increasing velocity. The resulting data can be organized in database tables using so-called indices. Indices are conceptual structures used for variable names, or facilitating conceptual subgroups

filtering, increasing the readability and accessibility of the final data set for reuse,

The six Variable Sections were: Socio-Demographic; Free-Time Activities; Respondent Belongings; Belongings of the Household; Media Use and Other (see Appendix 1). Each had its own harmonization challenges: socio-demographic variables invariable have follow-up questions while some social background variables have changed over time. The Media Use section had the highest variable volume with different media categories as a further subgrouping: Radio, TV, Press Media and Online Media. For Radio, Press Media and TV, we consulted well-established structuring units, since 'old media' have been studied in various contexts and their data processed accordingly. Further, there are single Media-Analysis datasets for these 'old media' categories, previously processed in academia, so these harmonization plans already exist (see Brentel et al., 2021). However, for the Online Media subgroup, the main object of interest, we needed to invent a variable structure from scratch that met FAIR principles.

5.1.2 Variable Structuring

For Online Media Use, there were a large number of variables without an established way to structure long lists of online offering's names (see Figure 2). A content analysis of those lists within an Excel spreadsheet revealed additional latent structures for online offerings, specifically 'Full Entities' (a whole website offering); 'Single Entities' (one Webpage belonging to a Full Entity); and 'Combined Outlines' (a set of media offerings that have been joined together in the measures of their use). As the research question sought to measure online news consumption at the most granular level available (Brentel, 2023; cf. Webster & Ksiazek, 2012), we extracted variables coded Full and Single Entities from the raw MA datasets to be harmonized and thereby reduce our harmonization work. We excluded the data for Combined Outlines, as it wasn't relevant to the research question and held little value for the research community.⁴⁴

The web entities included contained online offerings with a variety of content types, i.e., eBay and its e-commerce content, Yahoo for e-mails, or newspapers with information-oriented content. Our research interest was overall online-use behaviour, and detailed usage of information-oriented websites (especially in terms of thematic

⁴⁴ The Combined Outlines subgroup can be added to the data processing workflow at any time.

genre). We selected the Business-Model Type of a web entity as the next structuring level. We filtered on one of the most used typologies in economic research, Wirtz's typology (2018, pp. 307ff.) of four business models online: (e-)Commerce, Context – such as online search-engines –, Connection – e.g., websites with communication platforms and Content. By concentrating on those online entities with a 'Content business model', we filtered the information-oriented online offerings. The filtered Single Entities belonging to a Full Entity with a Content business model numbered between about 1.100 and 3.300 entities. This saved significant time and effort otherwise consumed in processing unrelated variables and constituted a reduction of around 25 percent in the overall variable count per year. At the same time, we found a holistic variable structuring to enable later processing of the unprocessed data segments in our research project if needed.

To analyze respondents' online use by genre, the genres were built by an inductive, data-driven half-automated text analysis in Excel of the entity names that a vendor gives the Single Entity assuming a meaningful has been given. Many of same, or quite similar, names were easy to cluster in word clouds and aggregate to genre categories (for details see Kampes, 2020; Brentel et al., 2021). The result was 24 thematic genre categories. These can also be aggregated into different genre typologies used in communication and media research, e.g., newspaper departments, thus are interoperable and re-purposable for different research interests using different theories and genre-concepts. We selected the Single Entity genre as a further structuring level.

Structuring Single Entities by genre categories makes comparative studies possible, and they can be linked to established offline and online genre classification in social science. For 24 categories in the Single Entities, we have on an average 150 variables per category; a number that was easy to handle within *CharmStats Pro* using a special feature we developed for handling large-scale data⁴⁵.

CharmStats Pro was a variable harmonization software solution developed at GESIS. The code was written by Martin Friedrichs in consultation with social scientist Dr. Kristi Winters.⁴⁶ It digitally traces a user's work and creates bespoke harmonized variable documentation outputs. Built from DDI metadata standards, it stores metadata on the study (e.g., study name, collection dates, collection area); question (e.g., multi-

⁴⁵ The "Virtual Variables" feature can match up to 300 variables to one "Truncal Variable".

⁴⁶ GESIS no longer supports this software, however it is available on GitHub (<https://github.com/charmstats>).

lingual question wording, show card response options), variable responses (options and corresponding labels); and values (add a comment to a response value) levels. It was a metadata-based workflow based on the structures of either SPSS or Open Office spreadsheets. These metadata are stored in and connected to each other using MySQL via *CharmStats Pro* (see Figure 4). This allowed metadata to be reused in efficient ways, including generating recoding structure graphs and auto-generating code in SPSS, Stata, Mplus and SAS. Its workflow allows users to make multi-level metadata connections, to import literature and/or references, and it has retrieval system for bespoke, hand-entered comments.

To structure the raw data, we developed an Excel-Sheet using macros to automatize the extension of the harmonization plans when adding more years of the Media-Analysis *IntermediaPlus*. Using a system of several Excel-Sheets, the structuring workflow was made transparent and more automatized. Due to Excel Formulas, cells will give the user errors if any information does not match due to mistakes made during the harmonization planning, such as naming variables over the years. This sped up and simplified the process step and conformed to quality management. Those cases requiring specification are identified because the workflow will capture when they are not unique and require individualized harmonization solutions, such as variable naming for single variables if automatized variable naming results in the same variable names being assigned to two different variables (see Brentel et al., 2021).

In summary, the *Longitudinal IntermediaPlus (2014-2016)* dataset has three levels for structuring online entities (see Figure 2). First is the type of offering online: Full Entity, Single Entity, and Combined-Outline entities. Second, we classified all Full Entity into the five business-model types following Wirtz (2018, pp. 307ff.): Commerce, Context, Connection, Communication and Content. Finally, we defined genres for the Single Entity belonging to Full Entity with a Content business model. This process downsized the total number of variables to about half of the number of entities for data processing and divided different online entities into smaller groups. This preliminary conceptual work guided the practical decisions we took in the software workflow. It also meant we could later organize our data processing as single projects in *CharmStats Pro* and organize our data documentation and database structure for the final, processed *Longitudinal IntermediaPlus (2014-2016)*.

5.2 Adapt the data

Next, we adapted the extant file formats into readable and workable files. This required hours of computer processing that involved handling Excel sheets with more than 2 GB of information. It was a significant amount of work to document the raw data and find meaningful conceptual structures applicable, while also making harmonization plans for the 18.000 variables to be pooled into our new *Longitudinal IntermediaPlus (2014-2016)* dataset. The lead author used a standardized SPSS script in conjunction with Python code to convert the binary formatted *.dat-files to *.sav-files to better read and work with the raw data. We documented this adaption step in detail by saving the standardized script to convert the raw data, ensuring re-usability and replication for other data users working with this binary format. Each year of the MA *IntermediaPlus* bundle was divided into several files. We established a structure for the raw data and the variable naming (see Figure 4). Each year of the MA *IntermediaPlus* bundle was divided in several files. The size of a single file can be between 1.3 and 25 GB.

5.3 Import and Harmonize

To meet the criteria of inter-subjectivity, traceability and provenance (CESSDA Training Working Group, 2017 – 2018; Team CESSDA Training, 2020), a digital method for large-scale data harmonization and documentation was required, one that would ideally also generate the recode text and the complete *Longitudinal IntermediaPlus (2014-2016)* data documentation. Using *CharmStats Pro*, we digitized the data pooling process while simultaneously 1) making the dataset reusable, 2) generating complete and traceable digital documentation, and 3) attaching variable recoding instructions for the software packages SPSS, STATA, MPlus and SAS.

5.3.1 Import

In the Import Step, all our study, question, variable, and value level metadata, along with bibliographical resources, important information and instructions detailing the harmonization decisions were imported into or added to *CharmStats Pro* using a combination of *.sav file importation and hand-entering information (see Brentel, 2021a for details). Figure 5 is an overview of the *CharmStats Pro* workflow.

5.3.2 Harmonization and Technical Innovations

In the Harmonize step, we created target variable-based project folders for the harmonization and data processing work. All data processing activities within the

software are traced and documented to be replicable and transparent. As a result, the data processing and harmonization is fully digitized and accessible. With over 300 Media Use variables all sharing identical metadata we needed a way to automatize the recoding process and generating the aggregated data-documentation to handle the volume and velocity. The solution developed by the lead author and Martin Friedrich advanced the technical capacity of the software to utilize looping as part of the data processing. As a result, a new feature was added to accommodate large-scale, looped, data processing for categorical and metric variables.⁴⁷ This allows users to connect the Source and Target Variable values using easy drop-down menus to that establish mappings (Figure 6) and then set up the looping process (called ‘Virtual Variable’ in *CharmStats Pro*) to generate recoding language for dozens of similarly structured variables. Alternatively, existing harmonized variable recoding files for long-term preservation and access can be documented by pasting it in as text (for a step-by-step guide for *CharmStats Pro* see Brentel, 2021a).⁴⁸

5.4 Produce Output: Apply Harmonization & Create Documentation

After our harmonization projects were completed in *CharmStats Pro* we auto-generated digital outputs in bespoke statistical package’s recoding languages, a graph to map the recodings, and/or a digitized report with the relevant metadata. This meant we could combine multiple target variable *CharmStats Pro* projects into bespoke codebook reports with complete data documentation. The complete documentation and structuring elements keywords are findable, making it is easier to understand the data and the conceptual structure. This information is needed to work with and repurpose the data.

5.4.1 Apply Harmonization

The lead author developed a more automatized workflow to supplement *CharmStats Pro* (see Figure 7). By adapting the raw data with the SPSS script in conjunction with Python code to convert the *.dat-files to *.sav-files, she created a

⁴⁷ *CharmStats Pro* is a Java-based, open-source software that stores to, retrieves from, and connects across harmonization metadata via a MySQL database.

⁴⁸ The *Pressmedia* and *Radio* bundle were originally harmonized in the same traceable and sustainable way. The new harmonized and longitudinal *Pressmedia (1955-2015)* and *Radio (1977-2015)* datasets can be downloaded from GESIS (https://search.gesis.org/research_data).

system of Excel Sheets using Excel Formulas and Macros⁴⁹ to ease the harmonization planning process. These data were then imported into CharmStats Pro for harmonization processing within the looping tool developed for our big data. Automated data-documentation outputs were then created from her newly developed Python script to apply the harmonization, and adapt the data structure into what would be the new dataset *Longitudinal IntermediaPlus (2014-2016)*.

We decided to use MySQL database as our new data storage format. The lead author developed a Python script to map the harmonization information using the re-coding from the *CharmStats Pro* MySQL database with the *.sav-files of the raw MA datasets (the *.sav-files were converted into *.csv format). This automatized the export process for the re-coding syntaxes and ran them on the single files of the raw MA datasets per year. The data was harmonized to the *new Longitudinal IntermediaPlus (2014-2016)* dataset and converted into our final storing structure of the MySQL database *charmstana*.

However, we could not use one column per variable and one row per respondent, the common structure in SPSS or STATA datasets. We had eight columns representing the respondents' identification (BefragtenID), the year the data was collected, the harmonized variable name (HVarName) and label (HVarLabel), the value of a respondent for the variable (HVarWert) and its label (HVarWerteLabel) as well as a marking whether the value is a missing value (is_missing) and the scale of a variable (HVarLevel), the *charmstana* database (see figure 9). The data of a respondent for the different variables is one below the other – so one row now represents the answer/value of one respondent for one variable⁵⁰.

5.4.2 Produce Documentation

To preserve construct validity, users must be able to find and track changes to question wording and response options. Additionally, variables with identical meta-information should be efficiently summarized using the conceptual structures as meta data, making the data documentation more readable. We needed data documentation useful for both primary and secondary data users. This was a challenge given there were more than 18,000 variables, a Variable Sections list (e.g., Media Use, Free-Time Activities, Socio-demographics) and a list of variables per Section (e.g., hh_wohnung

⁴⁹ This happened inside a project team lead by Céline Fabienne Kampes and the lead author. Céline Fabienne Kampes wrote the Excel Formulas. The Excel macros were computed by Niklas Fortmann. Conceptual structuring was developed jointly by Céline Fabienne Kampes and the lead author, supported by Principal Investigators Olaf Jandura and Olexiy Khabyuk (see Brentel et al., 2021).

⁵⁰ see Brentel (2021b) for a description of and access to conceptualizing and building the 'charmstana' script with a step-by-step manual.

[hh_flat], sex, GA_alst_CT_pi). Using a php-script the lead author developed to speed up the exportation-step from *CharmStats Pro*, she further automatized the generation of our digital data documentation and harmonization processes⁵¹.

We tracked variable changes across years and provide information for our harmonized variables (see Figure 9). Effort was devoted to designing and producing transparent, readable, and understandable documentation for secondary users. Hyperlinks for each variable were included in report output, linking the user back to the variable documentation so they need not scroll through the documentation for 18,000 variables. Reports included detailed researcher comments per variable, fetching from within the *CharmStats Pro* comment system, allowing for systematic inclusion of information vital for accuracy in later reuse and replications.

6. Publishing and Archiving

Our output, the *Longitudinal IntermediaPlus (2014-2016)* dataset, is a large, pooled, clearly structured, and reusable German longitudinal media-use data source, covering three years, various media channels, with more than 1 million people. As a result of this work, new research questions can be explored with this newly structured and digitally generated big dataset. The dataset's large N, both in terms of users and media outlets, allows for detailed analyses and enables complex statistical models to the level of specific groups. It is ideal for computational methods, e.g., network analysis and provides a longitudinal, large-scale survey data resource for communication research.

We archived the finalized data and full documentation, guaranteeing long-term access via persistent identifiers (ZA-number and DOI). After user registration, GESIS provides free access for academic use. The Data Archive (DBK) at GESIS provides archiving services and GESIS's SoWiDataNet hosts the research materials accompanying archived data. All the materials needed to replicate our data processing on the raw data is available at SoWiDataNet.⁵² Related research materials that have been published separately are hyperlinked with each other to assist with findability and accessibility.

⁵¹ see Brentel and Fortmann (2021) for a php-script description, a step-by-step manual and the script itself. Niklas Fortmann programmed the php-script.

⁵² DGPuK recommend DBK and SoWiDataNet. Both provide DOIs, metadata and keywords based on social science dictionaries.

This data project comprises of five total publications: the data itself at the DBK and one for each processing step (see figure 2) at SoWiDataNet. The latter includes harmonization plan files, log-files, step-by-step manuals, and the scripts used or adjusted to our needs, fulfilling the requirements for transparency and interoperability of our harmonization work. These are accompanied by a research article explaining the data and the data structuring, complemented by this article on our big data FAIR technical solutions, the digital process-cycle (see figure 9).

