NOTORIOUS

HAW • HBI • ISD

Cross-Plattform Social Media Forschung: Methoden der Daten-gewinnung und Analyse

Philipp Kessling, Mattes Ruckdeschel, Felix Victor Münch, Gregor Wiedemann

HAW HAMBURG

LEIBNIZ-INSTITUT FÜR MEDIENFORSCHUNG HANS-BREDOW-INSTITUT

ISD Institute for Strategic Dialogue

GEFÖRDERT VOM

Bundesministerium für Bildung und Forschung

Inhaltsverzeichnis

Autor*innen

Philipp Kessling ist Wissenschaftlicher Mitarbeiter im Media Research Methods Lab des Leibniz-Instituts für Medienforschung | Hans-Bredow-Instituts und forscht dort zu Methoden der Nachverfolgung von Informationsdiffusionsprozessen mittels Spurendaten. Zuvor forschte er am Competence Center Communication und dem Department Information der HAW Hamburg zur Verbreitung von Propaganda im digitalen Raum.

Mattes Ruckdeschel ist Wissenschaftlicher Mitarbeiter im Media Research Methods Lab des Leibniz-Instituts für Medienforschung | Hans-Bredow-Instituts und forscht dort zu maschineller Verarbeitung natürlicher Sprache und insbesondere Argument Mining. Er studierte Informatikingenieurwesen an der Technischen Universität Hamburg. Seine Masterarbeit schrieb er an der Language Technology Group der Universität Hamburg in Kooperation mit dem Nachrichtenmagazin DER SPIEGEL. Dort arbeitete er auch während des Studiums als Software Engineer an Open-Source-Software für Investigativjournalismus, in Zusammenarbeit mit der European Investigative Collaboration.

Felix Victor Münch ist Post-Doc im Media Research Methods Lab des Leibniz-Instituts für Medienforschung | Hans-Bredow-Instituts und leitet das (Social) Media Observatory, sein akademisches Interesse gilt der Netzwerkforschung, Big Social Data, Online-Medien und Theorien der Öffentlichkeit. Er war zuvor Research Fellow und Postdoc Researcher im Kolleg Algorithmed Public Spheres.

Gregor Wiedemann ist als Senior Researcher Computational Social Science am Leibniz-Institut für Medienforschung tätig und leitet dort das Media Research Methods Lab. Seine aktuellen Arbeitsschwerpunkte liegen in der Entwicklung von Verfahren und Anwendungen von Natural Language Processing und Text Mining für die empirische Sozial- und Medienforschung. Er studierte Politikwissenschaft und Informatik in Leipzig und Miami, USA. 2016 promovierte er im Fach Informatik in der Abteilung Automatische Sprachverarbeitung der Universität Leipzig. Im Anschluss arbeitete er als Postdoc in der Abteilung Language Technology der Universität Hamburg.

Partnerbeschreibungen

Hochschule für Angewandte Wissenschaften Hamburg

HAW HAMBURG

Nachhaltige Lösungen für die gesellschaftlichen Herausforderungen von Gegenwart und Zukunft entwickeln: Das ist das Ziel der HAW Hamburg – Norddeutschlands führender Hochschule, wenn es um reflektierte Praxis geht. Im Mittelpunkt steht die exzellente Qualität von Studium und Lehre. Zugleich entwickelt die HAW Hamburg ihr Profil als forschende Hochschule weiter. Menschen aus mehr als 100 Nationen gestalten die HAW Hamburg mit. Ihre Vielfalt ist ihre Stärke. Das Department Information und Medienkommunikation erforscht Informationssysteme und -prozesse sowie die Entwicklung des Internets. Studierende lernen, wie die Öffentlichkeit anhand von multimedialen Instrumenten mit Informationen versorgt wird – Kommunikation ist eine interdisziplinäre Wissenschaft und damit eine wichtige Schnittstelle für die Gesellschaft. Im Department Information der HAW Hamburg wird in der Gruppe von Prof. Dr. Stöcker zu den Themen Medienkompetenz, -rezeption und -produktion der digitalen Medien gelehrt und geforscht, u. a. in einem eigens eingerichteten Master-Studiengang „Digitale Kommunikation“ mit Lehrredaktion.

LEIBNIZ-INSTITUT
FÜR MEDIENFORSCHUNG
HANS-BREDOW-INSTITUT

Leibniz-Institut für Medienforschung | Hans-Bredow-Institut, Hamburg

Das Leibniz-Institut für Medienforschung | Hans-Bredow-Institut (HBI) erforscht den Medienwandel und die damit verbundenen strukturellen Veränderungen öffentlicher Kommunikation. Medienübergreifend, interdisziplinär und unabhängig verbindet es Grundlagenwissenschaft und Transferforschung und schafft so problemrelevantes Wissen für Politik, Wirtschaft und Zivilgesellschaft.

Das Forschungsgebiet des Leibniz-Instituts für Medienforschung | Hans-Bredow-Institut (HBI) ist die medienvermittelte öffentliche Kommunikation, unabhängig davon, auf welchen technischen Plattformen die Kommunikation stattfindet. Das Institut erforscht, wie bestimmte Formen der mediengestützten Kommunikation Lebensbereiche wie Politik, Wirtschaft, Kultur, Bildung, Recht, Religion und Familie mitprägen und zu strukturellen Transformationen beitragen. Mit der Problemorientierung der Forschung geht dabei ein besonderes Interesse an den jeweils „neuen" Medien einher, zu deren Verständnis und Gestaltung das Institut beitragen will.

ISD | Institute for Strategic Dialogue

ISD Germany

Seit 2006 ist das Institute for Strategic Dialogue (ISD) führend in der Analyse von koordiniertem Hass, Desinformation und Extremismus in all seinen Formen. Das ISD erforscht Demokratiegefahren nicht nur, sondern erarbeitet auch Lösungen, um diesen entgegenzutreten. Der in Amman, Berlin, London, Paris und Washington DC angesiedelte Think & Do Tank nimmt das gesamte Spektrum digitaler und analoger Entwicklungen in den Blick, mit dem Ziel, Freiheits- und Menschenrechte in den Mittelpunkt zu rücken. Das 2020 gegründete ISD Germany analysiert gesellschaftliche und politische Trends in den deutschsprachigen Ländern und in Europa aus einer globalen Perspektive. In enger Zusammenarbeit mit dem internationalen Team des ISD, bestehend aus weltweit angesehenen Analyst*innen, Politikberater*innen und Bildner*innen arbeitet die ISD Germany gGmbH an strategischen, innovativen und skalierbaren Lösungen gegen Extremismus, Hass und Desinformation. ISD Germany wird von zahlreichen Bundesministerien und namhaften Stiftungen gefördert.

Der vorliegende Bericht ist im Rahmen des vom Bundesministerium für Bildung und Forschung (BMBF) geförderten Projektes „Plattform-übergreifende Identifikation, Überwachung und Modellierung von Verbreitungsmustern von Desinformation (NOTORIOUS)" entstanden. Die inhaltliche Verantwortung liegt ausschließlich bei der Hochschule für Angewandte Wissenschaften Hamburg.

GEFÖRDERT VOM

Bundesministerium für Bildung und Forschung

Zusammenfassung

Dieser Report bietet einen Überblick über die im NOTORIOUS Projekt entwickelten und angewandten Methoden und Strategien zur Erhebung und Analyse von Mehrplattform-Social-Media-Daten mit einem Schwerpunkt auf der Aufdeckung von Desinformationsnarrativen. Zur *Datensammlung* erwiesen sich explorative Ansätze, basierend auf Akteur*innen, Inhalten und extern verlinkten Inhalten als notwendig und wurden entsprechend entwickelt und implementiert. Zur *Datenaufbereitung und -analyse* liegen die Vorteile der von uns entwickelten und vorgestellten Daten-Analyse-Methoden in ihrer Fähigkeit mit vergleichbar geringem Konfigurations-Aufwand, ausreichender Validität und hoher Effizienz große Textmengen zu verarbeiten und komplexe semantische Beziehungen zu erfassen. Hier hat sich die Verwendung semantischer Einbettungen von Texten als sehr effektiv herausgestellt, um Themen und Themenströmungen zu identifizieren. Auch die Erstellung von Ähnlichkeitsnetzwerken aus semantischen Einbettungen erwies sich als leistungsfähige Methode zur Visualisierung der Beziehung zwischen Themen und Akteur*innen. Semantische Ähnlichkeitsnetzwerke ermöglichten darüber hinaus eine Komplexitätsreduktion für Analyst*innen durch die Identifizierung von Clustern, die wir über einen semantischen Abgleich mit Fact-Check-Datenbanken mit bestimmten Desinformationsnarrativen in Verbindung bringen konnten. Mögliche zukünftige Entwicklungen basierend auf unseren Erkenntnissen sind die Integration multimodaler Ansätze und die Erschließung weiterer Datenquellen im Sinne einer kontinuierlichen Verbesserung der Erkennung und Bekämpfung von Desinformation.

