



Technische Universität Berlin
Fachgebiet Internet und Gesellschaft
Leitung: Prof. Dr. Bettina Berendt

Programmierpraktikum: Analyse von (Online) Social Data

Projektbericht

Echo Chamber-Effekt bei Videovorschlägen auf YouTube

Jonathan Brüders

Emin Egemen Hidiroglu

Jannik Leander Hyun-Ho Novak

Elena Osipova

Gunnar Scherdin

Gruppe 1

Betreuerin: Sonja Mei Wang

19. August 2023

Inhaltsverzeichnis

1 Abstract	3
2 Einleitung	4
2.1 Motivation	4
2.2 Ziele des Projektes	4
2.3 Stand der Forschung	4
2.4 Definition von Echo Chamber-Effekt	6
2.5 Auswahl von Seed Videos	7
3 Methoden	10
3.1 Datensammlung	10
3.2 Datenauswertung	11
3.2.1 Datenauswertung für Graphen	11
3.2.2 Datenauswertung für Topic Modelling	11
4 Ergebnisse der Analyse	14
4.1 Analyse von Graphen	14
4.2 Topic Modelling mit BERT	20
5 Diskussion der Ergebnisse	26
6 Konklusion	26
7 Limitationen und Anknüpfungspunkte für zukünftige Arbeiten	27
8 Quellen	28
8.1 Papers	28
8.1 Verwendete Software	30
9 Anhang	31

1 Abstract

In diesem Projekt untersuchten wir den Vorschlagalgorithmus von YouTube, um das Verhalten von Kanal-Empfehlungen innerhalb verschiedener thematischer Bereiche zu analysieren. Wir analysierten die Wechselwirkungen zwischen Kanälen, indem wir die gegenseitige Vorschlagshäufigkeit in den empfohlenen Videos untersuchten.

Im Projekt nutzten wir die "YouTube Data API" in Verbindung mit der Programmiersprache Python, um Daten zu sammeln. Anschließend erfolgte die Analyse dieser Daten mithilfe von Gephi und BERTopic.

Unsere Ergebnisse zeigen, wie der Algorithmus die Verbindungsmuster zwischen Kanälen beeinflusst und wie stark die Empfehlungen innerhalb bestimmter Themenbereiche verankert bleiben. Die Ergebnisse dieses Projekts verdeutlichen, dass die Themenvariation der vorgeschlagenen Videos je nach Thematik variiert. Diese Erkenntnisse tragen dazu bei, ein tieferes Verständnis für die Funktionsweise des YouTube-Algorithms zu entwickeln.

2 Einleitung

2.1 Motivation

Anfänglich ging es uns vorrangig darum, ein Bild der Youtube Landschaft zu bekommen, ähnlich wie bei diesem [Video](#), das die Beziehung zwischen deutschen Youtubern untersucht. Dieses Video stellte visuell sehr effektiv dar, was oft vermutet wurde: Die "Querdenker-Szene" ist extrem isoliert vom Rest von Youtube. Zusätzlich dazu gibt und gab es immer wieder Berichte über "Echochamber" oder "Filterbubbles" und das wollten wir näher untersuchen.

Nachdem wir uns tiefergehend mit dem Thema beschäftigt haben, fiel uns auf, dass oftmals die Frage offen blieb, ob eine beobachtete Tendenz zu Echo-Chambers vom Nutzer oder von Youtube erzeugt wurden. Um diese Frage beantworten zu können, kann es auch hilfreich sein, die "related videos" aus der Youtube API zu beziehen, da hier kein individuelles Nutzerverhalten Einfluss auf die Videos hat.

2.2 Ziele des Projektes

- Echo Chamber-Effekt bezüglich des Verhaltens des YouTube-Algorithmus analysieren.
- Mithilfe verschiedener Ausgangsvideos potenzielle Unterschiede in verschiedenen Themenbereichen in Bezug auf das Phänomen der Echo-Chambers zu identifizieren.
- Die gesammelte Information in Graphen konvertieren und analysieren, ob der Echo Chamber-Effekt vorhanden ist.
- Durchführung von Topic Modelling auf der Basis der gesammelten Informationen.

2.3 Stand der Forschung

Verschiedene Studien haben das Phänomen der Echo Chambers und den Einfluss der Videoempfehlungen von YouTube auf das Sehverhalten der Benutzer untersucht. In einer Studie, die auch als eine unserer Quellenstudien verwendet wurde, wurden 78 regelmäßige YouTube-Benutzer analysiert, während sie 10 beliebte Videos aus dem Bereich der gemeinnützigen Organisationen ansahen. Dabei wurden die sogenannten "Up-Next" Videos untersucht, die vom YouTube-Algorithmus empfohlen wurden. Die Ergebnisse zeigten, dass der Algorithmus in 57% der Fälle Videos vorschlug, die nichts mit der betrachteten

gemeinnützigen Organisation zu tun hatten. Die Unzusammenhänglichkeit mit dem entsprechenden Thema betrug 41%, was auf das Fehlen der Schaffung von Echo Chambers hinweist[8]. In einer zweiten Studie, bei der eine breitere Auswahl an Videos verwendet wurde und die im Inkognito-Modus durchgeführt wurde, erhöhte sich der Prozentsatz der Nicht-Empfehlung von Videos, die sich auf die gemeinnützige Organisation bezogen, auf 59%, und die Unzusammenhänglichkeit mit dem entsprechenden Thema betrug 44,6%. Es wurde auch festgestellt, dass Benutzer, je tiefer sie in den Rabbit Hole der Empfehlungen vordringen, umso mehr vom Thema abweichen.[8] Da die Daten für diese Studie jedoch nur für Videos von gemeinnützigen Organisationen erhoben wurden, ist sie begrenzt. Daher ist es möglicherweise nicht richtig, dasselbe für Videos mit anderen Themen anzunehmen. Aufgrund dieser Tatsache beabsichtigen wir, in unserer Studie auch das mögliche Vorhandensein von Echo Chambers in Bezug auf verschiedene Themen zu untersuchen.