7. Conclusion

In this article we explained the solutions we innovated to handle and manage unstructured big data. Although the Media Analysis dataset is specialized, this big data FAIR data management use case is intended to foster data sharing in communication science and promote a stronger institutionalized and professionalized research data provision. Our four-step, FAIR-principles guided process can be generalized to other big data projects in the following ways:

- **Structure and Adapt:** create clear and well-defined metadata structures to conceptually manage your source data's heterogeneity, to help you manage data volume, and anticipate and manage foreseeable technical challenges. Sketch out a mock-up of what you want your final documentation documents to look like, what information they will need to include, and reverse engineer your workflow to ensure each section is completed accordingly.
- **Import and Harmonize:** create and sketch out your harmonization metadata model, operationalize and specify your key concepts with the available metrics to realize your conceptual model's goals. Simultaneously, document this process for later inclusion in your data deposit submission and, ideally, digitize the data processing.
- **Apply Harmonization:** process the unstructured data into a smaller subset of conceptually structured data and segment your subsets to increase velocity in processing your big data.
- **Produce Documentation:** consult specialists at your intended data archive to advise you on how best to document the data in ways readable for humans and machines and generate variable documentation.

Researchers can meet the Findability criteria by including detailed metadata when they deposit their work at a data archive that can host large sized processed dataset and detailed data documentation for long-term preservation. This will ensure your data and full documentation are provided their own DOIs as unique and persistent identifiers, plus be assigned keywords that are searchable by humans and machines via various platforms.

Depositing your data with long term repositories, alongside data-related publications with its hardware support for long-term archiving will ensure your data's Accessibility for the future via readable data file formats and DOIs retrievability. It is also important to provide a detailed abstract, well-considered conceptual keywords for variable blocks, variable names, study, and data collection information that make it as easy as possible for future researchers to use your data.

Meet the criteria of Interoperability by making materials such as variable coding structures, input files and the main output databases available alongside the data and provide information on the software and the statistical packages used detail any technical solutions to facilitate replication and reuse.

Your newly structured big data will meet the criterion of Reusability if you take time to ensure the conceptual structuring the data result not only in the data structures you need but also provide theoretically relevant data structures that will be of most valuable to your research community. To evaluate whether your data processing is fully transparent, ask a peer to review your documentation to evaluate whether they think your documentation is understandable and your final data are fully reproducible.

References

- Brentel, I. (2021a). *Datenaufbereitung der MA-IntermediaPlus: Teil 2*. GESIS Data Archive. <https://doi.org/10.7802/2340>
- Brentel, I. (2021b). *Datenaufbereitung der MA-IntermediaPlus: Teil 3*. GESIS Data Archive. <https://doi.org/10.7802/2341>
- Brentel, I., & Fortmann, N. (2021). *Datenaufbereitung der MA-IntermediaPlus: Teil 4*. GESIS Data Archive. <https://doi.org/10.7802/2342>
- Brentel, I., Jandura, O., Kampes, F. C., & Khabyuk, O. (2021). *Datenaufbereitung der MA-IntermediaPlus: Teil 1*. GESIS Data Archive. <https://doi.org/10.7802/2339>
- Brentel, I., & Winters, K. (2021). The Longitudinal IntermediaPlus (2014–2016): A Case Study in Structuring Unstructured Big Data: Other Humanities. *Research Data Journal for the Humanities and Social Sciences*, 1–16. <https://doi.org/10.1163/24523666-06010001>
- Brentel, I., (2020, June 30). *How to manage Big Data (well) - A digitized process-chain*. CESSDA Training Days 2019 (CTD2019), Cologne, Germany. Zenodo. <https://doi.org/10.5281/zenodo.3923568>
- C. F. Kampes. Welche Genres existieren für Online-Medienangebote? Eine Analyse der Themenstruktur aus Anbietersicht [What genres exist for online media offerings? An analysis of the thematic structure from the perspective of media providers]. In W. Deiters, S. Geisler, F. Hörner, & A. K. Knaup (Eds.), *Die Kommunikation und ihre Technologien. Interdisziplinäre Perspektiven auf Digitalisierung* (pp. 13–44). Transcript Verlag.
- Corti, L., van den Eynden, V., Bishop, L., & Woollard, M. (2019). *Managing and sharing research data: A guide to good practice*. Sage.
- Dienlin, T., Johannes, N., Bowman, N. D., Masur, P. K., Engesser, S., Kümpel, A. S., . . . Vreese, C. de (2020). An Agenda for Open Science in Communication. *Journal of Communication*, 17(5), 1. <https://doi.org/10.1093/joc/jqz052>
- Draxl, C., & Scheffler, M. (2020). Big Data-Driven Materials Science and Its FAIR Data Infrastructure. In W. Andreoni & S. Yip (Eds.), *Handbook of Materials Modeling: Methods: Theory and Modeling* (pp. 49–73). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-44677-6_104
- Gudivada, V., Apon, A., & Ding, J. (2017). Data quality considerations for big data and machine learning: Going beyond data cleaning and transformations. *International Journal on Advances in Software*, 10, 1–20.

- Haim, M. (2021). Gütekriterien und Handlungsempfehlungen für die Entwicklung von Forschungssoftware in der Kommunikations- und Medienwissenschaft. *Medien & Kommunikationswissenschaft*, 69(1), 65–79. <https://doi.org/10.5771/1615-634X-2021-1-65>
- Hopp, F. R., Schaffer, J., Fisher, J. T., & Weber, R. (2019). iCoRe: The GDELT Interface for the Advancement of Communication Research. *Computational Communication Research*, 1(1), 13–44. <https://doi.org/10.5117/CCR2019.1.002.HOPP>
- Jarren, O. (2002). Forschung ist das Kerngeschäft... und Qualität das Ziel. Doch die Defizite sind groß. *Aviso*, 29, 2–3.
- Jungherr, A., Jürgens, P., & Schoen, H. (2018). 12 Twitter-Daten in der Wahlkampfforschung: Datensammlung, Aufarbeitung und Analysebeispiele. In A. Blätte, J. Behnke, K.-U. Schnapp, & C. Wagemann (Eds.), *Schriftenreihe der Sektion Methoden der Politikwissenschaft der Deutschen Vereinigung für Politische Wissenschaft. Computational Social Science: Die Analyse von Big Data* (1st ed., pp. 255–294). Baden-Baden: Nomos Verlagsgesellschaft.
- Jürgens, P., & Jungherr, A. (2016). A Tutorial for Using Twitter Data in the Social Sciences: Data Collection, Preparation, and Analysis. *SSRN Electronic Journal*. Advance online publication. <https://doi.org/10.2139/ssrn.2710146>
- Lauf, E. (2002). Freiheit für die Daten! Sekundäranalysen und Datenbestände in der deutschen Medien- und Kommunikationswissenschaft. *M&K Medien & Kommunikationswissenschaft*, 50, 247–260.
- Maroto, C. (2019). A Data Lake Architecture with Hadoop and Open Source Search Engines: Using Enterprise Data Lakes for Modern Analytics and Business Intelligence. Retrieved from <https://www.searchtechnologies.com/blog/search-data-lake-with-big-data>
- Mons, B., Neylon, C., Velterop, J., Dumontier, M., da Silva Santos, L. O. B., & Wilkinson, M. D. (2017). Cloudy, increasingly FAIR; revisiting the FAIR Data guiding principles for the European Open Science Cloud. *Information Services & Use*, 37(1), 49–56. <https://doi.org/10.3233/ISU-170824>
- Peter, C., Breuer, J., Masur, P. K., Scharrow, M., & Schwarzenegger, C. (2020, December 11). *Empfehlungen zum Umgang mit Forschungsdaten in der Kommunikationswissenschaft*. Retrieved from

https://www.dgpuk.de/sites/default/files/AG_Forschungsdaten%20Empfehlungen%20DGPuK_0.pdf

- Strippel, C. (2021). Forschungsinfrastrukturen für die Kommunikations- und Medienforschung im deutschsprachigen Raum. Initiativen, Bedarfe und Perspektiven. *M&K Medien & Kommunikationswissenschaft*, 69(1), 136–157. <https://doi.org/10.5771/1615-634X-2021-1-136>
- Team, C. T. (2020). Cessda data management expert guide. *CESSDA ERIC, Bergen, Norway*. Advance online publication. <https://doi.org/10.5281/zenodo.3820473>
- Tekiner, F., & Keane, J. A. (2013). Big Data Framework. In *Proceedings: 2013 IEEE International Conference on Systems, Man and Cybernetics: SMC 2013 : 13-16 October 2013 : Manchester, United Kingdom* (pp. 1494–1499). Los Alamitos, CA: IEEE Computer Society. <https://doi.org/10.1109/SMC.2013.258>
- Van Atteveldt, W., Margolin, D., Shen, C., Trilling, D., & Weber, R. (2019). A Roadmap for Computational Communication Research. *Computational Communication Research*, 1(1), 1–11. <https://doi.org/10.5117/CCR2019.1.001.VANA>
- Van Atteveldt, W., & Peng, T.-Q. (2018). When Communication Meets Computation: Opportunities, Challenges, and Pitfalls in Computational Communication Science. *Communication Methods and Measures*, 12(2-3), 81–92. <https://doi.org/10.1080/19312458.2018.1458084>
- Van Atteveldt, W., Strycharz, J., Trilling, D., & Welbers, K. (2019). Computational communication science| toward open computational communication science: A practical road map for reusable data and code. *International Journal of Communication*, 13, 20.
- Warin, T., & Sanger, W. (2014). Structuring big data: How financial models may help. *Journal of Computer Science and Technology*, 2, 1.
- Webster, J. G., & Ksiazek, T. B. (2012). The dynamics of audience fragmentation: Public attention in an age of digital media. *Journal of Communication*, 62, 39–56.
- Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J. J., Appleton, G., Axton, M., Baak, A., . . . Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, 3, 160018. <https://doi.org/10.1038/sdata.2016.18>
- Wirtz, B. W. (2018). *Electronic Business* (6., updated and extended edition). Wiesbaden, Heidelberg: Springer Gabler.

TEXT 4

Online media consumption in Germany: The role of political information

An analysis of German mass communication online

Status of the Article:

The article is currently getting published via the Social Science Open Access Repository (SSOAR)

Authors and Affiliation:

First author: Inga Brentel is a PhD student at the Department of Social Sciences at the Heinrich Heine University Düsseldorf, Germany.

Contact Corresponding Author:

Inga Brentel, Department of Social Sciences, Heinrich Heine University Düsseldorf, Universitätsstr. 1, 40225 Düsseldorf, Germany. E-Mail: inga.brentel@hhu.de

Funding:

This research was supported by the Digital Society research program funded by the Ministry of Culture and Science of the German State of North Rhine-Westphalia.

Declaration:

The author has no potential conflict of interests regarding this article.

Online media consumption in Germany: The role of political information

An analysis of German mass communication online

Abstract

Fragmented, thus, widely scattered, non-overlapping media-consumption patterns often are seen as a logical consequence of increasing numbers in online offerings, specialization and personalization, undermining a media-mediated common ground sufficient for democracy. Empirical evidence yet is missing maybe resulting from data lacking granularity in online-media consumption measured as aggregated online media offerings not detailing the level of single entities (subpages of a website).

Using social network analysis and the theoretical framework of news reading publics, this article exploratively analyses patterns of online-media consumption for ~4,000 single entities of commercially-driven, German websites and 339,423 people. A new methodological approach measuring overlapping media-consumption patterns accounting for individual online-media repertoires is suggested. Using community detection, two thematically driven online-groupings of overlapping audiences characterized by using/not using political- and digital-online-media offerings are identified. However, a total fragmentation in online-media patterns is missing: 43 percent of users observed are part of both news reading publics detected.

Keywords: online news consumption, audience fragmentation, network analysis, issue publics, media use online, audience duplication

Data Availability Statement: *The data underlying this article are available in GESIS Data Archive (DBK) at <http://dx.doi.org/10.4232/1.13530> or using the unique identifier ZA5769*

1. Introduction

Nowadays use of media has become an integral part of our everyday lives (Stark & Kist, 2020, p. 1138). Especially since the rise and diffusion of the internet, mediatization of societies increased. In Germany, 69 percent use the internet regularly, for 40 percent online media are the most important source of information (30 percent excluding social media) (Hölig et al., 2021). Particularly among young people (18 to 24 years), the popularity of online media is high: 70 percent see it as their main source of information. But also, for older people (45+ years) the relevance of internet continuously rises even though tv still is their primary source of information (Hölig et al., 2021). At the same time, the internet is seen as a “high choice” media (Mukerjee, 2021a, p. 3) with a high degree of self-selection due to the sheer number of online offerings available and the option of personalising media consumption online (Nechtushtai & Lewis, 2018; Stark, 2013, pp. 214f.). Therefore, many scholars see a high risk of audience fragmentation in the online sphere undermining “the democratic ideal of a public sphere, by drawing audiences into informational siloes, thereby reducing the common ground needed for informed debate, deliberation, information exchange, and political engagement (Habermas, 1989; Pariser, 2011; Sunstein, 2017; Williams & Delli Carpini, 2011)” (Mukerjee, 2021a, p. 3). Yet, empirical evidence for a siloing effect in online news consumption, where online audiences are highly fragmented, thus, “a situation where people increasingly use media they only share with small groups of like-minded individuals” (Fletcher & Nielsen, 2017, p. 476), is missing. By contrast, empirical studies on online news consumption find “massively overlapping” (Webster, 2014, p. 40) audiences rather than audience fragmentation. These studies are grounded in research on audience duplication, defined as “a situation where the audience for individual outlets may seem small and circumscribed, but most people in practice use many different media, and many media are used by people of many different persuasions” (Fletcher & Nielsen, 2017, p. 476). Or to put it more generally “the degree to which two media outlets share audience members” (Ksiazek, 2011, p. 237). However, audience duplication is analysed with very rough measures of online outlets and is usually restricted to a very limited number of online offerings: the level of measure often is the whole online outlet of a media brand like CNN.com (e.g. Fletcher & Nilsen, 2017; Ksiazek, 2011; Mukerjee et al., 2018; Taneja, 2020; Taneja et al., 2012; Trilling & Schönbach, 2013; Webster & Ksiazek, 2012), media categories like ‘local newspapers online’ (e.g. Olsen, 2020) or ‘Tabloid’ online media (e.g. Ormen, 2018). This aggregated level of measurement does not account for the high-choice options of users in online environments. Tendencies of audience fragmentation may be hidden in these aggregated metrics, as they do not capture the selective use of single subpages. Following, researchers indicated that “the inclusion of more discrete media resources” (Taneja et al., 2012, p. 965) as well as “[f]ar more ‘granularity’ – and a larger sample – is needed to

understand exactly what is being consumed” (Webster & Ksiazek, 2012, p. 51). Yet, sufficient data for more granularity in measuring online media consumption and corresponding analysis were missing.