1. Einleitung

Social-Media-Plattformen sind als Infrastruktur der (proprietären) digitalen Öffentlichkeit in aller Regel nicht isoliert, sondern sind über die Nutzendenschaft und auf der inhaltlichen und technischen Ebenen miteinander verwoben. Dazu tragen unter anderem die verschiedenen Ausrichtungen der Plattformen bei: X.com ist als Fotogalerie weniger geeignet, Instagram für Text nur über Umwege, Videos werden in der Regel nicht bei LinkedIn gehostet, sondern eher bei YouTube. Dadurch spannt sich ein plattformübergreifender digitaler Raum auf, in dem sich öffentliche Diskurse auf mehreren Ebenen abspielen.

Für die inhaltliche Analyse von (politischen) Diskursen ergeben sich daraus allerdings technische Herausforderungen, da die Erhebung und Aufbereitung digitaler Daten plattformspezifisch realisiert werden muss. Für die Crossplattform-Analyse müssen diese Daten dann vergleichbar gemacht werden, was durch die verschiedenen Affordanzen der Plattformen zu Eingriffen in die Inhalte führen kann. Dies kann wiederum für eine Analyse von Nachteil sein. Dementsprechend wohlüberlegt müssen die Entscheidungen zur Vereinheitlichung von Inhalten sein.

Das Ziel dieses Projektes war und ist es, diese Verbreitungsspuren für spezifische Inhalte – insbesondere Desinformation – in plattformübergreifenden Diskursräumen nachvollziehbar zu machen. Dafür wurde eine neue Analysemethode entwickelt, die die obigen Herausforderungen in Bezug auf Auswahl, Erhebung, Angleichung, Aufbereitung und Analyse von Daten aus mehreren Social-Media-Plattformen adressiert. Wir stellen die genutzten Strategien und Ansätze zur Auswahl von Akteur*innen, Inhalten und Plattformen vor und beleuchten den von uns verfolgten Ansatz. Veranschaulichend wird eine Fallstudie nachgezeichnet, die zu Anfang des Projektes als Pilotstudie zur Erprobung von Methoden gedient hat und seitdem kontinuierlich weiterentwickelt wurde (Bundtzen et al., 2022).

2. Datengewinnung und Herausforderungen

Eine der grundlegenden Entscheidungen, die für die Erhebung von Daten in mehreren Plattformen getroffen werden müssen, ist, auf welcher Grundlage die einzelnen Datenpunkte der Plattformen zusammengeführt werden. Zwei Strategien haben sich etabliert: Der akteursbasierte und der inhaltsbasierte Ansatz. Für ersteren werden im Vorhinein relevante Akteure definiert, z.B. Mitglieder von politischen Parteien, die auf verschiedenen Plattformen aktiv sind. Die erhobenen Spurendaten lassen sich dann zumindest teilweise über diese bekannten Akteur*innen verknüpfen. Dieser Ansatz ist vor allem dann praktikabel, wenn die Gruppen klar eingrenzbar und überschaubar sind, z.B. bei Mitgliedern einer parlamentarischen Fraktion im Gegensatz zu Unterstützern einer politischen Bewegung.

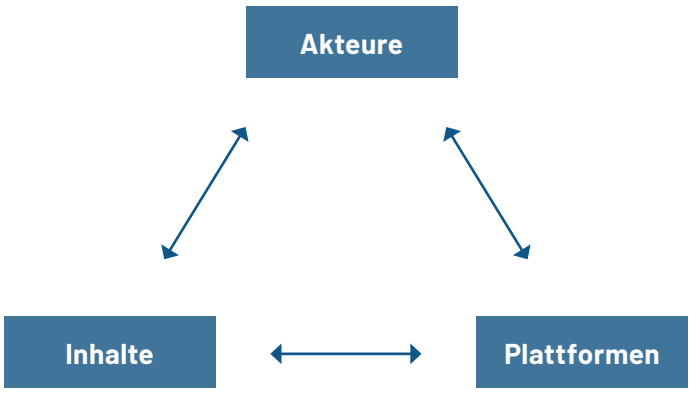

Abbildung 1: Schematische Darstellung der Übergänge zwischen Ansätzen der Mehrplattformdatenerhebung.

Der inhaltsbasierte Ansatz trifft keine Auswahl der Akteure, für die Daten erhoben werden sollen, sondern schränkt die Datenerhebung durch das Abfragen von inhaltlichen Markern, i.d.R. Stichworten, innerhalb der Spurendaten ein. So könnte es für ein Forschungsvorhaben interessant sein, Posts zu erfassen, die sich auf ein bestimmtes Ereignis beziehen: Nehmen wir als Beispiel die Besetzung des Hambacher Forsts, der in diesem Kontext auch als „Hambi" bezeichnet wurde, was als ein solcher Marker dienen kann. Für die stichwortbasierte Datenerhebung können in der Regel die Suchfunktionen der Zielplattformen – gegebenenfalls mit einer zeitlichen Einschränkung - genutzt werden. Durch die Kombination von Suchergebnissen auf mehreren Zielplattformen entsteht dann eine plattformübergreifende Datengrundlage, deren Auswertung eine strukturelle Vereinheitlichung erfordert (s.u.).

Herausforderungen bei der Datenerhebung

Bei der Datengewinnung wurden mehrere im Folgenden beschriebene Herausforderungen adressiert.

Unbekannte Datengrundlage

Zunächst muss die Einschränkung gemacht werden, dass Akteure und Inhalte öffentlicher Kommunikation selten klar abgrenzbar sind: Akteursgruppen sind im seltensten Falle trennscharf (wie z.B. Journalist*innen und Blogger*innen) und auch bei Stichwörtern gibt es Unschärfen und tote Winkel, die ausgeleuchtet werden müssen. Gerade Gruppen, die fernab des Mainstreams agieren, nutzen eigenes Vokabular, um in der jeweiligen Gruppe über Sachverhalte zu sprechen. Dies wird erst durch eine intensive Auseinandersetzung mit erhobenen Daten sichtbar. Impfgegnerkreise brachten z.B. das Wort *schlumpfen* als Synonym für impfen hervor, um abfällig über diesen Vorgang kommunizieren zu können. Gerade bei explorativen Studien, bei denen die zu erforschende Materie zu großen Teilen noch nicht erfassbar ist, kommt es unter Umständen zu Forschungsfragen, bei denen entweder die zu beleuchtenden Akteure oder auch von Akteuren gesetzte Themen (noch) nicht bekannt sind. Erhebungen wachsen dadurch iterativ und sind nie vollständig, insbesondere wenn mehrere Plattformen miteinbezogen werden sollen.

Im Laufe des Projektes hat es sich als robust herausgestellt, Diversifikation zu betreiben und eine Methodenkombination anzuwenden, um dieses Problem zu mitigieren: So ist es möglich, aus den Inhaltsdaten einer akteursbasierten Datensammlung die relevanten Themen, die diese Akteure in ihrem Diskurs behandelt haben, zu extrahieren. Genau so können auch für beobachtete Themen relevanten Akteure bestimmt werden. Eine Variation dieses explorativen Vorgehens ist es, auch über URLs verknüpfte Inhalte mitzuanalysieren, z.B. auf dem Domain-Level. So können auch andere

relevante Plattformen und Akteure entdeckt werden. Auf der technischen Ebene besitzt jeder Post und Artikel, jedes Bild und Video auf den Plattformen einen sogenannten Universal Ressource Locator (kurz: URL), umgangssprachlich einen Link, mit dem die Inhalte abgerufen werden können. Durch diese technische Ebene lassen sich auch ohne plattform-interne Metadaten Verbreitungsspuren von Inhalten nachzeichnen. URLs erlauben eine erste Annäherung an diese Verbreitungsspuren, aber haben Limitationen: Texte und Bilder sind schnell kopiert und an anderer Stelle wieder hochgeladen. Akteure und deren Accounts sind mitunter sehr flüchtig und selbst große Plattformen können überraschende Strategieänderungen vornehmen. Eine Iteration aller drei Ansätze erlaubt jedoch unseres Erachtens ein ausreichend umfassendes Mapping der betrachteten, digitalen Öffentlichkeit.

Unsichere und unterschiedliche Standards für Datenzugang

Die Erhebung von Social-Media-Daten ist auf die Verfügbarkeit von Datenzugriffen durch Plattformbetreibende angewiesen. Dieser Zugang steht allerdings vermehrt im Konflikt mit den Interessen der Betreiber, die als Reaktion Zugänge erschweren oder unterbinden. Als Beispiel kann hier das *Ad Observatory* der New York University angeführt werden, deren Zugängen und sogar private Accounts der Forschenden durch Meta Inc. gesperrt wurden, weil die Forscher*innen aus Metas Sicht gegen Richtlinien der Plattform verstoßen hatte (Clark, 2021; DeLong, 2021).