This study will follow the request of more granularity based on a new and uniquely comprehensive as well as detailed big-data data source on online media use in Germany, the *IntermediaPlus 2014-2016* (Brentel et al., 2020; Brentel & Winters, 2021), which has been prepared for this purpose and made accessible for academia. With an explorative, data-driven approach as proposed by Mukerjee (2021a) online news consumption in Germany is analysed. The data relies on tracking and survey data allowing to measure the actual use of commercial online media at the level of single subpages. It comprises the media use of about 1.6 million people in Germany for more than 4.000 online media offerings (Brentel & Winters, 2021). The explorative study is oriented on the research question what patterns of online-media consumption can be identified: the extent of audience fragmentation vs. duplication and patterns of overlapping online media use meaning the composition of audiences grounded on the subpages-level. It offers insights on:

- the status of audience fragmentation vs. audience duplication in online media consumption in Germany;
- a description of patterns of online media consumption considering supply- and user-side, factors, i.e. the content genre and socio-demographic metrics, that are suspected determinants of audience fragmentation;
- further, in this study an alternative, innovative method to calculate audience duplication is suggested and applied continuing the methodological discussion of Webster, Taneja (2018) and Mukerjee with collaborators (2018b).

This study presents an overall mapping for online news consumption of commercial mass media online including online-born as well as legacy media to integrate case studies or to find new interesting online-media user groups for further research such as media repertoires.

2. Audience fragmentation and the internet

“One of the most widely observed consequences of the growth in digital media is audience fragmentation. As more offerings are delivered on broadband networks and more choices are available ‘on-demand’, patterns of consumption become more widely distributed.” (Webster & Ksiazek, 2012, p. 1)

This is one of the central concerns and a prominent hypothesis discussed in audience fragmentation literature (Stark, 2013, p. 199, 201; Schweiger, 2007, pp. 303f.). Considering the increasing importance of online media as source of information (see Hölig et al., 2021), a high audience fragmentation online challenges the idea of public sphere that “comprises in

essence the communicative institutions of a society, through which facts and opinions circulate and by means of which a common stock of knowledge is built up as the basis for collective political action” (McNair, 2003, pp. 20f.). With communicative institutions he meant mass media that are “the main source and focus of a society’s shared experience” (McNair, 2003, p. 21). Thus, an overlapping media use is needed to form a common ground of values and shared understandings of reality in societies (Stark, 2013, pp. 200f.; Jandura & Weiß, 2017, pp. 104). Especially, with the rise of internet a total fragmentation of public sphere often associated with filter bubbles or echo chambers is concerned. It is rooted in the differentiation of media offerings online, a fragmentation on the supply side (Kampes & Brentel, 2020, pp. 17ff.). With more offerings, media users have more options of selection and to fragment themselves among these different media offerings into smaller audiences (e.g. McQuail, 2010, pp. 444ff.; Stark, 2013, p. 199, 201; Webster & Ksiazek, 2012, pp. 40f.; Mahrt, 2019, pp. 46). Another cause origin from the possibility of personalization online, often associated with a “siloing effect on public conversations” (Nechushtai & Lewis, 2018, p. 306)⁵³. Media supply and advertising industry can use personalization “guiding people’s choice” (Webster, 2018, p. 99), e.g. tracing users and fostering homogeneous audiences. Also, users can personalize their media diet as precise media selection is easier online than in offline context: a user can choose a specific content without browsing through the whole outline of a media outlet like it is the case for newspapers or magazines. Further, media users have a very active role online (Bonfadelli, 2002, p. 72) and a high degree of self-selection (Mukerjee et al., 2018a, pp. 27f.; Bruns, 2019b): there is no program of media content that starts once a person is online like for tv or radio; user need to look for information, actively visit websites and scroll through or click on content (cf. Bonfadelli, 2002, p. 72). Following, media use online is very divers in possibilities and its nature – a high-choice medium (Mukerjee, 2021a, p. 3; Ormen, 2018, pp. 1f.; Zillien & Hargittai, 2009).

Another central assumption underlying the fear of total fragmentation and therewith dysfunctionality of public sphere online roots in the theory of cognitive dissonance and selective exposure. Accordingly, users will turn towards like-minded media and built like-minded communities. That is the fundament of echo chambers (Sunstein, 2007) in many cases combined with the approach of segmentation research. It describes a segmentation of society that is often reflected in fragmented media use, reinforced by media offerings that increasingly specify along lifestyles and interests of smaller target groups, thus generating further segmentation. Disintegration and reinforced (digital) divides in society are seen as consequences (McQuail, 2010, p. 446; Schweiger, 2007, pp. 303f.). As Webster & Ksiazek (2012, pp. 49f.) summarize: “Writers have labelled these audience formations gated

⁵³ Nechushtai & Lewis (2018) offer a systematic overview of the literature on filter bubbles and fragmentation.

communities, sphericules, echo-chambers, cyberbalkans, redmedia-bluedmedia, or, less judgmentally niches and microcultures (Anderson, 2006; Gitlin, 1998; Iyengar & Hahn, 2009; Sunstein, 2007; Turow, 1997, 2006; Van Alstyne & Brynjolfsson, 2005)". All studies have in common that "public attention is limited" (Webster & Ksiazek, 2012, p. 41) while digital media increase in numbers⁵⁴.

3. Three approaches to audience fragmentation

Audience duplication research has started in the 1960s with the diffusion and differentiation of tv-programs and has been used to analyse audience fragmentation. Most measures have been driven by media research improvement of tv-programs and advertising reasons (Mukerjee, 2018, pp. 28f.). At the same time, research on audience fragmentation increased. Webster and Wakshlag (1983) have developed a model to predict and explain tv-program choice including different determinants of program choice and audience fragmentation vs. duplication. They have involved determinants on the supply-side like availability of devices, programs, etc. as well as factors of individuals, the demand-side, like available time, needs, interests, etc.. McQuail (1997) evolved a model of audience formation describing four stages of audience fragmentation. Thereby, audience formation oscillates between two poles: concertation – a state of absolute audience duplication – and break-up of audiences, its disintegration due to absolute audience fragmentation – the absent of audience duplication (*cf.* Fletcher & Nielsen, 2017, p. 480; Stark, 2013). McQuail (1997, pp. 137f.; 2010, pp. 444f.) saw the differentiation of media as a central determinant of audience fragmentation since with increasing diversification of tv-channels the audience formation may develop from a unity model via pluralism and core-periphery models to a break-up model – reflecting a total fragmentation. Concluding, for audience fragmentation two main strings of factors can be identified: market factors such as the differentiation of media offerings online and user-sided factors which underlie the assumption of the need of selectivity in media use. Thus, audience fragmentation and duplication are formed by a duality of market and user factors. Consequently, different approaches to analyse audience patterns have aroused. Webster and Ksiazek (2012) have clustered audience fragmentation literature into two types: media-centric and user-centric studies while suggesting a third, the audience-centric perspective combining both.

3.1 Media-centric

In literature it is the most prominent perspective. It takes a rather structural perspective where factors on the macro-level, those of the media market, determine audience formations – either as audience fragmentation or audience duplication (Taneja et al., 2012, p. 952; Webster & Ksiazek, 2012, p. 42). It "tallies total attendance across outlets or products. This mode of

⁵⁴ For the case of Germany this has been empirically validated by Kampes, 2020a, 2020b, *forthcoming*.

analysis is typified by trend lines, long tails, and power law distributions” (Webster & Ksiazek, 2012, pp. 39f.). Media consumption is in media-centric studies often measured on an aggregated level (Taneja et al., 2018, p. 952), e.g. the reach of media outlets, viewpoint is the media not users (Fletcher & Nielsen, 2017, pp. 480). Online consumption usually is measured as page impressions or the number of unique users of a media offering. Further, metrics of media market concentration on different levels like companies or marketers based on the media reach like the Herfindahl-Hirschman Index (HHI) or Gini coefficient are used to analyse fragmentation on the supply-side (Webster & Ksiazek, 2012, p. 43; e.g. Kampes, 2020a, 2020b, *forthcoming*) as well as diversification of media outlets (e.g. Kampes & Brentel, 2020). Those studies show which media offerings are popular but are missing who is using what. They lack an adequately measure for audience fragmentation and cannot map combinations of media outlets used together – the extent of audience duplication remains unclear. Homogenisation effects in audience stay hidden, indications on segmentation and polarization of audiences cannot be analysed (Webster & Ksiazek, 2012, pp. 44f.).

3.2 User-centric

The individual combinations of different media offerings, often in a cross-media perspective, is the focus of studies with a user-centric perspective. They aim to find and describe “the media repertoires of individual consumers” (Webster & Ksiazek, 2012, p. 40). Usually cluster analysis is a preferred method to analyse audience fragmentation and media consumption on the micro level. Popular in such studies is a description of a typical user or user types by defining their media repertoires. Repertoires, thereby, “are subsets of available media that individuals use on a day-to-day basis” (Webster & Ksiazek, 2012, p. 45; see e.g. Bächler, 2017; Ng & Taneja, 2019). Early studies focussed on tv use and ‘channel repertoires’ (Webster & Ksiazek, 2012, p. 45), lately often cross-media use is considered for media repertoires. However, the individual media consumption usually is based on self-report in survey (Taneja et al., 2012, p. 956). Methodologically, the description of media repertoires is often limited to the number of outlets used not giving information about what is used exactly.⁵⁵ Further, the number of outlets to be reported is very limited either to the most important media sources or a fixed list that cannot represent today’s high-choice media environment (Mukerjee et al., 2018, p. 27). Additionally, self-reported data has various pitfalls (Mukerjee et al., 2018, p. 27; Olsen, 2020, p. 9; Ormen, 2018, pp. 4f.) like social desirability or memorability. Furthermore, this form of data collection mismatches the way news is consumed in the digital age – such as incidental use (Boczkowski et al., 2018) or specific research needs using keywords to find a suitable medium assisted by platforms distributing users across different media outlets (Bruns, 2019b, pp. 2ff.). Opposed to

⁵⁵ Often in self-reported survey data the number of media outlets used on a typical day or week is asked (e.g. Hölig et al., 2021).

rather habitual media use, e.g. known from consumption of newspapers where users (regularly) buy a preferred media outlet⁵⁶. Therefore, reliability and validity of such measures for online media consumption are questioned (Mukerjee et al., 2018, pp. 27, 49; Olsen, 2020, p. 9; Ormen, 2018, p. 4).

Moreover, the aggregated level of (online) media measures are failing to cover the granularity of today's media selection possibilities. Especially considering that media repertoires are seen as "one of several 'coping strategies' people have for finding preferred content in an increasingly complex media environment" (Webster & Ksiazek, 2012, p. 45) and thus, a potential indicator of audience fragmentation. Theoretically this micro-level perspective often is based on the theory of selective exposure and the assumption that media consumption patterns become more and more specialized due to selection pressure evoked by (high) media choice (*cf.* Webster & Ksiazek, 2012, p. 45; Stark, 2013, pp. 214f.). Sometimes it is combined with the uses and gratification theory arguing that media use habits, and therewith media consumption patterns, are a product of different needs people have and want to satisfy by their media choice (gratification) (Mukerjee, 2021a, p. 11; Webster & Wakshlag, 1983; Taneja et al., 2012, pp. 952f.). Typically, the socio-demographic characteristics age, education and gender of users are used to describe differences in media use, such as media repertoires or across user types (Webster & Ksiazek, 2012, p. 45; Schweiger, 2007, pp. 270-278). Studies with a user-centric approach hardly include the perspective of audience overlaps nor abstract findings on media repertoires and user types for an understanding of public audiences (*cf.* Webster & Ksiazek, 2012, p. 45).

3.3 Audience-centric

Due to the shortcomings of one-sided perspectives on audience fragmentation described above, Webster and Ksiazek (2012) suggested a combination of both: The audience-centric perspective focussing on "the extent to which public attention is dispersed across the media environment" (Webster & Ksiazek, 2012, p. 45), thus, lying on the macro-level. "It is user-centric in that it reflects the varied repertoires of audience members, which are aggregated into measures that summarize each audience" (Webster & Ksiazek, 2012, p. 45) and media-centric as it analyses (overlapping) audiences on the level of media outlets (*ebd.*). Hence, it accommodates the duality of media market (structure) and media users (agents) based on the concept of Giddens' theory of structuration (1984). The agents act within a given structure, here the available online media offerings that can be selected by users for their individual consumption. The media market then follows the demand of users, formed by their individual

⁵⁶ Of course, online versions of print outlets can be bought and their online offerings consumed in a more habitual way including the whole outlet. This may primarily be the case for subscribers – a special group of online users.

preferences of media consumption. “As agents use media, they reproduce and alter the structural features of the environment” (Taneja et al., 2012, p. 953 on Giddens’ theory). The supply side, the structure of the online media market, and demand side, composed of online media users as the agents, are interdependent.

Online media consumption is shaped by this self-reinforcing duality of supply and demand side. Therefore, an audience-centric approach is needed to analyse audience fragmentation vs. duplication incorporating both sides. Audiences here are characterized “by the other media they use” (Webster & Ksiazek, 2012, p. 45) and methods of audience duplication research are used to find out the extent of audience overlap across different media offerings used (*cf.* Webster & Ksiazek, 2012, p. 46). For a detailed literature review on audience duplication research see Fletcher and Nielsen (2017) and Mukerjee with collaborators (2018a). Most recent studies on online media consumption have concentrated on social network analysis as suggested by Webster and Ksiazek (2012, pp. 46ff.), especially when applying an audience-centric approach to analyse fragmentation. They either used the suggested measure for audience overlap Ksiazek developed in 2011 (e.g. Taneja et al., 2018; Taneja & Webster, 2016) or that of Mukerjee and colleagues proposed in 2018 as a better, alternative way to build networks of overlapping audiences (e.g. Olsen, 2020; Mukerjee, 2021a). This study will resort to the method of social network analysis to apply the audience-centric approach, but will use a third option to build networks of overlapping audiences considering the role a media offering has in the individual repertoire of a user.

4. Research Approach

This study wants to shed light on the online-media consumption patterns in Germany examining the status of audience fragmentation vs. duplication. In addition, the composition of overlapping audiences will be used to investigate supply- and demand-sided factors that link or separate those audiences as an indication for determinants of audience fragmentation in Germany. Following, the three leading research questions are:

RQ1: What are the patterns of online-media consumption?

RQ2: How fragmented is the online audience?

RQ3: How are fragmented vs. overlapping audiences online composed?

Supplementary the role of online-subpages offering general interest issues, the political “agora” (Neuberger & Lobigs, 2010, p. 132), in comparison to the general public is considered. As for democratic countries like Germany the news consumption of political information is crucial for the (political) integrative function of media and media-mediated political discourse (McNair, 2003; Schweiger, 2007, pp. 302f.). Thus, audience fragmentation is seen more critical in relation to media offerings with political content.

4.1 Theoretical and methodological Approach

The audience-centric perspective is used to meet the duality of audience fragmentation vs. audience duplication considering supply- and demand-sided factors as well as the proposed network analysis to measure and analyse overlapping audiences online. Further the theoretical framework of news reading publics suggested by Mukerjee (2021a) is applied. A news reading public is “a group of individuals who share substantial overlap in their news consumption patterns” (Mukerjee, 2021a, p. 7). Also, it can be characterized “as a set of news (or more generally media) outlets that are consumed by certain groups of people who either seek specific gratifications, or automatically gravitate towards by virtue of shared societal experiences” (Mukerjee, 2021a, p. 7).

The framework of news reading publics for analysis on online media consumption has been developed as a theoretical concept related to theories of media use such as uses and gratification theory, cultural proximity, social identity and the idea of issue publics. These theoretical approaches make “assumptions about what the dimensions of audience fragmentation are” (Mukerjee, 2021a, p. 8). Analysis grounded on one of these theories often formulate hypothesis and group or aggregate data accordingly narrowing the potential of such analysis, e.g. pre-defining social groups or segments to identify their media-repertoires. “The news reading publics framework flips this idea on its head, by not imposing a preconceived model on the data, but instead, allowing the empirical evidence to inform and validate what such a model could be” (Mukerjee, 2021a, p. 8f.).

As the aim of this paper is to understand how fragmented vs. overlapping audiences online are composed to draw a conclusion which factors might shape audience fragmentation, this framework is promising. Similar to Mukerjee this study seeks “to see if the structure of audience behavior that emerges, is in line with what we would expect if these theoretical considerations were valid” (2021a, p. 18) instead of testing causal inferences of one or the other theory on media use.

Moreover, the theoretical approach of news reading publics is suitable for international comparative studies due to its characteristic of a “loosely organised collective” without a fixed theory that has to be applied. The framework of news reading publics can be flexibly adapted to national or regional circumstances by adding or removing theoretical concepts. So, this case study on German online media consumption is internationally connectable as the framework of news reading publics allows “us to identify the structural similarities and differences in the consumption landscapes between countries.” (Mukerjee, 2021a, p. 9).

However, the term "news reading publics" is misleading, as it refers not only to political news, but to the consumption of content-related media in general⁵⁷. Following the theoretical framework of news reading publics and an audience-centric perspective, different attributes of users as well as structural factors of the media market are considered in the analysis on online news consumption in Germany.

4.1.1 Supply-sided explanations for audience fragmentation

In research with a media-centric perspective it is assumed that market concentration forms audience fragmentation – offerings with a high reach produce audience duplication while those of the long tail, i.e. with a low media reach, are associated to be conceptualized to meet special interests of like-minded audiences. Not only interests but also demographic attributes of users are considered by the media market to allow “advertisers to profile individuals [...] creating specialized audiences for different products, thereby ‘segmenting’ what was hitherto a ‘whole’ national audience (Anderson, 2008; Turow, 1998, 2011)” (Mukerjee, 2021a, p. 3). The case of Germany can deliver interesting insights here as the online media market has the structure of a long-tail (Kampes, 2020b): a fragmentation of offerings in the sense of a differentiation and thus an increase in online offerings (Kampes, 2020a), especially those that are entertainment-oriented; an opposite development of concentration, i.e. an increase in the concentration of reach on the offerings that already have a high reach (Kampes, 2020b); and a concentration of online providers and marketers (Kampes, *forthcoming*). This supports the assumption of differentiation into niche offerings, in which a 'segmenting' effect regarding the audience is assumed (Webster & Ksiazek, 2012, pp. 48ff.; McQuail, 2010, p. 446), in order to remain attractive and thus competitive for advertisers, especially with the low reach of the niche offerings of the "tail". Because with a segmented audience of a specific niche offer, advertisers can place (personalised) ads according to their target group (Taneja, 2020; Schweiger, 2007, p. 303).

4.1.2 Demand-sided explanations for audience fragmentation

In literature with a user-centric perspective the socio-demographic prepositions of users are analysed in conjunction with media consumption and audience fragmentation (Webster & Ksiazek, 2012, p. 45; Taneja et al., 2012; Schweiger, 2007, pp. 270-278). Mostly, age, gender and education are included and seen as factors of audience fragmentation with high potential. Sometimes income and socio-economic status are added to measure for financial resources a person has, e.g. to buy new technical devices, for a societal position of a user or a proxy for employment status that can reference on technical or media literacy skills as well as for time

⁵⁷ media whose goal and business model are to generate and disseminate content, as opposed to e-commerce, for example (e.g. Kampes, 2020; see Wirtz, 2008 for more information on business models)

availability (Bonfadelli, 2002, pp. 72ff.; Zillien & Hargittai, 2009, cf. p.275). Further, audience segments in advertising are usually built upon these socio-demographic metrics of users and potential target groups (Bonfadelli, 2002, p. 71). With the internet and the possibility of personalisation, this is another explanation for audience fragmentation online (see Nechtushtai & Lewis, 2018; Pariser, 2011). The majority of studies that explain media use refer to the uses and gratification theory and/or selective exposure. In both cases, it is assumed that different predispositions, interests, partisanship and time availability (needs) lead to different media use to gain gratification. Individuals choose different media preferences according to their preferences and attitudes (selective exposure) (Webster & Wakshlag, 1983, pp. 437ff.; Mukerjee, 2021a, pp. 5f.). In the high-choice environment of online media, this leads to audience fragmentation (Ormen, 2018, pp. 1f.; Zillien & Hargittai, 2009). Explanation is given in different (media) habits, time availability, knowledge, media literacy and resulting gratification needs, especially in case of gender, age, education and socio-economic status (Peiser & Jandura, 2015; Schweiger, 2007, pp. 271-275). But also, affiliation to a social identity, social status, socialisation and the accompanying media use patterns and habits (Peiser & Jandura, 2015; Schweiger, 2007, pp. 271-275).

So, news reading publics differentiated by socio-demographic factors could on the one hand be a result of and explained by the supply-sided long-tail development found in Germany for 2014 to 2016 (Kampes, 2020b) with emerging differentiation in (entertainment-oriented) online offerings, that are mostly online-born (Kampes, 2020a, 2020b). Combined with the assumption of 'segmenting' effects in the audiences based on targeting specific user groups so that the niche offerings remain attractive for advertisers despite the rather small media reach (Schweiger, 2007, pp. 303f.; Taneja, 2020). On the other hand, such patterns of audience duplication reflected in news reading publics could arise from different social identities, habits and gratifications, e.g. from a generational effect where younger people may have different habits and gratification needs in media use than elder people as they grew up in different media environments (Peiser & Jandura, 2015; Schweiger, 2007, pp. 274f.); as well as different socialization, e.g. between women and men, still shaped by divergent gender roles (Schweiger, 2007, pp. 271f.).

4.2 Issue publics as explanation of audience fragmentation

Another promising theoretical explanation is that of *issue publics* meaning that audience formation fragments into thematically specialized audiences (Habermas, 2009, p. 157) that don't overlap at all or have at least a rather low audience duplication (Webster & Ksiazek, 2012, pp. 44.f). Thus, news reading publics online would form according to *content genres*, representing the interest needs of users. Using a media-centric perspective, it is rooted in an economically driven differentiation among content genres to fit assumed audience segments

and simplify personalized advertising on the supply side (Taneja, 2020; McQuail, 2010, pp. 432, 444; Schweiger, 2007, pp. 303f.). A user-centric perspective would see an audience divided along thematically different media as a result of a high-choice environment – meeting the media-centric argument of audience fragmentation as a consequence of thematic differentiation – leading to a high degree of self-selection driven by different interests (Fletcher & Nielsen, 2017, pp. 477ff., 491; Habermas, 2009, p. 157; Zillien & Hargittai, 2009) and needs of the users (Webster & Wakshlag, 1983, pp. 439f.).

4.3 The German online media market

The case of Germany is interesting here as online offerings with different thematically focus are distinguished in their market structure: online-born offerings have a rather high reach and are rather entertainment-oriented while offerings with a political focus are provided by legacy media only (Kampes & Brentel, 2020). At the same time consolidation of providers and marketers in the online media market are observed for Germany, especially between 2014 and 2016 (Kampes, *forthcoming*). These market structures support the idea of a long-tail development with a rising differentiation of offerings and an increase in niche-offerings in Germany (Kampes, 2020b), that can endorse issue publics as well as ‘segmenting’ audiences in favour of personalization and advertisers. On an international level, Germany can be seen as a high-income democracy with a strong state intervention in the media system due to the public service media, which is well-funded. With public service media a media landscape with high audience duplication is associated. However, public service media in Germany is not strongly represented online (Fletcher & Nielsen, 2017, p. 483). In this analysis public service media online is not considered.

In addition, Germany is characterised by a decentralised media system due to federalism and accordingly a strong representation of regional media (Fletcher & Nielsen, 2017, p. 483), also online (Kampes, 2020a). This tends to support strong regional identities, both in terms of social identity and cultural proximity, which can foster audience fragmentation. This contradiction between centralisation through public service media and decentralisation through federal structures was also reflected in the results of the study by Nilsen and Fletscher (2017) on audience fragmentation vs. duplication in cross-media audiences: in an international comparison with France, Spain, Denmark, the UK and the USA, Germany could not be clearly classified. On the one hand, like Denmark and the UK, it had a low density, i.e. in the network of overlapping media use, only a comparatively small proportion of the possible connections representing shared media use could be found (ebd., p. 487). on the other hand, a low value was observed for diameter, which speaks for a lot of overlapping use (ebd., p. 488), as in the case of the USA, Spain and France. Furthermore, large differences in audience duplication structures were found between media types. Thus, overlapping use among offline media is lower than among online media, which has the highest value of overlapping use in terms of

density in an international comparison (ebd., pp. 488, 490). Therefore, the study in Germany is particularly interesting for the online sector with regard to audience duplication, especially because studies are lacking and so far, the focus has been on a cross-media view with strongly aggregated measurements of media use as well as case studies without online media.

5. Data: MA IntermediaPlus 2014-2016

As described already above, a sufficient data source is needed to analyse audience fragmentation, especially when considering the theoretically framework of news reading public. The “empirical data must be available at a sufficiently granular level” (Mukerjee, 2021, p. 9) and should rely on behavioural trace data on a large scale. Initial analysis on media consumption recommended a more detailed look with corresponding data (Webster & Ksiazek, 2012, p. 51) that were not available at the time. This study now can deliver a more granular analysis on online media consumption for the case of Germany based on new, highly detailed data source: The data of *IntermediaPlus 2014-2016* (Brentel et al., 2021) comprises about 4,000 online offerings per year and includes most of the German online offerings in the commercial sector. Here the most recent year, 2016, is considered.

For this analysis we are only interested in online media offerings following a content business model⁵⁸ as those are the ones potentially providing citizens with information. It allows a very detailed analysis of online media consumption patterns regarding commercial media offerings in Germany as a media offering is defined on the level of a single subpage. At the same time, the single subpage with a content-related business model can be assigned to a marketer, a provider, a media brand and its content-genre. The use of a medium is measured by tracking data and fused with survey data. Due to in-App and on-site surveys unique users are identified for different devices used to enter online media and socio-demographic parameters collected. The survey data is based on a representative sample for the German population older than 14 years and includes about 30,000 respondents. One year of the *IntermediaPlus* is based on two surveys, one during spring and one during autumn. Accordingly, online tracking data on a daily basis is matched from the three-month period of each survey data collection. With the behavioural data pitfalls of self-reported data on media consumption are overcome, too (Ormen, 2018, pp. 4f.).

5.1 Operationalisation

The analysis will use Social Network Analysis (SNA) as recommended for an audience-centric research on audience fragmentation vs. duplication, particularly when using large-scale data.

⁵⁸ The data source comprises information on the use of tv, radio and print media. However, an analysis of cross-media use would require aggregation of media measurements to account comparability of usage.

Further a projection of a two-mode network will be executed as described later to create a one-mode network representing overlapping audiences between media offerings. The two-mode network is measured by nodes that either are respondents or online media offerings on the level of sub-pages that have a content business model and a commercial background being part of the *IntermediaPlus 2014-2016* data.

Edges: The edges, also called ties, represent a connection between two nodes in a network. Nodes here can be media offerings or people. A connection between a node symbolizing a medium and a node that embodies a user exists when a person uses a media offering, i.e. the individual media use. Media use is measured on a daily basis within a three-month period and comprises the actual use of a single media offering by a person.

Thematic genre: The *thematic genre* was gained from an inductive text analysis conducted for the names of the single subpages to define genre-categories (Kampes, 2020, pp. 32ff.). The result are 23 thematic genres: advice, car, career, culture, digital, economy, entertainment, family, finance, forum, games, health, knowledge, lifestyle, breaking news, news, newsletter, politics, regional, soccer, sports, style and travel as well as four additional categories: brand name, homepage, video and other including landing pages (Kampes, 2020; Brentel et al., 2021). Thus, thematic genre is measured with 27 categories. In this study the following thematic genre are seen relevant for *issues of general interest* as they are “relating to matters of information and culture that are of wide interest and concern in a society, without being addressed to any particular individual” (McQuail, 2010, p. 5): economy, knowledge, breaking news, news, politics, regional and homepage for media outlines coming from press media or having a breaking-news section on their homepage, like e-mail providers (e.g. t-online.de - Homepage).

The data of socio-demographic metrics is based on the answers given by the respondents in CATI surveys (Brentel et al., 2020, 2021). In total for 2016 the data comprises 339,423 respondents and 274,193 respondents are considered as online users.

Gender: Gender is a dichotomous variable comprising ‘male’ and ‘female’.

Age: The age of a respondent is measured with twelve categories in four-year steps starting at 14 to 19 years and having ‘70 or older’ as final cohort.

Education: The formal education is measured in four categories: ‘elementary/primary/lower secondary school’, ‘secondary school’, ‘high school/matriculation standard’ and ‘university’. Here they are aggregated to three categories for low, mid and high educational level, whereby the category ‘matriculation standard’ and ‘university’ have been aggregated as a high education.

Income: The variable income is based on the income of a respondent and is measured in 500-Euro steps in eight categories. Beginning with ‘no income’ and ‘less than 500 Euros’ while ending with ‘3,000 Euros and more’.

6. Method

6.1 Analysis of news consumption patterns

Analysis with behavioural trace data on a large scale is challenging due to the “noise” included in such data complicate to find meaningful results (Mukerjee, 2021, p. 9). Social Network Analysis (SNA) is a sufficient method for behavioural, large-scale data on the micro-level helping to reduce complexity and filter for noise. Further, it is appropriate to analyse audience fragmentation and detect overlapping audiences. To reduce complexity the two-mode-network is projected to a one-mode network comprising media offerings and representing overlapping audience structures. As in previous studies on audience duplication, the media-projection will be used (*i.a.* Webster & Ksiazek, 2012; Taneja et al., 2012; Mukerjee et al., 2018a; Olsen, 2020). Figure 1 exemplifies the principle of such a projection. It shows how projection can help to identify patterns (communities of overlapping audiences) drawn from behaviour (media use) of individuals (users). In a media-projection, media offerings are connected if they have been used by the same user and thus share their audience to some extent. The strength of the share in audience, the audience duplication, is determined by the shared ties (media use) of the two-mode network.