Die in Deutschland gebräuchlichen Social-Media-Plattformen unterscheiden sich darüber hinaus nicht nur in der Art und Weise, für welche Art von Inhalten diese gedacht sind (beispielsweise Instagram als Foto- und Videoplattform, Twitter/X.com als Kurznachrichtendient), sondern auch darin, wie Forscher*innen Daten aus diesen Plattformen für Vorhaben gewinnen können. Twitter war hier lange, durch offene und gut dokumentierte Datenzugänge mittels einer Programmierschnittstelle (API), die meistbeforschte Plattform. Auch die einfache und klare Datenstruktur Twitters erwies sich als Vorteil. Meta hatte als Betreiber von Instagram und Facebook mit Crowdtangle bisher ebenfalls eine nutzerfreundliche API zur Verfügung gestellt. Inhalte dieser großen, kommerziellen Plattformen ließen sich dementsprechend vergleichsweise leicht erheben. Beide Schnittstellen wurden aber mittlerweile durch derzeit weniger zugängliche Verfahren ersetzt. Eine große Herausforderung der im Projekt angewandten Methoden wird daher in der Zukunft die Datengewinnung für weitere Anwendungen sein, da die Plattformen der bisherigen Untersuchungen den Datenzugang nach und nach erschweren oder restriktiv monetarisieren. So wurde die Academic-API von X/Twitter mittlerweile eingestellt und durch eine mit in öffentlicher Forschung üblichen Budgets nicht bezahlbare Enterprise-API ersetzt. Meta als Betreiber von Crowdtangle hat hat die Plattform Mitte August 2024 eingestellt.

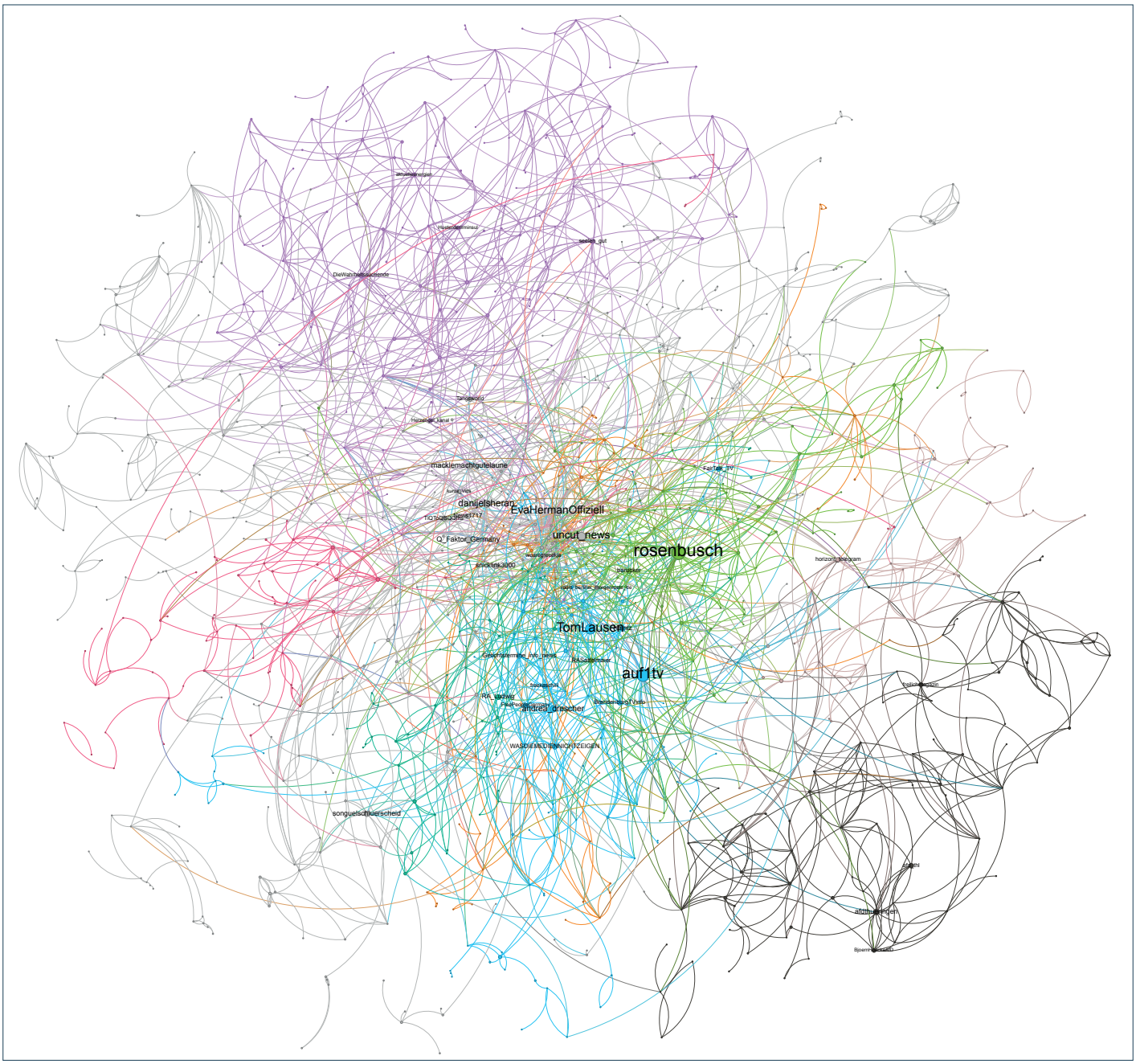

Abbildung 2: Weiterleitungen zwischen Telegram Kanälen als Netzwerk visualisiert. Jeder Knoten repräsentiert einen Kanal, die Dicke der Verbindungen gibt die Anzahl von Weiterleitungen zwischen den Kanälen wieder.

Während beide Plattformen neue Zugänge für die Forschung geschaffen haben oder schaffen, wird der Zugang erfahrungsgemäß immer restriktiver. Ob der Data Services Act (DSA) der Europäischen Union hier Abhilfe schafft, bleibt abzuwarten. Neuere plattformübergreifende Datenerhebungen sind zurzeit nur mit deutlich höherem Aufwand und dennoch nur geringerem Umfang realisierbar.

Sonderfall Telegram

In den Jahren der Covid-19-Pandemie hat Telegram in Deutschland einen starken Zuwachs an Nutzenden erfahren. Durch die meist ausbleibende Inhaltsmoderation auf der Plattform und die Möglichkeit, nicht nur Chats zwischen Einzelpersonen, sondern auch sog. *One-to-many* oder *broadcast* Kanäle anzulegen, ist Telegram eine Anlaufstelle für verschwörungsideologische, rechtsextreme Akteure und politische Splittergruppen geworden (Wiedemann, Schmidt, et al., 2023). Für die Datenerhebung im NOTORIOUS Projekt war Telegram also von großem Interesse, allerdings musste eine eigene technische Lösung zur Datenerhebung entwickelt werden, da sich die Funktionsweise von Telegram von den anderen Plattformen technisch unterscheidet. Zwar bietet Telegram auch offene Zugänge mittels einer Programmierschnittstelle – Nachrichten von öffentlichen Kanälen sind bspw. sogar frei im Internet (also ohne spezielle App) abrufbar – doch erlaubt die Plattform keine globale Stichwort-Suche in allen Kanälen. Das heißt, dass thematisch fokussierte Datensammlungen hier ohne weiteres nicht möglich sind.

In der Literatur werden verschiedene Möglichkeiten für das Sampling von Akteuren für die Erhebung von Spurendaten vorgeschlagen, um diesen Umstand zu umgehen, wobei hier vor allem die Erhebung für einzelne Studien im Fokus stehen (Jost et al., 2023). In diesem Projekt wurde für die Urbarmachung von Telegram für die Forschung ein Indexdatensatz für die Volltextsuche mittels eigener Software angelegt. Dafür wurde ein spezialisierter Scraper entwickelt (Kessling & Münch, 2022/2023). Diese Software sammelt gezielt Weiterleitungs- und Reputationsdaten von Posts und Kanälen, welche Aufschlüsse über die Makrostruktur von Aufmerksamkeit auf der Plattform erlauben. Eine weiterer Scraper erlaubt die groß-skalierte Sammlung von Inhaltsdaten ausgewählter Kanäle (Kessling & Münch, 2023), die auch kontinuierlich aktuell gehalten werden können. Mit dieser Kombination wurde zuerst ausgehend von den 1.000 reichweitenstärksten Telegram-Kanälen zurückverfolgt, welche Nachrichten anderer Kanäle dort geteilt wurden. Über die Abfragen über das öffentliche Web-Interface der Plattform, lassen sich auf diese Weise innerhalb kürzester Zeit die groben Zusammenhänge der Plattform erfassen und kartieren. Abbildung 2 zeigt ein solches Weiterleitungsnetzwerk, das im Oktober 2022 erstellt wurde. Mittels der generierten Kanallisten wurden für alle erfassten Kanäle alle Nachrichten ab dem 1.1.2019 rückwirkend erhoben. Dieser Datenbestand fasst im Moment 143.403.843 Postings von 15.396 internationalen Kanälen, davon ungefähr 2.100 deutschsprachige Kanäle.[1]

[1] Da Telegram selbst keine automatische Detektion der Sprache einer Nachricht oder eines Kanals durchführt, wie es andere Social Media Plattformen tun, wurde die Sprachendetektion in Nachgang der Datenerhebung vorgenommen. Auch einer Stichprobe von zehn Nachrichten pro Kanal wurde mit einem spezialisierten Sprachmodell und der fasttext-Programmbibliothek die wahrscheinliche Sprache der Texte erfasst (Grave et al., 2018).