There are several options to measure the weight of such ties and consequently to calculate the strength of audience duplication between the media offerings. In literature a discussion about the best way to consider the weight of shared use in a media projection was held (see Mukerjee et al., 2018a, 2018b; Webster & Taneja, 2018) to measure audience duplication. Both approaches of projection follow the idea of filtering “deviation from-random duplication” (Ksiazek, 2011, p. 240). Randomness is defined as the expected overlap considering the media reach of two media outlets. The underlying assumption is, that only duplication above the “degrees of freedom” is significant duplication (Mukerjee et al., 2018a, p. 33). Thus, they want to filter the media-projected network only for audience overlap above the “degree of freedom”. To do so the media reach of two media outlets is taken and “the degree to which the observed duplication between two outlets differs from the expected duplication between those outlets” (Ksiazek 2011, p. 240) calculated. Ksiazek (2011) assesses the χ^2 calculations using the percentage of media reach for each outlet. So far, the idea proposed by Ksiazek (2011) is dominating but is questioned by Mukerjee and colleagues in 2018 (2018a, 2018b). Their main criticism is “[t]he decision to use percentages instead of frequencies disregards, in other words, the substantive differences in reach. Hence, this approach treats as equivalent the least and the most accessed outlets, thus offering a misleading interpretation of the strength of the overlap – a weakness that affects the rest of the analyses applied to the network” (Mukerjee et al., 2018a, p. 32). They propose to use the frequencies as it is done in a χ^2 test instead of using percentages for the media reach of outlets and resort to the calculation basis of phi

correlation to find significant audience duplication. They call it the “phi correlation approach” (Mukerjee et al., 2018a, p. 35).

6.2 Alternative Measure of audience duplication

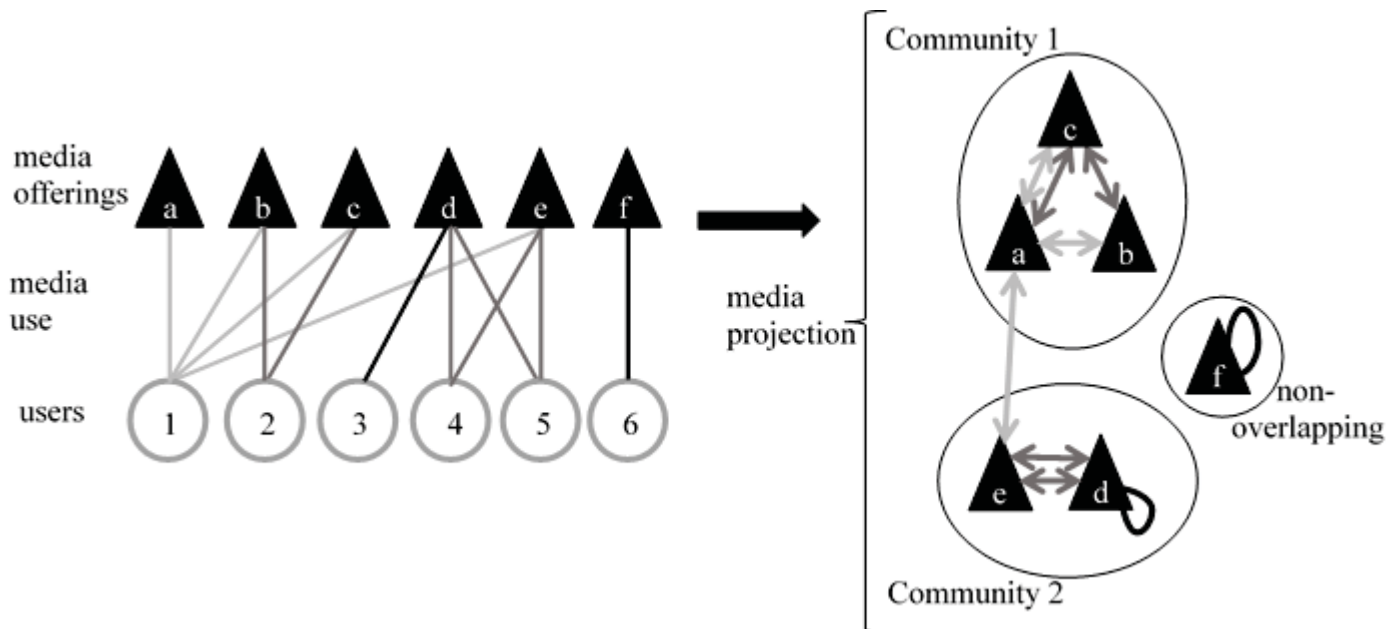
At first glance the approaches using χ^2 statistics or phi correlations seem comprehensible to find ‘significant’ overlap in media use and filter for the noise of big-data. It accounts the potentially large dispersion and thus ‘irregular’ appearing use that distinguishes the usage of the medium Internet from legacy medium. However, it does not measure the weighting of an outlet within the (online) repertoire of a user determined by the frequency of use and the number of outlets used. Thus, it fails to measure audience duplication from a user-centric perspective accounting for the role a media offering has in individual information seeking online considering the number of media offerings used and their frequency of use by an individual online. SNA-projection techniques would allow a more user-centric measure of audience duplication and respectively fragmentation.

Hence, distinct to Ksiazek (2011) and Mukerjee with colleagues (2018) this study proposes a new, innovative way to measure audience overlaps in media-projected networks accounting for the role of an online offering in the online repertoire based on the frequency of use and repertoire size. Since the role that a media offer plays in the individual media repertoire is of great importance in the evaluation of audience duplication and fragmentation. Especially, in context of (questioned existence of) media-mediated common ground in societies (Pariser, 2011; Sunstein, 2007). As the behaviour is measured as the frequency of use instead of used vs. not used. This granularity should be considered in the projection since it makes a difference if one uses a media offering once or several times and thus the relevance of an overlap with another user differs (see figure 1). For example, if two people use the same two media offerings online and this is the only media offerings they use, their overlap should count as relevant, regardless if the audience duplication of these two media is above the “degree of freedom” based on their media reach – see *user 4* and *5* in figure 1 forming *community 2*.

Consequently, the sparse matrix methods are used (Lietz, 2020a) here to differentiate between strong and weak ties, i.e. the audience duplication measured by overlapping use of two media offerings by two or more individuals: The weight of an overlap between two media is determined by the importance of these media in the individual (online) repertoires; Each user enters the calculation with “1”, so that a normalisation and comparability is established. A user who uses many media (and that a lot) – like *user 1* in figure 1 – enters the calculation of the overlap for the media-projection equally as a user who only uses two media, e.g. *user 4* in figure 1. So, an overlap between two media is strong if it is used by many users and depending on the importance both media play in the individual consumption patterns (see Lietz, 2020a, 2020b for further methodological details and matrix calculation). In the end, for a projected network, the entire network could be filtered for strong vs. weaker connections, but are not

excluded in advance. Thus, the overlap of audiences always depends on the importance of a medium in the individual usage patterns.

Figure 1: exemplification of media-projection to identify media consumption patterns



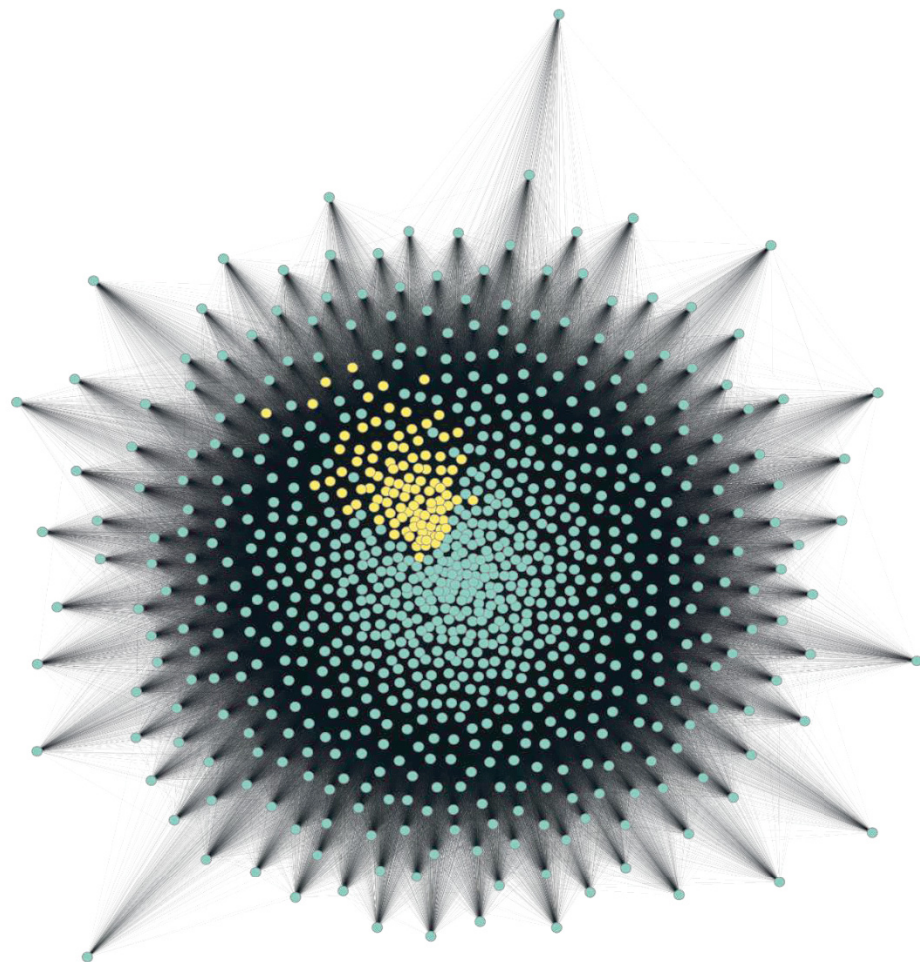
Finally, a community detection in SNA is executed for the media-projected network, common in audience duplication research to identify overlapping audiences (Mukerjee, 2021b, p. 1). The community detection is based on similar statistical mechanisms like cluster analysis: based on the ties between two media offerings, that represent overlapping use of at least two people, similarity of these pairs is calculated and assigned to a community accordingly. “Communities, if identified in the network, would imply the presence of certain outlets sharing significantly higher audience overlap than two outlets selected at random” (Mukerjee, 2021a, p. 21). As exemplified in figure 1 for the media offerings a and e, due to the granularity of measure for overlap in the media-projection, two media offerings that are used jointly (with a low frequency of use) by only one person (*user 1*) with a fairly large online repertoire can be assigned to different communities if they have a stronger share with other media offerings – for e it is d and in case of a it is c with b in figure 1. The algorithm of Louvain is used for the community detection, considered to best fit large-scale networks (Barabasi, 2016) and “work remarkably well even in the presence of substantial noise, when many of the others fail to identify any community” (Mukerjee, 2021b, p. 9). A detected community can be interpreted as a news reading public as “the presence of news reading publics is reflected in these clustered communities” (Mukerjee, 2021a, p. 21), but is not equivalent. However, “news outlets corresponding to a single news reading public [are expected] to be more tightly connected to each other, than to outlets corresponding to other news reading publics” (Mukerjee, 2021a, p. 21).

7. Results

7.1 RQ 1: Patterns of online news consumption – community detection for overlapping audiences

Two communities of overlapping audiences could be identified in the media-projection (see figure 2) indicating two news reading publics online in Germany for 2016. Already 1,025 media offerings (53.7 %) were excluded due to the media-projection as they were not or not jointly used with other media. Thus, the base for community detection were 883 media offerings (46.3 %) which are used with others; 88.5 percent of them (n=781) are part of the first community, the blue one in figure 2, and 11.6 percent (n=102) are assigned to the second community, the yellow one. Consequently, the blue community is by far the largest when it comes to the number of media offerings comprised. As the visualization of figure 2 implicates the

Figure 2: media-projected network for overlapping online use in Germany in 2016



communities of overlapping audiences are not completely separated from each other but connected. Following, the two news reading publics identified have a strong audience duplication within their community, but also have overlapping audiences with each other based on weaker ties. Like *community 1* and *2* in figure 1 share the *user 1* as part of their audience,

but hold stronger connections, i.e. a higher proportion of shared audience, with the other media of their community.

7.2 RQ 2: Fragmented or overlapping patterns – An analysis of distribution of users across the communities

Thus, a total fragmentation in the online audiences among the yellow and blue community does not exist. Instead, they share parts of their audiences: when analysing the demand-side of the media-projection three user groups can be identified – those using exclusively media offerings of the blue community (6.3 %; n = 11,509) or those of the yellow community (50.8 %; n = 92,877), plus, users that have a share in both communities (42.9 %; n = 78,433).⁵⁹ However, there is a certain degree of audience fragmentation between the exclusive users of the respective community. Concluding, it is evident that a separation of audiences cannot be identified due to the audience duplication found forming the two communities and even across them. It should be mentioned that around 33.3 percent of people do not use the tracked online subpages and therefore do not overlap with the audiences of the communities.

7.3 RQ3: Composition of communities: issue publics – thematically driven news reading publics

7.3.1 Thematic genre

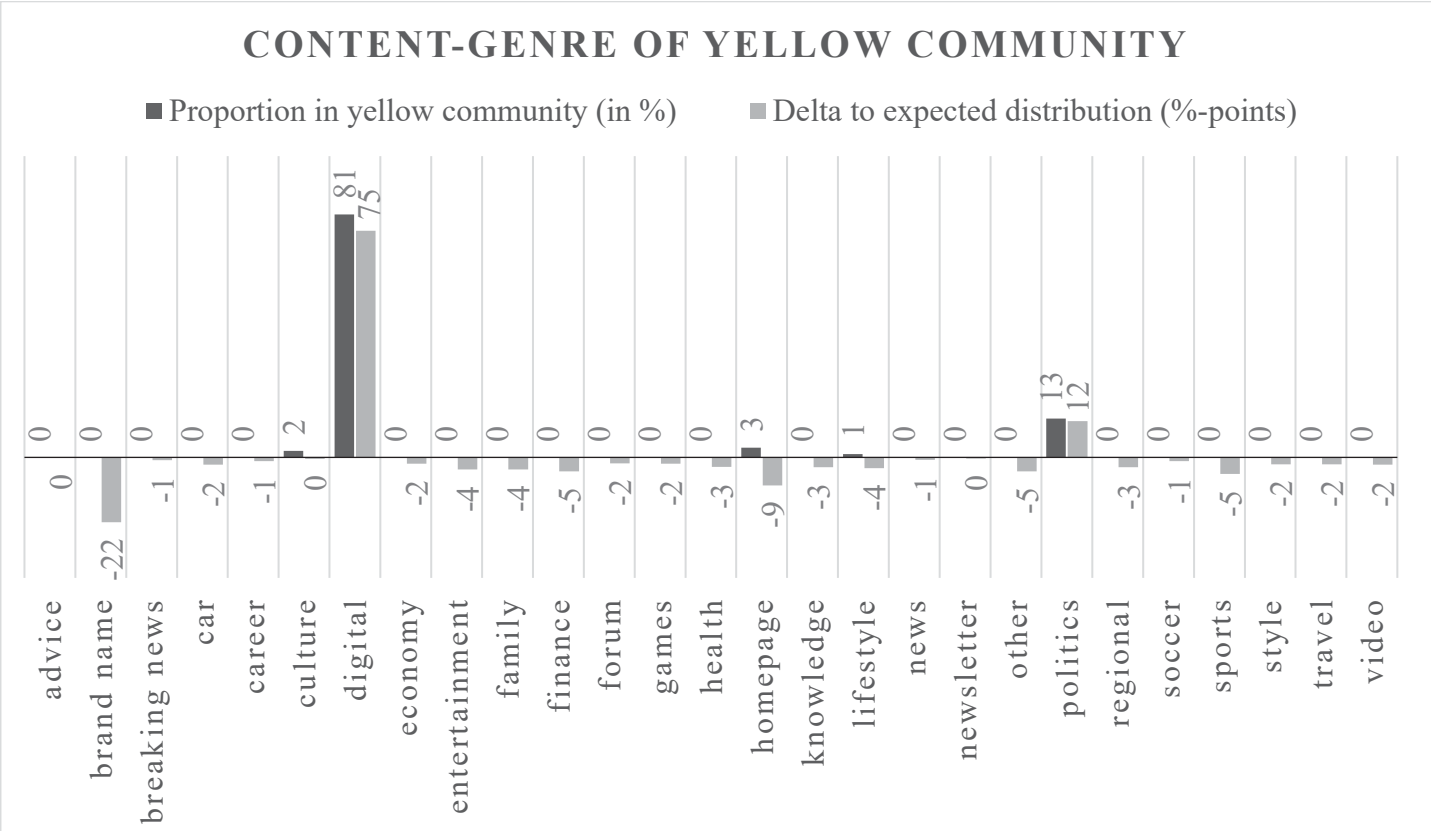
After examining the distribution of supply and demand characteristics in the communities, it becomes clear that issue publics, i.e. thematically driven news reading publics, can most distinctly explain the clustering of audience duplication. Only five of the 27 content-genre categories occur in the yellow community: digital, politics, homepage, culture and lifestyle (see figure 3). Almost all online offerings assigned to the digital and politics genre – further called digital- and politics-offerings – are part of the yellow network community meaning they have an extremely high audience duplication. People using politics-offerings, over-proportionally seem to use digital-offerings too but not so much offerings with different thematical focus. In contrast, especially digital-offerings are almost absent in the blue community, only one offering is assigned to the blue network (see figure 4). Also politics-offerings are rare having a share of 0.5 percent in the blue community.

The high audience duplication of digital- and politics-offerings indicated by the yellow network community is surprising as from a market-centric perspective those are the genres that are most different and can be seen typically for online-born (digital genre) on the one side of the

⁵⁹ There are also people without overlaps in their use / not using subpages which is 33.3 percent of the two-mode data; they are excluded from the calculations of the numbers for the shares in overlaps.

market and offline-originated (politics genre) on the other side (Kampes & Brentel, 2020). Thus, the clustering of digital- and politics-offerings into one network community seems best to be explained by interest – consequently the explanation of issue publics may differentiate the online public in Germany at least in case of audiences of digital- and politics-offerings, that are strongly overlapping. It could result from interest lead selection (selective exposure) or out of the need of information seeking (gratification).

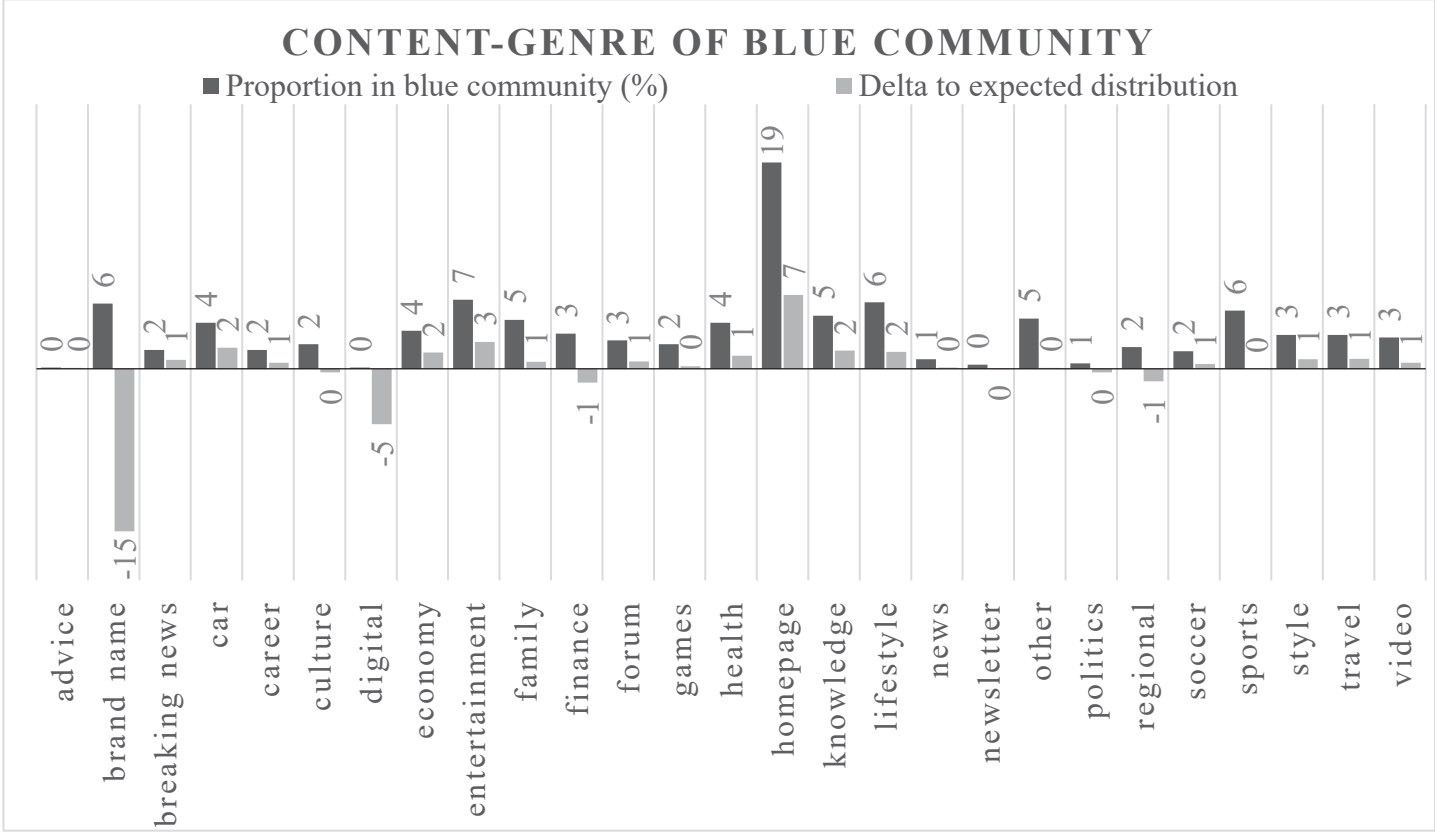
Figure 3: Distribution of content-genre in the yellow community of the media-projection showing the deviation (delta) to an expected distribution based on the overall distribution in the data



The news reading public indicated by the blue network community seems to be united by the use of online offerings assigned to the genre homepage (homepage-offerings) that make about 19 percent of the community’s offerings lying five percentage points above the expected share assuming an equal distribution across all online offerings in the data (n = 1908). Further, offerings assigned to entertainment (+ 3 %-points), car, economy, lifestyle and knowledge (+ 2 %-points) are above the expected share; while offerings assigned to family, other, sports and entertainment also seem to be comparatively popular among users in this group of overlapping audience with a share between five and six percent meeting or being above the expected share (see figure 4). However, characteristic for the news reading public reflected by the blue community in the media-projection seem to be a thematically rather unspecific use of online

media offerings oriented towards superficial, fairly broad information seeking indicated by the use of homepage-offerings and entertainment-offerings (see figure 4).

Figure 4: Distribution of content-genre in the blue community of the media-projection showing the deviation (delta) to an expected distribution based on the overall distribution in the data



This is in favour of an explanation based on the long-tail theory: It indicates a high overlapping use among niche offerings that are said to be rather among entertainment-oriented offerings (Kampes, 2020b) instead of an assumed fragmenting effect. Further it could be interpreted as an interest driven news consumption pattern where individuals unite various gratification needs that result in the use of different special-interest offerings. Or to say it with the words of Alexander Bruns (2019a, 49): “As a long history of media and cultural studies inquiry has made abundantly clear, we each have multiple identities, which we perform according to the social contexts we find ourselves in – and, especially in online and social media environments, those contexts are more and more often collapsing onto one another.”

7.3.1 Socio-demographic metrics

Considering socio-demographic metrics of users, further, the gender seems to determine if users belong exclusively to the blue community (women are + 49 %-points above the expected share) or to the yellow community (men are +16%-points above the expected share). As displayed in figure 5, for users belonging to audiences of both network communities it is pretty balanced (+ 8%-points towards women).

Figure 5: Distribution of gender among user groups showing the deviation (delta Δ) to an expected distribution based on the overall distribution in the data

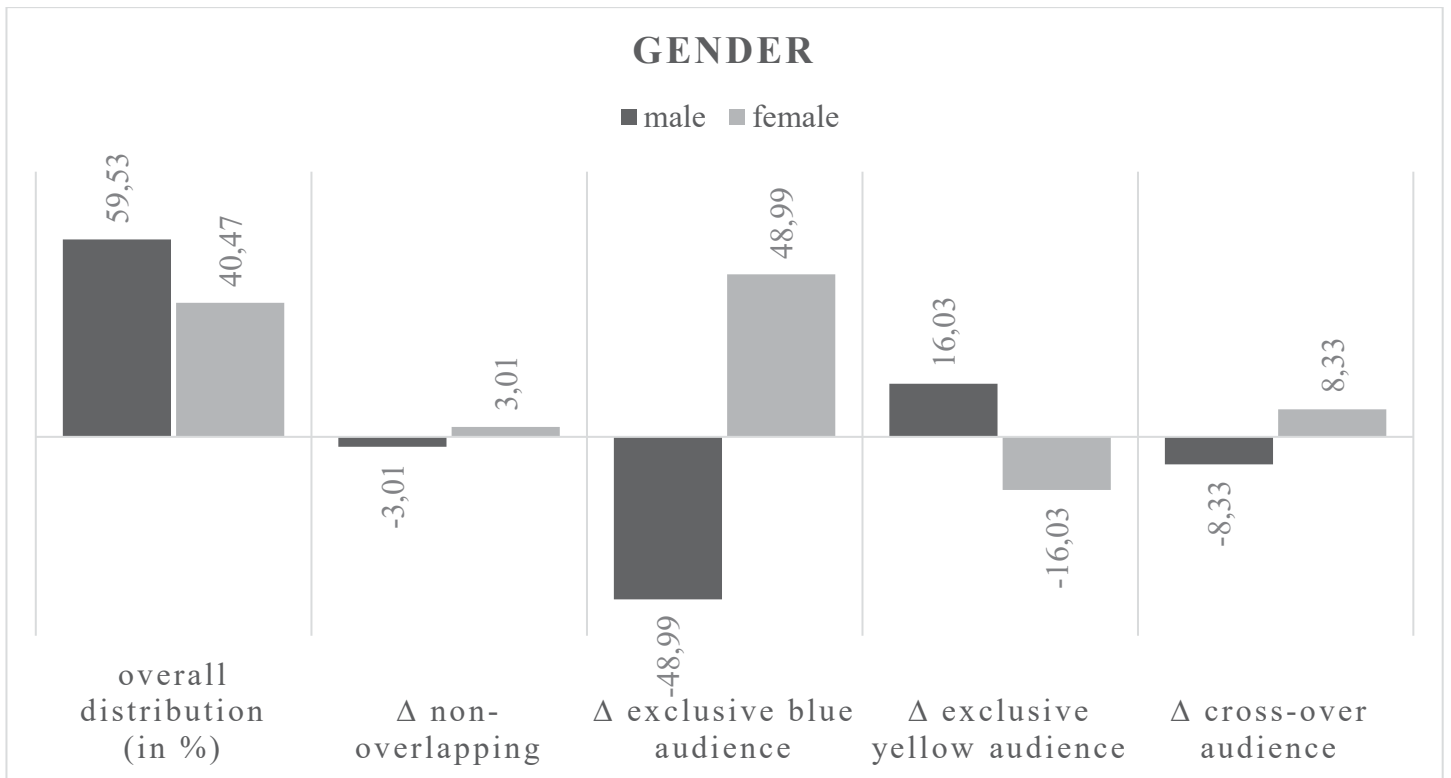
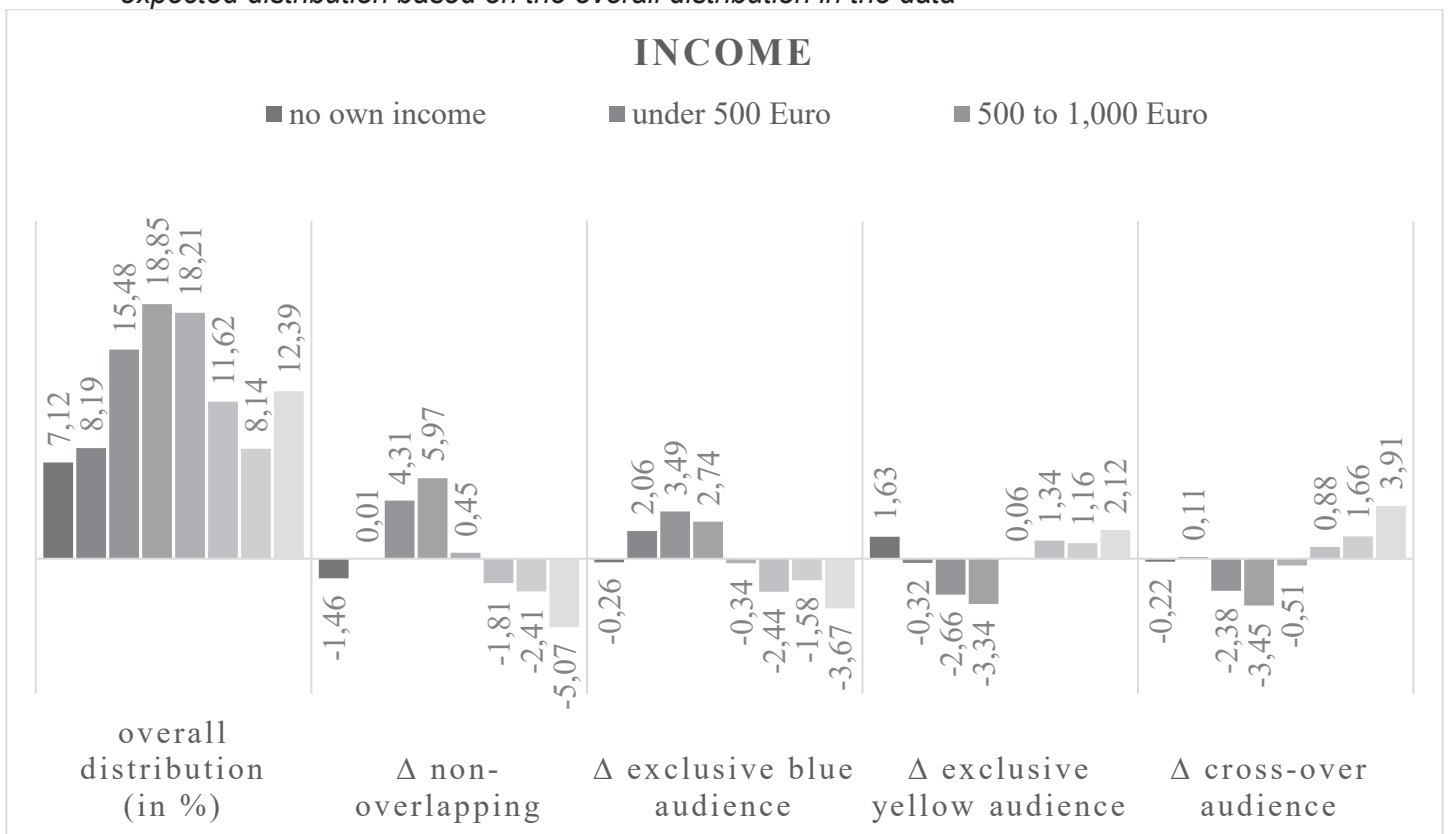