3. Datenaufbereitung und -analyse

Eine Besonderheit der Mehrplattformanalyse sind die unterschiedlichen Affordanzen, die die Plattformen den Nutzenden zur Verfügung stellen. Ein zentraler Baustein auf allen Plattformen ist die Verwendung von Texten entweder als primäres Element, wie auf Twitter oder Facebook, oder zu mindestens als sekundäres Element, wie etwa eine Bildunterschrift auf Instagram. Diese Postingtexte werden im Projekt als primäre Analyseeinheit genutzt.

3.1. Datenvorbereitung/Preprocessing

Eine weitere Herausforderung für die Textverarbeitung von Posts verschiedener Plattformen sind die unterschiedlichen Textlängen typischer Posts (auch teilweise eingeschränkt durch Maximallängen, wie bei X.com). Um zu gewährleisten, dass die primäre Analyseeinheit vergleichbar bleibt, mussten längere Posts, wie sie für Telegram und Facebook üblich sind, in Absätze zerteilt werden, die dann weiterverarbeitet werden können. Die Arbeit mit kompletten Posts als atomare Texteinheit ist insbesondere mit Telegram-Daten insofern schwierig, dass häufig viele verschiedene Themen in einem Post behandelt werden und somit die Auflösung als einzelne Posts die Vergleichbarkeit mit anderen Plattformen erschwert.

3.2 Methoden für die Cross-Plattform Analyse: Text Embeddings und Netzwerkanalyse

Um die Komplexität einer großen Anzahl an Einzeltexten auf inhaltlich vergleichbare und durch Analyst*innen überschaubare Kategorien zu reduzieren, werden oft sog. Clustering- oder Topic-Modeling-Verfahren eingesetzt. Für das Clustering von Posts verschiedener Plattformen wurden von uns zwei methodische Paradigmen miteinander verknüpft: Das Natural Language Processing (NLP) und die Netzwerkanalyse.

Moderne Verfahren des NLP machen es möglich, ohne manuell kodierte Daten die Ähnlichkeit zwischen zwei Texten zu errechnen, wenn Sprachmodelle (Language Models, LMs) mit großen Datenmengen für diese Aufgabe vortrainiert wurden. Die berechneten Ergebnisse dieser Modelle, *Embeddings* genannte numerische Repräsentationen der Eingabetexte, geben deren Lage in einem angenommen, semantischen Vektorraum wieder, wobei Texte ähnlichen Inhalts dann auch in diesem hochdimensionalen Raum nahe beieinander liegen. Die Ähnlichkeit zwischen zwei Texten wird dann auf Basis der sog. Kosinus-Ähnlichkeit auf einer Skala zwischen 0 und 1 angegeben (Reimers & Gurevych, 2019). Für unsere Methode sind prinzipiell alle Sprachmodelle, die solche Texteinbettungen erstellen, verwendbar, sodass auch spätere Verbesserungen in diesem Bereich genutzt werden können. Auch eine Verwendung mit multimodalen Inhalten (also Bild, Audio, oder Video) ist mit einer vorherigen Umwandlung in Texte möglich.[2] Wir verwenden seit diesem Jahr ein kleines, mehrsprachiges Modell, das eine gute Performance aufweist (Wang et al., 2024).

[2] Vgl. Abschnitt **6. Ausblick**.

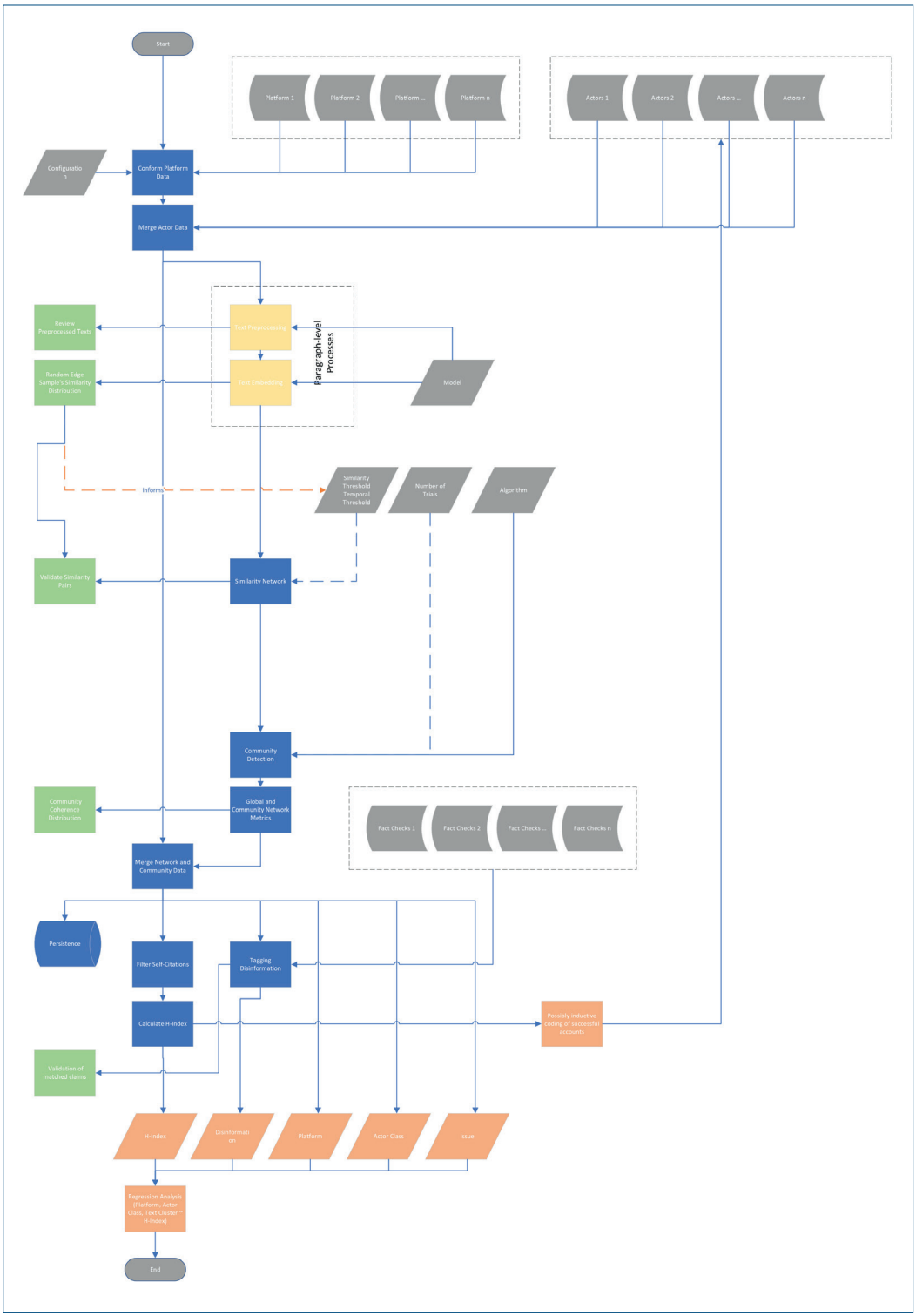

Abbildung 4: Beispielablauf einer Mehrplattformdatenanalyse mit semantischem Clustering, Desinformations- und Akteursabgleich.

Das zweite von uns genutzte methodische Paradigma ist die Netzwerkanalyse. Netzwerke sind besonders gut geeignet, Komplexität zum Zweck der Übersicht auf der Makro-Ebene zu reduzieren und gleichzeitig die Details der Einzelbeobachtungen für die qualitative Analyse zu erhalten (Hidalgo, 2016; Münch, 2019). Mit der semantischen Ähnlichkeit zwischen Posts erhalten wir eine gewichtete Relation, die genutzt werden kann, um ein Netzwerk zwischen Posts aufzuspannen. Eine Mindestähnlichkeitsschwellenwert, für den zwei Analyseeinheiten als ähnlich genug für eine Verbindung derselben gelten, kann als Hyperparameter der Methode gewählt werden. Das durch diese Verbindungen aufgespannte Netzwerk kann dann mit verschiedenen etablierten Methoden der Netzwerkanalyse interpretiert werden. Beispielsweise können inhaltlich zusammenhängende Posts mittels *community detection* Algorithmen wie etwa *infomap* (Edler et al., 2017) gefunden werden, um so thematische Zusammenhänge über mehrere Social-Media-Plattformen aufzudecken und über die Zeit nachzuverfolgen.