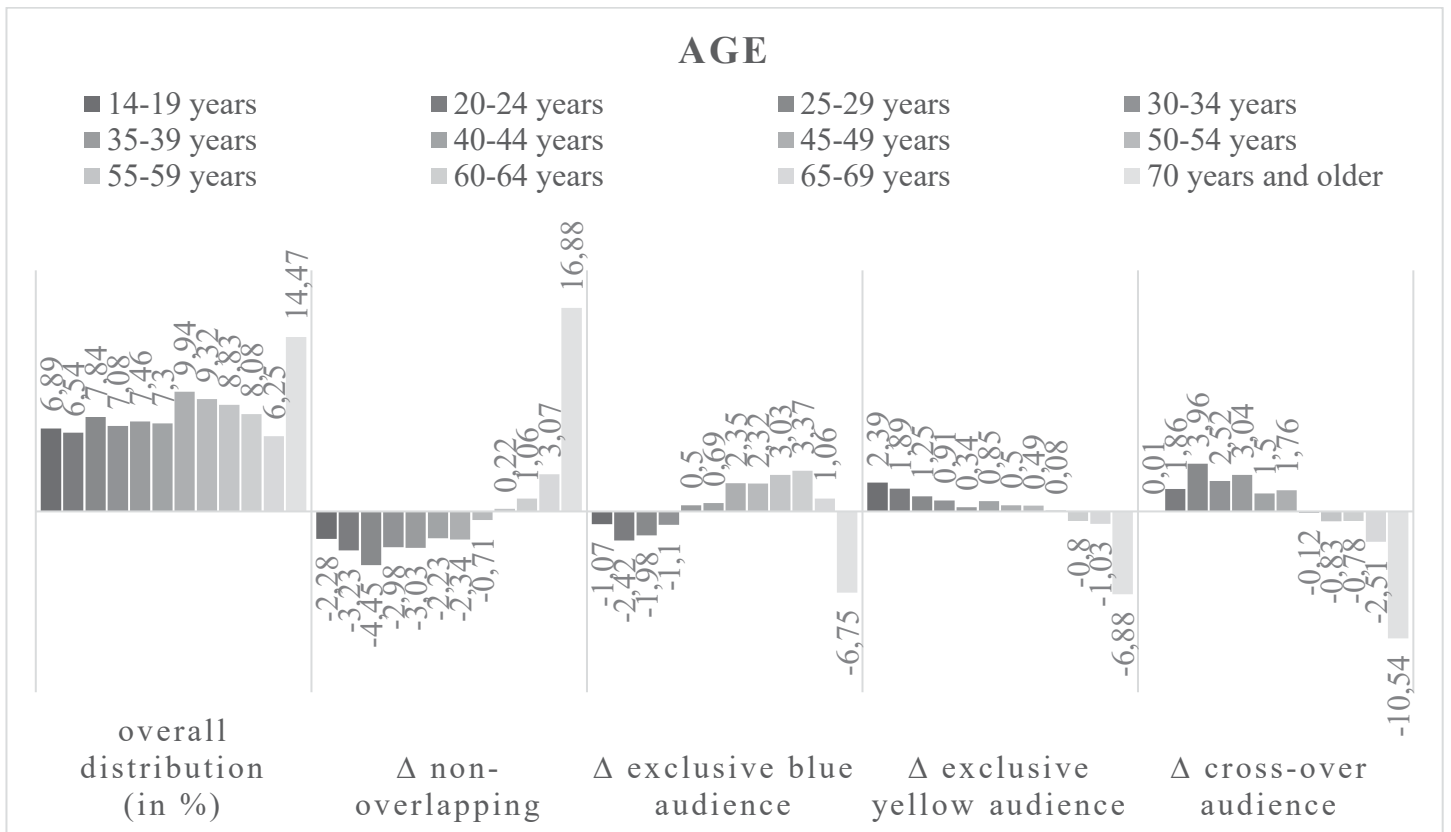
Figure 6: Distribution of income among user groups showing the deviation (delta Δ) to an expected distribution based on the overall distribution in the data⁶⁰



⁶⁰ Note that 6.9 percent didn't answer the question on their income and are not displayed in the graph.

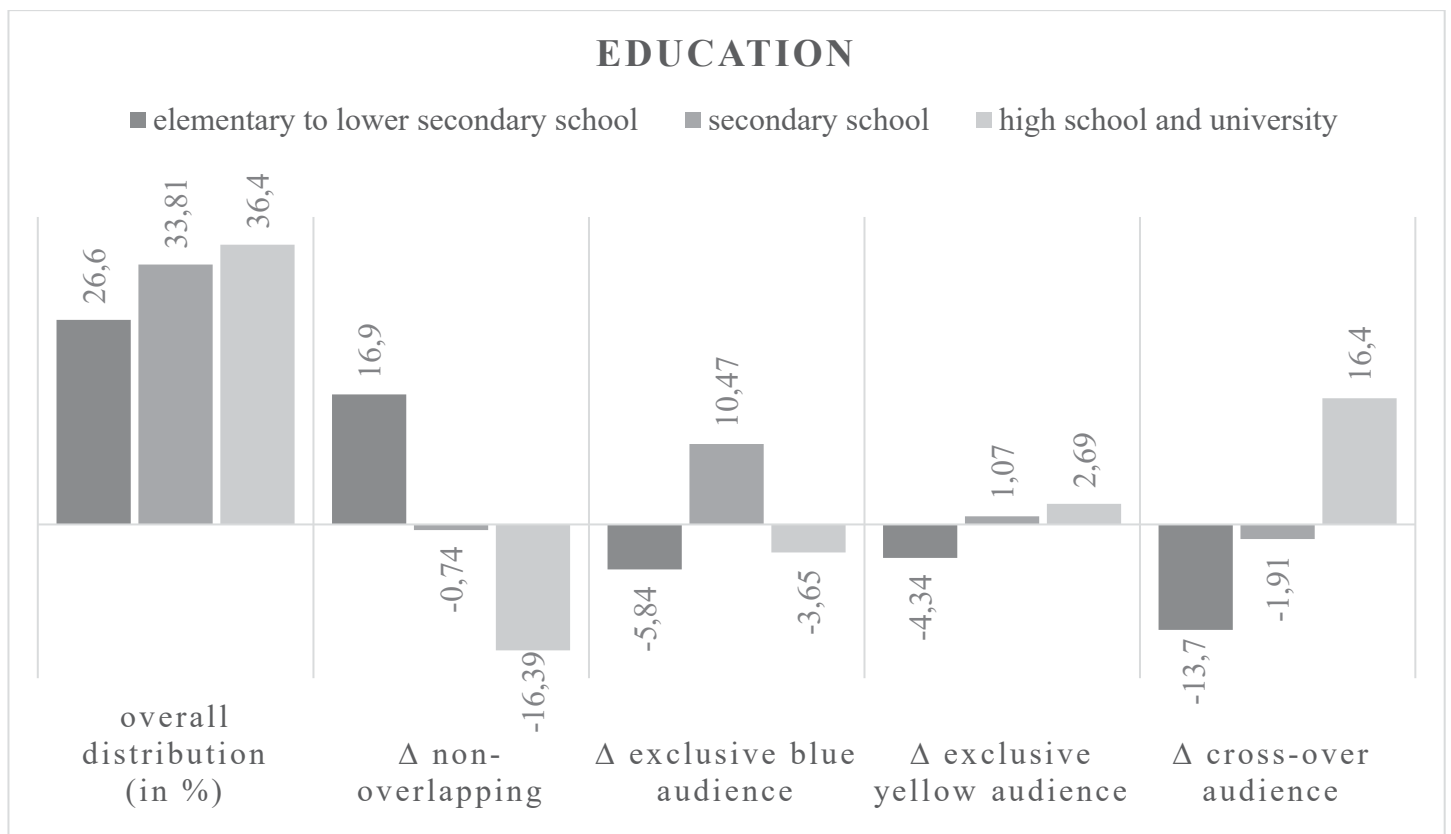
The age and educational level as well as in parts income (see figure 6) appear to impact the media use patterns and thus overlapping audience structures, too, particularly when it comes to whether media are used (with others) at all: elder people (55+) often are non-(overlapping) users, especially when they are older than 70 years (deviation of + 17 %-points) like figure 7 shows. Same applies to people with a low educational level (deviation of + 17 %-points), see figure 8.

Figure 7: Distribution of age among user groups showing the deviation (delta Δ) to an expected distribution based on the overall distribution in the data



For the distribution of the exclusive communities, the picture is rather mixed, with a tendency for younger people towards the yellow community (14 to 54 years with a total deviation of + 9 %-points, that is decreasing with rising age) than towards the blue community where the 35-year olds to 69-year olds are represented above the expected shares (with a total deviation of + 13 %-points). Further, a tendency for the medium educated towards the blue community (with a deviation of + 11 %-points) can be observed. Users who belong to both communities (cross-over users) and correspondingly have overlaps with both distinct audience groupings, are mainly high educated people. However, the cross-over users do not contribute to a fragmented pattern, but rather increase the degree of audience duplication as they are audience to the online offerings of both identified communities.

Figure 8: Distribution of education among user groups showing the deviation (delta Δ) to an expected distribution based on the overall distribution in the data



The high divergence of men and women in exclusively using online offerings of one network community and thus detached news reading publics when ignoring cross-over users could be explained by different socialisation or to put it differently: a result of shared societal experiences forming different media consumption habits (see Peiser & Jandura, 2015) as well as different interest, e.g. men are said to use more political information and digital content than women (Witting, 2018, p. 469; Marth, 2019, p. 125) as they are supposed to have on average a higher political interest (Peiser & Jandura, 2015). The same applies to the differences in media use patterns, which find expression in the audience duplication structures of the network communities, among younger vs. older people. One possible explanation is that younger people have already grown up with the internet and thus may have acquired different media use habits than older people who were socialised without the internet; this is a promising explanation, especially when it comes to understanding the large proportion of non-users among the 70+ age group. Growing up in different media environments may also have led to different cultural proximities or social identities, which is a further explanation for the age gap in the communities of overlapping media use, the news reading publics of digital- and political-online offerings vs. those with a fairly unspecified and more entertainment-oriented focus. Nevertheless, the audience behaviour that emerges in the clustered communities and can be interpreted as the two news reading publics observed in the German online sphere for 2016, seems rather to be explained by the theoretical considerations of issue publics; because the

two identified communities are mainly marked by the different thematic orientation, which in particular seems to be shaped by the overlapping usage patterns of the audience of the yellow community. Even if differences in the socio-demographics of the respective users are recognisable, more complex connections, such as intersectionality, seem to be present here for the formation of overlapping usage patterns. Only in the case of the difference between users vs. non-users are there clear indications of a plausible explanation based on age.

8. Fragmented or overlapping patterns of online news consumption in Germany? – A summary

With the spread of internet and its technical possibilities of personalization as well as increasing differentiation of online offerings, the discussion about audience fragmentation rekindled. Due to the high number of available media offerings online with low entry barriers, audience fragmentation is seen as a logical consequence resulting from media consumption patterns that “become more widely distributed” (Webster & Ksiazek, 2012, p. 1). Resulting informational silos are feared by scholars (e.g. Sunstein, 2007), undermining a media-mediated common ground of shared knowledge, beliefs and values sufficient for democracy (Stark, 2013; Habermas, 1989; Williams & Delli Carpini, 2011). However, this study, too, cannot find empirical evidence for these fears, as was the case with previous studies. Although this study had a much greater level of granularity in the data due to behavioural trace-data and thus analysis to better measure and identify audience fragmentation (Ormen, 2018, p. 5), which according to the fragmentation thesis is occurring along the increasingly thematically differentiated online media (Stark, 2013, pp. 199f.), a large overlap in the audiences was again identified. Instead of “the information cocoon fragmenting our society and democracy that Sunstein, Pariser, and even Obama warned of (...) communities of interest that accumulate around shared topics and identities, but do not detach altogether from the world beyond” (Bruns, 2019a, p. 49) have been identified for the patterns of media consumption among German, commercially-driven online offerings for the year of 2016.

Using social network analysis, audience duplication between media offerings was made visible on the basis of individual usage patterns and groupings of overlapping media usage were identified by community detection. Given the duality of audience fragmentation of supply- and demand-side an audience-centric perspective was applied to investigate the composition of the identified network communities. The framework of news reading publics has proven useful, especially the explanatory approach of issue publics, where overlapping use is characterised by interest-driven use and is formed along certain thematic areas. hence two news reading publics were identified: One is characterised by a strong audience duplication of digital- and politics-offerings. while the other comprises rather entertainment-oriented media and more

superficial information offerings, like homepage-subpages, and is thematically less specific characterized.