Vergleiche mit der mittlerweile weit verbreiteten Topic-Modellierungs-Methode BERTopic (Grootendorst, 2022) zeigen einen Vorteil des netzwerkbasierten Clusterings in der Performance und der Handhabbarkeit großer Datenmengen. Um BERTopic auf einem Datensatz von mehreren Millionen Posts anzuwenden, bedarf es selbst mit der Rechenbeschleunigung durch die Verwendung von Beschleuniger-Hardware, wie etwa Grafikkarten (Graphics Processing Units, GPUs), mehr als 12 Stunden Rechenzeit. Dies ist insofern problematisch, dass BERTopic durch seinen sequenziellen Aufbau aus verschiedenen statistischen Methoden mit unabhängig einzustellenden Parametern häufig erst durch die intensive Auseinandersetzung mit der Methode und wiederholten Berechnungen brauchbare Ergebnisse liefert. Diese Notwendigkeit besteht bei einer direkten Umwandlung in ein Netzwerk und angeschlossen Analysemethoden in der Regel in diesem Maße nicht.

Ein weiterer Vorteil der Nutzung von semantischer Einbettung mittels eines Netzwerks ist die Anschlussfähigkeit für weitergehende Analyseschritte. So können netzwerkanalytische Methoden angewandt werden, um besonders repräsentative Posts für bestimmte Themen zu finden. Ebenso lassen sich klassische Methoden der automatisierten Sprachverarbeitung, wie Schlagwortextraktion mittels der $\chi 2$ Statistik anwenden.

3.3 Verknüpfungen zu externen Datenquellen

Es lässt sich auch recht einfach eine *semantische Suchmaschine* implementieren, wenn von Analyst*innen formulierte, prototypische Texte als Suchanfragen ebenfalls in den Vektorraum eingebettet und für die Ähnlichkeitssuche genutzt werden. Hier sind auch automatisierte Abfragen für externe Daten möglich. Im Folgenden werden zwei externe Datenquellen vorgestellt, mit denen wir die Analyse von Projektdaten weiter vertiefen konnten.

Desinformationsbestimmung
Die semantische Suche bietet einen guten Startpunkt für eine explorative, qualitative Analyse in umfangreichen Korpora von Social-Media-Posts. Im Projekt wurde eine Datenbank für die Verknüpfung von Personen und eine Datenbank mit professionell durchgeführten Fact-Checks verwendet, um den Einfluss prominenter Accounts bei der Verbreitung von Desinformation zu untersuchen. Um zu bestimmen, ob Posts Aussagen enthalten, die als Desinformation verdächtig markiert werden sollten, werden die einzelnen Textanalyseeinheiten eines Datensatzes mit einer Datenbank von Faktenchecks (z.B. der Google FactCheck Claim Search API: https://toolbox.google.com/factcheck/apis) verglichen. Die Einbettungen für die Kernaussagen der Faktenchecks werden dafür ebenfalls berechnet. Danach kann die Ähnlichkeit zwischen den Inhalten des Faktenchecks und des Posts wie oben beschrieben verglichen werden.

Influencer-Bestimmung auf mehreren Plattformen
Das Hans-Bredow-Institut in Hamburg pflegt eine langfristig angelegte Datenbank öffentlicher Sprecher*innen (DBöS), in der Metainformationen zu den Social-

Media-Kanälen typologisierter öffentlicher Akteure abgelegt sind (Schmidt et al., 2023). Eine Verknüpfung dieser Informationen mit Plattformdaten macht es möglich, öffentliche Sprecher wie Nachrichtenmedien, Parteifunktionäre oder Journalisten plattformübergreifend zu beobachten.

Im Projekt wurde dieser Datenbestand um weitere Datenbanken ergänzt: Um auch Wirtschafts- und Verbandsinteressen beobachtbar zu machen, wurden das Lobbyregister des Bundestages (https://www.lobbyregister.bundestag.de/startseite, Stand 2023) ausgewertet und die Social-Media-Accounts der dort registrierten Personen, Firmen und Verbände zusammengeführt. Ebenfalls wurde eine Datenbank von alternativen Medien angelegt.

Ergänzend zu diesen Beispielen einer Eingrenzung von Influencer-Accounts durch manuelle Erstellung von Datenbanken und externen Datenquellen besteht die Möglichkeit, diejenigen Accounts, die innerhalb eines Datensatzes Einfluss beweisen, datengetrieben zu erfassen. Dafür lassen sich verschiedene Metriken formulieren: Accounts, die besonders früh auf Themen aufgesprungen sind, die viel zitiert wurden (etwa durch die Detektion in einem Ähnlichkeitsnetzwerk) oder die eine hohe Anzahl an Erwähnungen innerhalb eines Datensatzes aufweisen. Auch die unter Umständen abweichenden Nutzernamen in verschiedenen Plattformen lassen sich durch Messung von Ähnlichkeiten zwischen Nutzernamen aufdecken und so Verknüpfungen über Plattformen hinweg erstellen.

4. Infrastruktur

Die im ersten Bericht dieses Projekts (Bundtzen et al., 2022) skizzierte technische Architektur wurde sukzessive skaliert und für die Datenerhebung in mehreren Plattformen ausgelegt. Diese wurde in die Forschungsinfrastruktur des Social Media Observatory des Forschungsinstituts gesellschaftlicher Zusammenhalt am Leibniz-Institut für Medienforschung – schematisch dargestellt in Abbildung 5 – integriert und dokumentiert (Wiedemann, Münch, et al., 2023). Für relevante Themen, wie etwa die Invasion Russlands in der Ukraine, wurden auch große Datensätze als Forschungsinfrastruktur nicht nur im Projekt genutzt, sondern auch der Allgemeinheit zur Verfügung gestellt (Münch & Kessling, 2022).

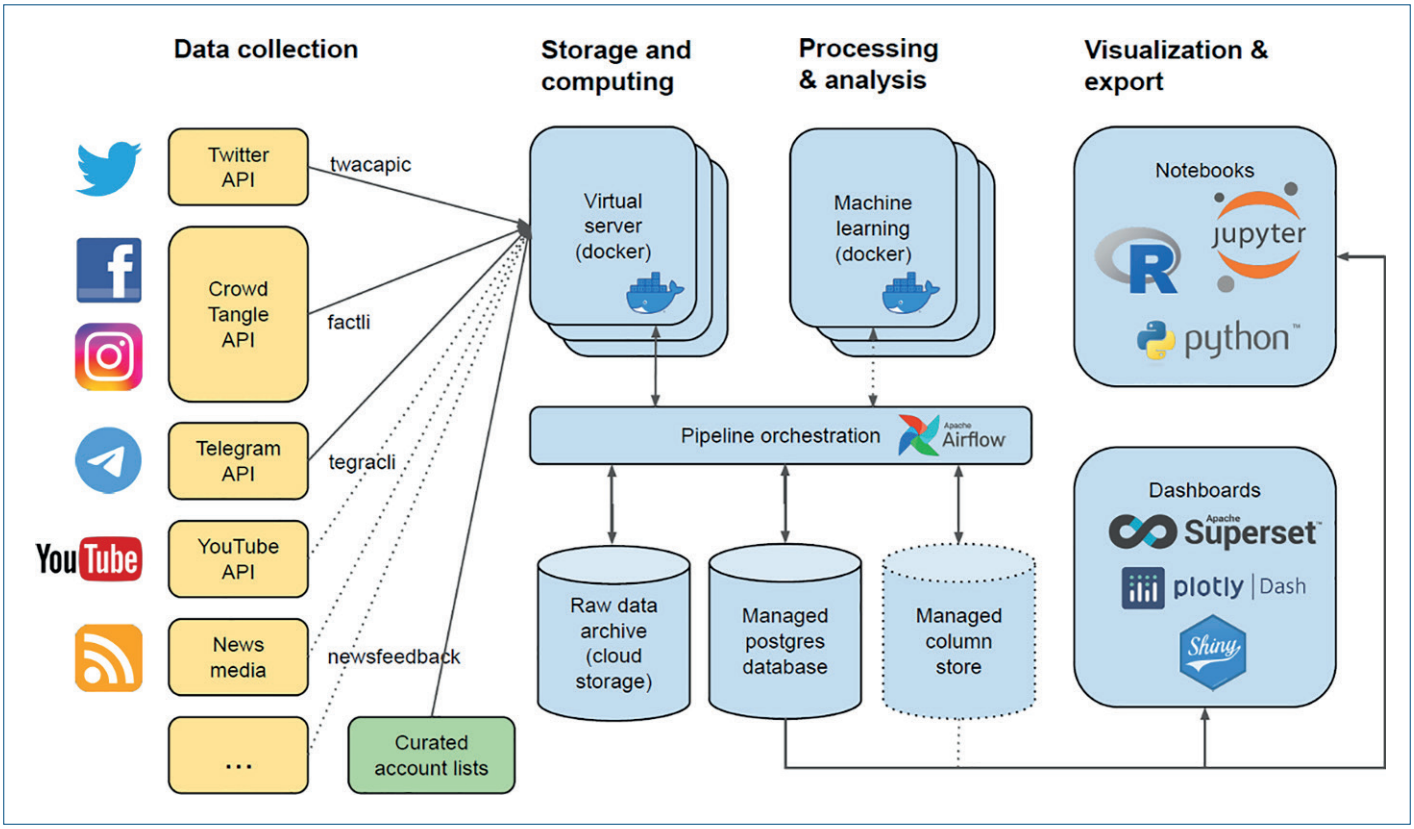

Abbildung 5: Schematische Darstellung des Dateninfrastrukturkonzepts des Social Media Observatory des Forschungsinstituts gesellschaftlicher Zusammenhalt am Leibniz-Institut für Medienforschung (Wiedemann, Münch, et al., 2023)

Dockered Pipelines und reproduzierbare Workflows

Um die Reproduzierbarkeit von automatisierten Methoden zu gewährleisten, ist es ratsam, eine Container-Umgebung – in unserem Fall eine Docker-Umgebung – zu erstellen, in der alle notwendigen Pakete für die Datenauswertung kontrolliert installiert werden. Docker ist ein Tool, um Laufzeitumgebungen zu definieren und automatisiert virtuelle Maschinen mit vordefinierten Einstellungen zu starten. Einzelne Arbeitsschritte sind in Jupyter-Notebooks festgehalten, die sukzessiv aufeinander aufbauend aus zuvor erhobenen Rohdaten eine Clusteranalyse durchführen.

Analyseplattform für schnellen Datenzugriff

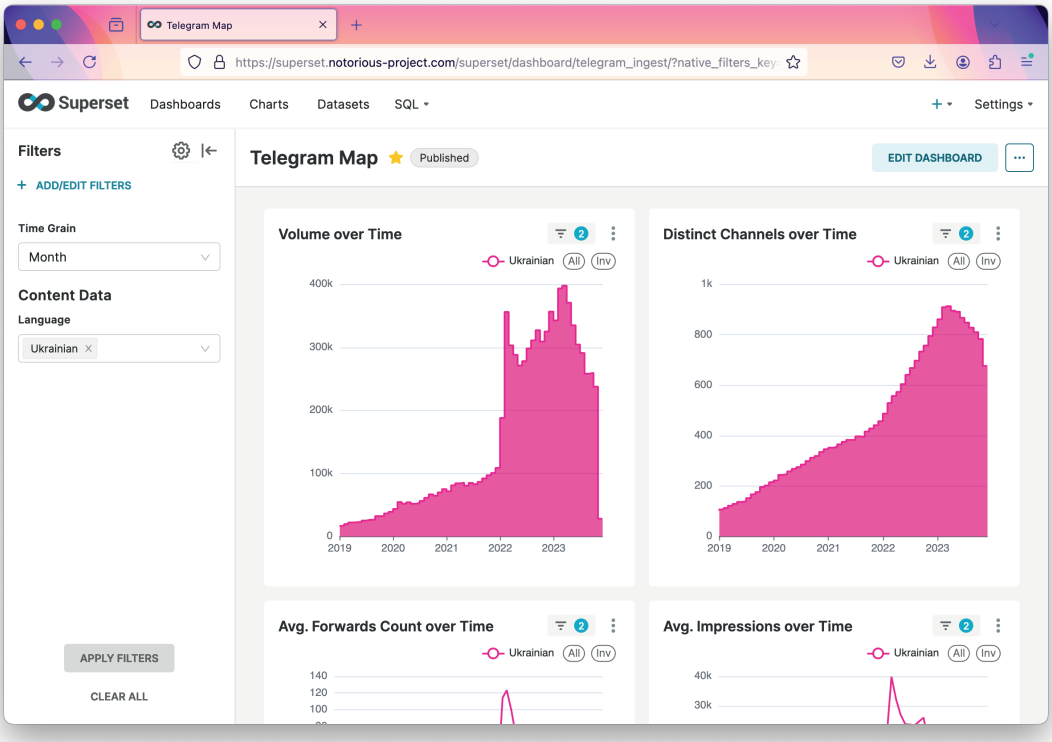

Abbildung 6: Beispiel eines Superset -Dashboards. Hier gezeigt werden Parameter des oben genannten Telegram-Datensatzes mit der Anzahl der monatlich gepposteten Nachrichten, der Anzahl der monatlich aktiven Kanäle, sowie der durchschnittlichen Weiterleitungs- und Anzeigeanzahlen pro Monat. Links finden sich die Filtereinstellungen, hier können weiter Sprachen zu oder abgewählt werden.

Um ein fortlaufendes Monitoring und interaktive Datenanalysen direkt auf den erhobenen Daten zu implementieren, bieten sich web-basierte Datenanalyseplattformen an. Im Projekt wurde nach einer Erkundung des Marktes die Free and Open Source (FOSS) Lösung Superset, die von der Apache Foundation entwickelt wird, genutzt. Diese Plattform erlaubt den nutzerfreundlichen, zentralen Zugriff auf in verschiedenen Datenbanksystemen hinterlegte Datenbestände und dadurch die interaktive Analyse mittels SQL-Abfragen, zum Beispiel um Arbeitsdatenstände zu generieren und im Projekt zu teilen. Dabei müssen Analyst*innen nicht unbedingt in der Abfragesprache geübt sein, sondern können auch über ein grafisches System Abfragen generieren und die entstandene Aggregation als Plots visualisieren. Mehrerer solcher Visualisierungen lassen sich als *Dashboards* zusammenstellen und über Filtereinstellungen parametrisieren. Eine lokale Installation dieses Systems ist nicht notwendig, da es zentral für eine Forschungsgruppe oder eine Institution gehostet und für Forschende zur Verfügung gestellt wird.

5. Anwendungsbeispiel: mRNA-Rapper

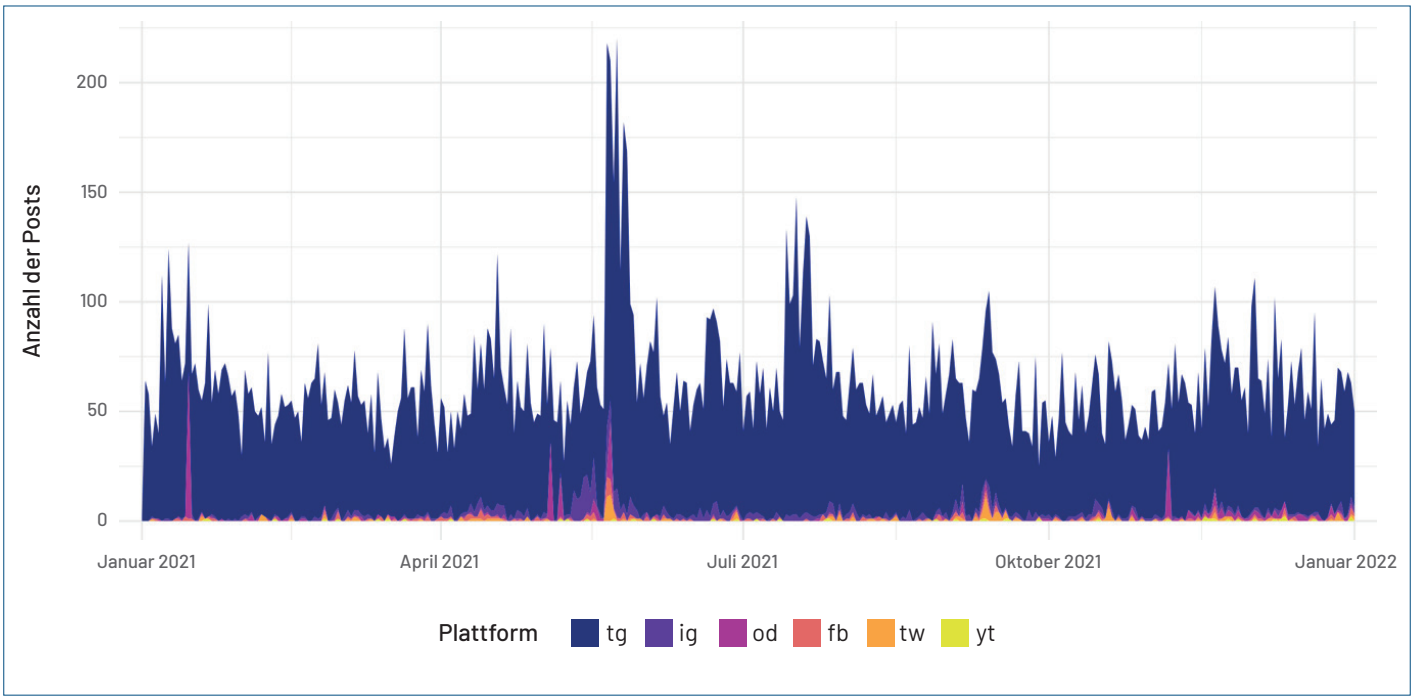

Abbildung 7: Zeitleiste mit der Anzahl der täglich geposteten Nachrichten auf sechs Plattformen des Rapbellion-Kollektivs.