The granularity of the data and the new methodological approach to assess media projection networks for overlapping usage allows for a more accurate description of the identified network communities than before; investigating audience duplication and fragmentation more detailed. Now it is easier to identify the user groups that form these audience duplication communities. For example, it became apparent that about six percent of the users exclusively use the media offerings of the blue network, although this comprises significantly (over eight times more) more online offerings than the yellow network community which is used by almost 51 percent. One could say, these two user groups detached in their media consumption. However, 43 percent of users, the cross-over user group, are consuming offerings from both network communities. This massively increases audience duplication and makes a total fragmentation of the blue and yellow network communities representing the two identified news reading publics unlikely. Nevertheless, within the blue network community, there is a greater potential for further fragmentation, as many online offerings are included, but a much smaller proportion of users are using them (the cross-over users also have a higher proportion in the yellow cluster). The users of the yellow cluster, on the other hand, have more potential to split off from other audience groups, as they have a much narrower scope of topics that they use online. However, as these are general interest media, policy-specific information, this is less of a concern. In total, it seems as if the concentration of overlapping audiences for media offerings of the yellow network is highest, concluding it can be interpreted as the core of the online public sphere regarding to McQuails' model of audience formation (1997, p. 138; 2010, pp. 444f.). Due to the high number of media offerings in the blue network community, characterized as thematically fairly disperse and less users, especially in relation to that of yellow media offerings, it can be argued to be the periphery of the online sphere.

Nevertheless, this is a first explanatory study on online news consumption patterns in Germany with this granularity and more research is needed to draw sound conclusions about audience fragmentation online in Germany. Particularly, because a new methodological approach was used to measure overlaps in media use as well as an innovative theoretical framework just recently suggested by Mukerjee (2021a). Further limitation lies in the data that misses information on public service media and social media, thus it offers a clear delimited section of digital media available. Moreover, audience fragmentation is subject to processuality, here a snapshot of 2016 is given and further analysis on more recent years needed. As well as more detailed analysis of news consumption patterns using the option of filtering for strong vs. weak ties in the network and conduct more community detection or other methods to examine the user groups more, e.g. zooming on the level of their individual media repertoires, considering intersectionality to research segmenting effects or cross media consumption. Continuing

studies considering other levels of fragmentation, like the reception of consumed content, are inevitable, because “[i]n a hyperconnected yet deeply polarised world, the most important filter remains in our heads, not in our networks: it is the cognitive filter that makes us reject some ideas out of hand” (Bruns, 2019a, p. 61). This means people can consume the same media offerings, but interpret and use it differently; mainstream media can also be used to monitor them and “attach new disinformation” towards them (Bruns, 2019a, p. 49).

References

- Anderson, C. (2006). *The long tail: Why the future of business is selling less of more*. Hachette Books.
- Giddens, A. (1984). *The constitution of society: Outline of the theory of structuration*. Univ of California Press.
- Bächler, J. (2017). Medienrepertoires und Fragmentierung: Impulse zur Deutung der Medienkrise. In W. A. Meier (Ed.), *Medienstrukturen: Vol. 11. Abbruch - Umbruch - Aufbruch: Globaler Medienwandel und lokale Medienkrisen* (1st ed., pp. 319–342). Nomos Verlagsgesellschaft mbH & Co. KG. <https://doi.org/10.5771/9783845276663-319>
- Barabási, A. L. (2016). *Network science*. Cambridge: Cambridge University Press. <http://networksciencebook.com/>
- Boczkowski, P. J., Mitchelstein, E., & Matassi, M. (2018). “News comes across when I’m in a moment of leisure”: Understanding the practices of incidental news consumption on social media. *New Media & Society*, 20(10), 3523–3539. <https://doi.org/10.1177/1461444817750396>
- Bonfadelli, H. (2002). The Internet and Knowledge Gaps. *European Journal of Communication*, 17(1), 65–84. <https://doi.org/10.1177/0267323102017001607>
- Brentel, I., Kampes, C. F., & Jandura, O. (2020). *Meta-Information des Samples der Media-Analyse Daten: IntermediaPlus (2014-2016)*. <https://doi.org/10.7802/2030>
- Brentel, I., & Winters, K. (2021). The Longitudinal IntermediaPlus (2014–2016): A Case Study in Structuring Unstructured Big Data: Other Humanities. *Research Data Journal for the Humanities and Social Sciences*, 1–16. <https://doi.org/10.1163/24523666-06010001>
- Brentel, I., Kampes, C. F., & Jandura, O. (2021). *The Longitudinal IntermediaPlus Data Source (2014-2016)*. ZA5769. GESIS. <https://doi.org/10.4232/1.13530>
- Bruns, A. (2019a). *Are filter bubbles real?* Polity Press. <https://ebookcentral.proquest.com/lib/ulbd/reader.action?docID=5887595>
- Bruns, A. (2019b). Filter bubble. *Internet Policy Review*, 8(4). <https://doi.org/10.14763/2019.4.1426>
- Trilling, D. & Schoenbach, K. (2013). Patterns of News Consumption in Austria: How Fragmented Are They? *International Journal of Communication*, 7(0), 25. <https://ijoc.org/index.php/ijoc/article/view/1769>
- Fletcher, R., & Nielsen, R. K. (2017). Are News Audiences Increasingly Fragmented? A Cross-National Comparative Analysis of Cross-Platform News Audience Fragmentation and Duplication. *Journal of Communication*, 67(4), 476–498. <https://doi.org/10.1111/jcom.12315>

- Gitlin, T. (2002). Public sphere or public sphericules? In *Media, ritual and identity* (pp. 178–184). Routledge.
- Habermas, J. (2009). *Europe: The flatering project*. Polity, Cambridge.
- Habermas, J. (1989). *The Structural Transformation of the Public Sphere: Vol. 9. print*. <https://doi.org/10.1017/S0010417500017527>
- Hölig, S., Hasebrink, U., & Behre, J. (2021). *Reuters Institute Digital News Report 2021: Ergebnisse für Deutschland*. <https://doi.org/10.21241/SSOAR.73637>
- Iyengar, S., & Hahn, K. S. (2009). Red media, blue media: Evidence of ideological selectivity in media use. *Journal of Communication*, 59(1), 19–39.
- Kampes, C. F., & Brentel, I. (2020). The German online media market: Online-born information offerings and their audiences – A shift towards digital inequalities? *World of Media. Journal of Russian Media and Journalism Studies*, 1(4), 5–34. <https://doi.org/10.30547/worldofmedia.4.2020.1>
- Kampes, C. F. (2020a). Welche Genres existieren für Online-Medienangebote? Eine Analyse der Themenstruktur aus Anbietersicht. In W. Deiters, S. Geisler, F. Hörner & A. K. Knaup (Hrsg.), *Die Kommunikation und ihre Technologien. Interdisziplinäre Perspektiven auf Digitalisierung* (S. 13–43).
- Kampes, C. F. (2020b). Tail or no Tail? Applicability of the Long Tail Theory to the German Online Media Market. *Central European Journal of Communication*, 13(3), 371–389.
- Kampes, C. F. (forthcoming). *Two-Sided German Online Media Market at its Interlinking Actors: Structural Competition of Marketers and Media Providers*.
- Kösters, R. (2020). *Medien als Mittler im Konflikt? Der Streit um die Migration im Spiegel der Berichterstattung*. Düsseldorf: Dissertationsschrift an der Heinrich-Heine-Universität Düsseldorf.
- Ksiazek, T. B. (2011). A Network Analytic Approach to Understanding Cross-Platform Audience Behavior. *Journal of Media Economics*, 24(4), 237–251. <https://doi.org/10.1080/08997764.2011.626985>
- Lietz, H. (2020a). Drawing impossible boundaries: Field delineation of Social Network Science. *Scientometrics*, 125(3), 2841–2876. <https://doi.org/10.1007/s11192-020-03527-0>
- Lietz, H. (2020b). compsoc—Notebooks for computational sociology. Retrieved December 6, 2021 from <https://github.com/gesiscss/compsoc>.
- Mahrt, M. (2019). *Beyond Filter Bubbles and Echo Chambers: The Integrative Potential of the Internet*. ifpuk - Institute for Media and Communication Studies at FU Berlin. <https://doi.org/10.17174/dcr.v5.0>
- McNair, B. (2003). *An introduction to political communication* (3rd ed.). *Communication and society*. Routledge.
- McQuail, D. (1997). *Audience analysis*. SAGE Publications.

- McQuail, D. (2010). *McQuail's mass communication theory* (6. ed.). SAGE.
- Mukerjee, S. (2021a). Rethinking Audience Fragmentation Using a Theory of News Reading Publics: Online India as a Case Study. *Pre-Print*. Advance online publication. <https://doi.org/10.31219/osf.io/3f47z>
- Mukerjee, S. (2021b). A systematic comparison of community detection algorithms for measuring selective exposure in co-exposure networks. *Scientific Reports*, *11*(1), 15218. <https://doi.org/10.1038/s41598-021-94724-1>
- Mukerjee, S., Majó-Vázquez, S., & González-Bailón, S. (2018a). Networks of Audience Overlap in the Consumption of Digital News. *Journal of Communication*, *68*(1), 26–50. <https://doi.org/10.1093/joc/jqx007>
- Mukerjee, S., Majó-Vázquez, S., & González-Bailón, S. (2018b). Response to Webster and Taneja's Response to "Networks of Audience Overlap in the Consumption of Digital News". *Journal of Communication*, *68*(3), E15-E18. <https://doi.org/10.1093/joc/jqy022>
- Nechushtai, E., & Lewis, S. C. (2019). What kind of news gatekeepers do we want machines to be? Filter bubbles, fragmentation, and the normative dimensions of algorithmic recommendations. *Computers in Human Behavior*, *90*(8), 298–307. <https://doi.org/10.1016/j.chb.2018.07.043>
- Neuberger, C., Lobigs, F., Herbers, M. R., Karthaus, A., & Nuernbergk, C. (2010). *Die Bedeutung des Internets im Rahmen der Vielfaltssicherung: Gutachten im Auftrag der Kommission zur Ermittlung der Konzentration im Medienbereich (KEK). Schriftenreihe der Landesmedienanstalten: Vol. 43. Vistas.*
- Ng, Y. M. M., & Taneja, H. (2019). Mapping User-Centric Internet Geographies: How Similar are Countries in Their Web Use Patterns? *Journal of Communication*, *69*(5), 467–489. <https://doi.org/10.1093/joc/jqz030>
- Olsen, R. K. (2020). Connecting People? Understanding Media's Role as Democratic Resources for People in Digitally Advanced Local Environments. *Digital Journalism*, *8*(4), 506–525. <https://doi.org/10.1080/21670811.2019.1679029>
- Ørmen, J. (2019). Not So Distinct After All: Assessing Social Stratification of News Users on the Web. *Journalism Studies*, *20*(11), 1653–1670. <https://doi.org/10.1080/1461670X.2018.1539342>
- Pariser, E. (2011). *The filter bubble: How the new personalized web is changing what we read and how we think*. Penguin.
- Peiser, W., & Jandura, O. Political Media Use. In *Donsbach (Hg.) 2015 – The international encyclopedia of communication* (pp. 1–6).
- Schweiger, W. (2007). *Theorien der Mediennutzung: Eine Einführung. Springer eBook Collection Humanities, Social Science*. VS Verlag für Sozialwissenschaften. <https://doi.org/10.1007/978-3-531-90408-5>

- Stark, B. Fragmentierung Revisited: eine theoretische und methodische Evaluation im Internetzeitalter, 199–220. <https://doi.org/10.5771/9783845249278-199>
- Stark, B., & Kist, E. L. (2020). Mediennutzung. In J. Krone & T. Pellegrini (Eds.), *Handbuch Medienökonomie* (pp. 1137–1163). Springer Fachmedien Wiesbaden. https://doi.org/10.1007/978-3-658-09560-4_57
- Sunstein, C. R. (2017). *#Republic*. Princeton University Press. <https://doi.org/10.1515/9781400884711>
- Sunstein, C. R. (2007). *Republic.com 2.0*. Princeton, NJ: Princeton University Press.
- Taneja, H. (2020). The Myth of Targeting Small, But Loyal Niche Audiences: Double-Jeopardy Effects In Digital-Media Consumption. *Journal of Advertising Research*, 60(3), 239–250. <https://doi.org/10.2501/JAR-2019-037>
- Taneja, H., Webster, J. G., Malthouse, E. C., & Ksiazek, T. B. (2012). Media consumption across platforms: Identifying user-defined repertoires. *New Media & Society*, 14(6), 951–968. <https://doi.org/10.1177/1461444811436146>
- Taneja, H., Wu, A. X., & Edgerly, S. (2018). Rethinking the generational gap in online news use: An infrastructural perspective. *New Media & Society*, 20(5), 1792–1812. <https://doi.org/10.1177/1461444817707348>
- Turow, J. (1997). *Breaking Up America: Advertisers and the New Media World*. Chicago: Univ. Chicago Press.
- Turow, J. (2006). *Niche Envy: Marketing discrimination in the digital age*. The MIT Press. <https://doi.org/10.7551/mitpress/5016.001.0001>
- van Alstyne, M., & Brynjolfsson, E. (2005). Global village or cyber-balkans? Modeling and measuring the integration of electronic communities. *Management Science*, 51(6), 851–868.
- Webster, J. G. (2018). 6. Audience Behavior. In P. M. Napoli (Ed.), *Mediated Communication* (pp. 91-104). De Gruyter. <https://doi.org/10.1515/9783110481129-007>
- Webster, J. G. (2014). *The Marketplace of Attention: How Audiences Take Shape in a Digital Age*. MIT Press.
- Webster, J. G., & Ksiazek, T. B. (2012). The dynamics of audience fragmentation: Public attention in an age of digital media. *Journal of Communication*, 62(1), 39–56.
- Webster, J. G., & Taneja, H. (2018). Building and Interpreting Audience Networks: A Response to Mukerjee, Majo-Vazquez & Gonzalez-Bailon. *Journal of Communication*, 68(3), E11-E14. <https://doi.org/10.1093/joc/jqy024>
- Webster, J. G., & Wakshlag, J. J. (1983). A theory of television program choice. *Communication Research*, 10(4), 430–446.
- Weiß, R., & Jandura, O. (2017). Medien und gesellschaftlicher Zusammenhalt. In O. Jandura, M. Wendelin, M. Adolf, & J. Wimmer (Eds.), *Zwischen Integration und Diversifikation*:

- Medien und gesellschaftlicher Zusammenhalt im digitalen Zeitalter* (pp. 11–31). Springer Fachmedien Wiesbaden. https://doi.org/10.1007/978-3-658-15031-0_2
- Williams, B. A., & Delli Carpini, M. (2011). *After Broadcast News: Media Regimes, Democracy, and the New Information Environment*. Cambridge University Press.
- Witting, T. (2018). Digitale Ungleichheiten. In E.-U. Huster, J. Boeckh, & H. Mogge-Grotjahn (Eds.), *Handbuch Armut und soziale Ausgrenzung* (pp. 457–477). Springer Fachmedien Wiesbaden. https://doi.org/10.1007/978-3-658-19077-4_20
- Zillien, N., & Hargittai, E. (2009). Digital Distinction: Status-Specific Types of Internet Usage. *Social Science Quarterly*, 90(2), 274–291. <https://doi.org/10.1111/j.1540-6237.2009.00617.x>