Für den Anwendungsfall einer Studie, die zu Beginn des Projektes gestartet wurde, wurden für ein Kollektiv von Musikern, die Protestsongs gegen Pandemie-Maßnahmen veröffentlichten, und deren Umfeld die geposteten Nachrichten aus vier Plattformen erhoben. Für Facebook und Instagram ließen sich die Nachrichten über CrowdTangle erheben, für Twitter/X.com wurde die von der Plattform bereit gestellte Research API genutzt. Für Telegram wurden die Daten ebenfalls über die plattformeigene API erhoben. Um diesen Datensatz zu komplettieren, wurde geprüft, welche anderen Plattformen von dem Kollektiv genutzt wurden. Nach einer Auswertung der in den Posts eingebetteten Links, wurden YouTube und Odyssee als ebenfalls relevante Plattformen identifiziert und in die Datenerhebung mit einbezogen. Für die Datenerhebung auf YouTube konnte ebenfalls eine API genutzt werden, während die Inhalte von Odysee über einen automatisierten Browser gescraped wurden. Darauffolgend wurden die skizzierten Methoden zur Datenaufbereitung und -analyse angewendet. Hier ergaben sich aus den über 20.000 gesammelten Posts, knapp 9.000, die nach der Textbereinigung noch mindestens fünf Worte umfassen und in die Analyse miteinbezogen werden. Zwischen diesen Posts wurden 166.000 Ähnlichkeiten zwischen Textteilen dieser Posts identifiziert (siehe Abbildung 8). Mit infomap wurden mittels dieser Verbindungen Posts zu Clustern zusammengefasst. Nach der Einteilung in Cluster, lassen sich für diese besonders relevante Schlagworte und durch ein Sprachmodell erstellte Zusammenfassungen der Clusterinhalte bestimmen (Tabelle 1).

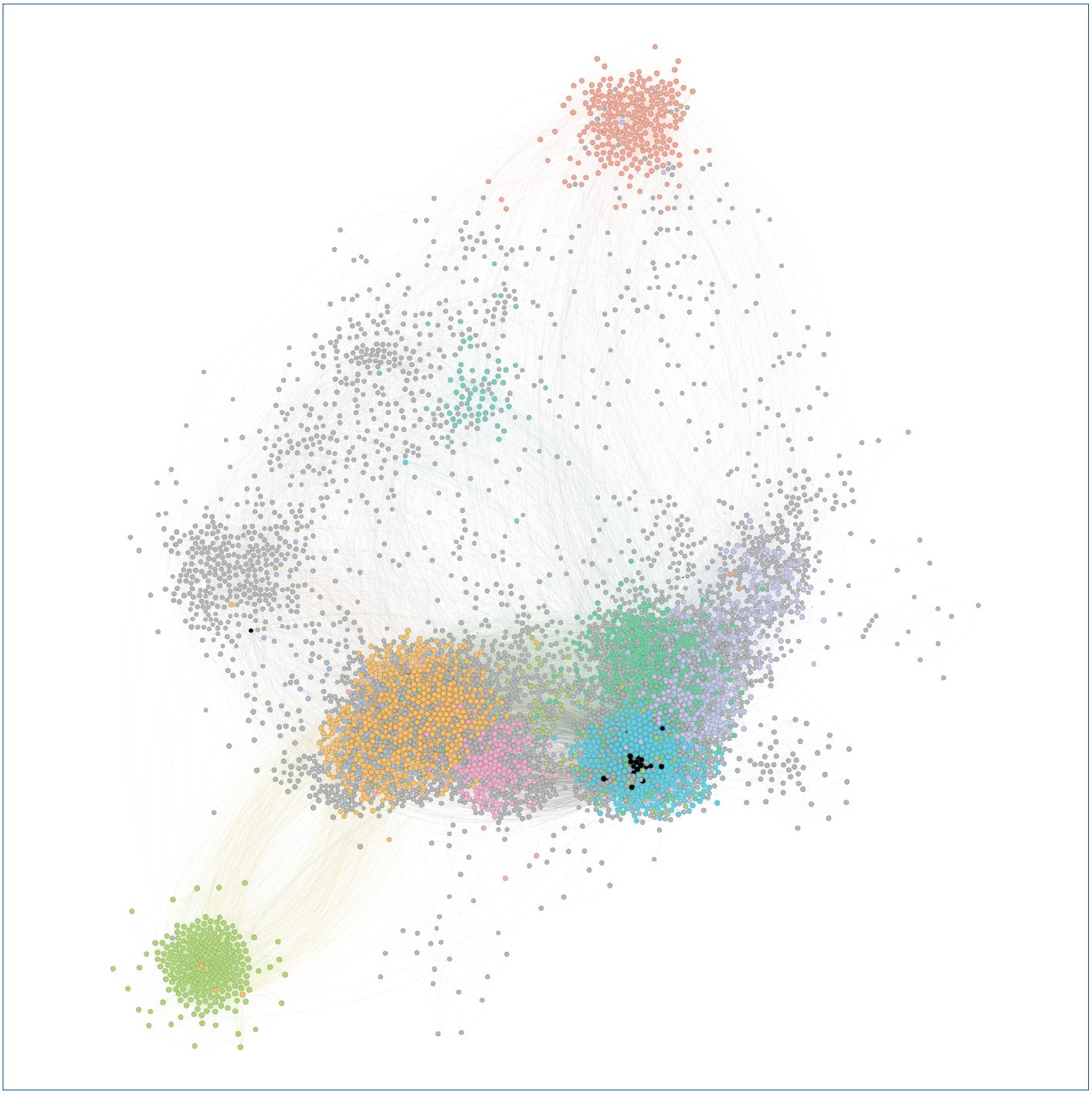

Abbildung 8: Visualisierung der Ähnlichkeitsbeziehungen zwischen einzelnen Postings. Die Farben markieren Clusterzugehörigkeiten: Grau – Sonstige, Pink – 2:1, Grün – 1:3, Blau – 1:2, Dunkelgrau – 1:1, Orange – 1:4, Rot – 2:2, Türkis – 2:3, Flieder – 3:2.

Tabelle 1: Übersicht von Schlagworten, die nach dem Clustering der Inhalte aus den Texten, extrahiert wurden. Ausgewählt wurden die zehn größten Cluster. Mit einem Sprachmodell Llama3 wurden jeweils zehn relevantesten Posts der jeweiligen Cluster zusammengefasst[3].

Cluster	Kohärenz	Größe	Schlagworte	Zusammenfassung
1:1	24,3 %	1.242	*lapaz, links, stark, support, danke*	Verschiedene Kanäle motivieren zur Beteiligung an Diskussionen, zum Nachdenken und zur Unterstützung durch Abonnements und Spenden.
2:3	17,6 %	767	*denkt, ganze, faktenfriedenfreiheit, letzte, dran*	Die Texte befassen sich mit Demonstrationen, politischem Aktivismus, Corona-Maulkorb und Verschwörungstheorien.
1:3	26,0 %	724	*geimpften, ärzte, impfung, studie, impfen*	Die Texte äußern, dass viele Impferfahrungen negativ und mit hohen Sterberaten und Nebenwirkungen verbunden sind. Ethisch gesehen sei das Impfen falsch, da es nebenwirkungsreich und unnötig ist.
2:2	20,0 %	679	*unterstützung, paypal, info, unterstützen, thomas*	Die Texte behandeln unterschiedliche Themen: Schulleitung in Zeiten von Corona, Politiker Möllemanns Tod, Schulpflicht ohne Impfung, friedlicher Widerstand, Ethikratsvorsitzende über Impfangebote und Staatsfernsehen sowie diverse Aufrufe zu Sendungen, Videos und Aktionen.
1:4	13,7 %	629	*ungeimpfte, covid, impfung, geimpfte, geimpften*	Die Texte äußern skeptische bis ablehnende Positionen zur COVID-19-Impfung. Sie werden darin zusammengefasst, dass die Impfung keine Immunität bringt und unnötig ist.
2:1	29,8 %	467	*amazon, mach, zensur, gelöscht, music*	Mehrere Musiker rufen zum Streaming und Teilen eines Lieds auf, da sie von verschiedenen Online-Plattformen gelöscht wurden. Sie wollen damit in die Charts kommen.
2:4	22,7 %	405	*impfung, impfstoff, covid, impfungen, zusammenhang*	Die Texte scheinen sich größtenteils auf negative Nebeneffekte von Impfungen gegen COVID-19 zu konzentrieren.
1:2	66,8 %	214	*banane, paypal, pianoacrossthe-world, kanäle, info*	Ein Lied von Björn Banane über Verschwörungstheorien während des Lockdowns mit Links zu seinen Social-Media-Kanälen und Spendenmöglichkeiten.
2:9	20,0 %	165	*lockdown, merkel, pandemie, prozent, reset*	Die Regierung nutze Lockdowns, um Menschen zum Impfen zu zwingen. Kritiker werfen ihr vor, dass ihre Strategie kontraproduktiv sei und die Pharmaindustrie profitiere.
1:5	35,9 %	145	*amazon, itunes, apple, music, unzensiert*	Rapbellions erreichen mit einem Lied einen großen Erfolg in den deutschen Charts, trotz Unterdrückung durch den Mainstream. Es erscheinen bald weitere Lieder.

[3] Die generierten Zusammenfassungen wurden u.U. gekürzt und grammatikalisch korrigiert.

6. Ausblick

6.1 Akteurs-Cluster nach Verhaltensmustern

Practice Mapping ist eine neue Methode, um Netzwerke nach verschiedenen, frei definierbaren Attributen (Features) zu clustern (Bruns et al., 2024). Dafür werden aus Metainformationen Feature-Vektoren erstellt, und somit die Ähnlichkeit zwischen verschiedenen Akteuren anhand der Features errechenbar gemacht. Dadurch können beliebige Metainformationen zum Clustern von Netzwerkdaten genutzt werden. Für die Fragestellung nach besonderen Akteuren für die Verbreitung von Verschwörungstheorien könnten beispielsweise Verweise (Links) auf unterschiedliche Nachrichtenportale oder auch die Verwendung von besonderen Emojis verwendet werden. Diese Features müssen mittels geeigneter qualitativer Methoden erarbeitet werden. Erstellte Netzwerke ließen sich dann mit den Ergebnissen der bisherigen Methoden verknüpfen, um tiefgreifendere qualitative Analysen durchzuführen.

6.2 Multimodale Inhalte

Da die verschiedenen Plattformen mehr Modalitäten als Text anbieten, ist das Potenzial für weitere Analysen, die Bild- Video- und Audiodaten mitverarbeiten können sehr hoch – vor allem, weil Bilder und Videos für die Erzeugung großer Reichweiten und schneller Informationsverbreitung genutzt werden. Eine Herausforderung ist, diese Inhalte so abzubilden, dass sie in das Clustering-Verfahren eingewoben werden können. Insbesondere im letzten Jahr wurde die Entwicklung und Verbreitung von leistungsstarken KI-Methoden, die akkurate Bildbeschreibungen ohne Trainingsdaten erstellen können, vorangetrieben. Auch Audio-zu-Text-Modelle sind mittlerweile so leistungsstark, dass sie mit annehmbarer Fehlerrate automatisiert zur Datenkonvertierung genutzt werden können. Die Einbindung solcher Methoden in die Datenerhebungspipelines würde den Informationsumfang und damit unsere Analysen erheblich aufwerten. Hierbei bleibt aber zu bedenken, dass diese zusätzlichen Prozessschritte sehr ressourcenaufwendig und damit auch mit höheren Kosten verbunden sind.

Fazit

In diesem Report haben wir aufgezeigt, wie aus Sicht der Computational Social Science Daten aus mehreren Social-Media-Plattformen erhoben werden können und welche Strategien und Lösungen in diesem Problemraum anwendbar sind. In der Praxis mehrerer Fallstudien wurden diese Strategien angewendet und verifiziert. Die Entwicklung einer Methode zur Modellierung von Themen und Themenströmungen mittels semantischer Einbettungen von Text und daraus abgeleiteten Ähnlichkeitsnetzwerken steht dabei als Werkzeug für die Aufdeckung von Desinformationsnarrativen im Fokus.

In Zukunft werden diese und ähnliche Methoden durch Austausch von Einzelkomponenten, wie etwa neueren und leistungsfähigeren Sprachmodellen, an Präzision und Skalierbarkeit gewinnen, während für die Verarbeitung multimodaler Inhalte, z.B. durch die vorherige Übersetzung in Texte, ebenfalls Anwendungen denkbar sind.

Referenzen

Bruns, A., Kasianenko, K., Paddinjaredath Suresh, V., Dehghan, E., & Vodden, L. (2024). *Untangling the Furball: A Practice Mapping to the Analysis of Multimodal Interactions in Social Networks* [In preparation].

Bundtzen, S., Matlach, P., Kessling, P., Stegers, F., & Ziock, J. (2022). *Influencer:innen und Verschwörungsrapper.* Institute for Strategic Dialogue Germany. https://isdgermany.org/influencerinnen-und-verschwoerungsrapper/

Clark, M. (2021, August 4). Research Cannot Be the Justification for Compromising People's Privacy. *Meta*. https://about.fb.com/news/2021/08/research-cannot-be-the-justification-for-compromising-peoples-privacy/

DeLong, L. A. (2021, August 21). Facebook Disables Ad Observatory; Academicians and Journalists Fire Back—NYU Center for Cyber Security %. *NYU Center for Cyber Security*. https://cyber.nyu.edu/2021/08/21/facebook-disables-ad-observatory-academicians-and-journalists-fire-back/

Edler, D., Bohlin, L., & Rosvall, M. (2017). Mapping higher-order network flows in memory and multilayer networks with Infomap. *Algorithms*, *10*(4), 112. https://doi.org/10.3390/a10040112

Grave, E., Bojanowski, P., Gupta, P., Joulin, A., & Mikolov, T. (2018). *Learning Word Vectors for 157 Languages* (arXiv:1802.06893). arXiv. https://doi.org/10.48550/arXiv.1802.06893

Grootendorst, M. (2022). *BERTopic: Neural topic modeling with a class-based TF-IDF procedure* (arXiv:2203.05794). arXiv. https://doi.org/10.48550/arXiv.2203.05794

Hidalgo, C. A. (2016). Disconnected, fragmented, or united? A trans-disciplinary review of network science. *Applied Network Science*, *1*(1), 6. https://doi.org/10.1007/s41109-016-0010-3

Jost, P., Heft, A., Buehling, K., Zehring, M., Schulze, H., Bitzmann, H., & Domahidi, E. (2023). Mapping a Dark Space: Challenges in Sampling and Classifying Non- Institutionalized Actors on Telegram. *M&K Medien & Kommunikationswissenschaft*, *71*(3–4), 212–229. https://doi.org/10.5771/1615-634X-2023-3-4-212

Kessling, P., & Münch, F. V. (2023). *Leibniz-HBI/tegracli: V0.2.6* [Software]. Zenodo. https://doi.org/10.5281/zenodo.8043363

Kessling, P., & Münch, F. V. (2023). *Ponyexpress* [Python]. Leibniz-Institut für Medienforschung | Hans-Bredow-Institut. https://github.com/Leibniz-HBI/ponyexpress (Original work published 2022)

Münch, F. V. (2019). *Measuring the Networked Public – Exploring Network Science Methods for Large Scale Online Media Studies* [PhD thesis, Queensland University of Technology]. https://doi.org/10.5204/thesis.eprints.125543

Münch, F. V., & Kessling, P. (2022). *Ukraine_twitter_data.* https://doi.org/10.17605/OSF.IO/RTQXN

Reimers, N., & Gurevych, I. (2019). *Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks* (arXiv:1908.10084). arXiv. https://doi.org/10.48550/arXiv.1908.10084

Schmidt, J.-H., Merten, L., & Münch, F. V. (2023, März 28). *OSF | DBoeS-data*. https://osf.io/sk6t5/

Wang, L., Yang, N., Huang, X., Yang, L., Majumder, R., & Wei, F. (2024). Multilingual E5 Text Embeddings: A Technical Report. *arXiv preprint arXiv:2402.05672*.

Wiedemann, G., Münch, F. V., Rau, J. P., Kessling, P., & Schmidt, J.-H. (2023). Concept and challenges of a social media observatory as a DIY research infrastructure. *Publizistik*. https://doi.org/10.1007/s11616-023-00807-6

Wiedemann, G., Schmidt, J.-H., Rau, J., Münch, F. V., & Kessling, P. (2023). Telegram in der politischen Öffentlichkeit. Zur Einführung in das Themenheft. *M&K Medien & Kommunikationswissenschaft*, *71*(3–4), 207–211. https://doi.org/10.5771/1615-634X-2023-3-4-207

HAW HAMBURG

LEIBNIZ-INSTITUT FÜR MEDIENFORSCHUNG HANS-BREDOW-INSTITUT

ISD Institute for Strategic Dialogue

GEFÖRDERT VOM

Bundesministerium für Bildung und Forschung