In einer anderen Forschungsarbeit lag der Fokus auf dem Teilen von YouTube-Videos in ideologischen Subreddits, um die Ideologie jedes Videos vorherzusagen. Eine Sammlung von über 50.000 Videos wurde als Trainingsdaten für einen natürlichen Sprachklassifizierer verwendet, damit die Ideologie jedes YouTube-Video bestimmt werden kann. Danach wurden in einer experimentellen Studie echte YouTube-Benutzer eingeschrieben und erhielten ein Ausgangsvideo sowie eine Regel zur Auswahl der empfohlenen Videos um den Einfluss von User-Verhalten auf empfohlene Videos zu vermeiden. Dieser Ansatz sollte den Einfluss des YouTube-Algorithmus isolieren und untersuchen, ob er Benutzer in Echo Chambers oder in bestimmte ideologische Richtungen lenkt. Die Wahl der Benutzer wurde aus dem Prozess entfernt, und die empfohlenen Videos wurden anhand ihrer vorhergesagten Ideologie bewertet. Die Methoden zur Bestimmung der Ideologien der Videos und Vermeidung des User-Verhaltens sind eigentlich wichtig für die Analyse von Echo-Chambers aber gleichzeitig auch aufwändig. Deswegen würde es sinnvoll das Youtube Algorithmus unbeschränkt und unisoliert zu analysieren und weniger Fokus auf die politische Neigung des Algorithmus zu geben.

Darüber hinaus hat eine andere Studie eine Analyse von Daten aus einer repräsentativen Stichprobe von über 300.000 Amerikanern von Januar 2016 bis Dezember 2019 durchgeführt, um die politische Kategorisierung von YouTube-Videos anhand der Besuche von Benutzern auf bestimmten URLs zu untersuchen. Eines unserer Ziele zu Beginn der Studie war es, politische und nicht-politische Videos unterscheiden zu können. Dafür haben wir analysiert, wie Videos in anderen Studien als politisch oder nicht-politisch gekennzeichnet wurden. Um uns mit anderen Studien zu vereinheitlichen, die unterschiedliche Klassifizierungssysteme verwenden, wurden die Kategorisierungen auf sechs

Kategorien abgebildet, wie sie in einer unserer Quellstudien verwendet wurden: stark links (fL), links (L), Mitte (C), anti-woke (AW), rechts (R) und stark rechts (fR). Diese Klassifikationen stammen aus zwei früheren Studien [6] und [7]. Die Ergebnisse deuten darauf hin, dass es begrenzte Belege für die weit verbreitete Behauptung gibt, dass YouTube Benutzer dazu verleitet, radikalere politische Inhalte zu konsumieren, egal ob links oder rechts. Stattdessen kann YouTube als ein weites Meer an Informationen, Fehlinformationen und einer Vielzahl anderer politischer und nicht-politischer Inhalte betrachtet werden. Da unsere Studie ähnliche Muster in politischen Videos aufgezeigt hat, war es logisch, dass wir uns stärker auf das mögliche Vorhandensein von Echo-Chambers im Allgemeinen konzentrieren, anstatt uns auf spezifische ideologische oder politische Faktoren zu beschränken.

2.4 Definition von Echo Chamber-Effekt

Im Rahmen unserer Projektarbeit haben wir viel recherchiert, um den Echo-Chamber-Effekt angemessen zu definieren. Hierfür haben wir verschiedene Definitionen aus unterschiedlichen wissenschaftlichen Papers analysiert und miteinander verglichen, um eine präzise und treffende Definition zu formulieren, die unseren Projektzielen gerecht wird.

Unsere Projektziele in Bezug auf Echo Chambers waren:

- Wir wollten den Echo Chamber-Effekt bezüglich des Verhaltens des YouTube-Algorithmus analysieren und nicht bezüglich des Verhaltens und der individuellen Überzeugungen einer Person.
- Um den Echo Chamber-Effekt zu erkennen, war es von Bedeutung, festzustellen, dass die große Mehrheit der vorgeschlagenen Videos dasselbe Thema wie das Ausgangsvideo behandelt und/oder von demselben YouTube-Kanal veröffentlicht wurde.

Basierend auf unseren Recherchen konnten wir folgende Definitionen des Echo Chamber-Effekts ermitteln:

- “Echo Chambers sind Umgebungen, in denen die Meinung, die politische Einstellung oder das Befinden der Nutzer zu einem Thema durch wiederholte Interaktion mit gleichaltrigen Nutzern oder Quellen, die ähnliche Tendenzen und Einstellungen vertreten, verstärkt wird.” (Quelle: [1])

- “Bei Echo Chambers sind Einzelpersonen Inhalten ausgesetzt, die ihre bestehenden Überzeugungen verstärken. Echo Chambers wurden in Themenbereichen wie Politik, Wissenschaft und Verschwörungstheorien dokumentiert, in denen Verbraucher oft stark abweichende Ansichten haben und nach Inhalten suchen, die ihre Ansichten bestätigen, und Plattformen können dann algorithmisch Inhalte empfehlen, die diese bestätigen.” (Quelle: [2])
- “Politische Echo-Chambers sind die Bestätigung bereits bestehender Überzeugungen und den Ausschluss widersprüchlicher Überzeugungen innerhalb einer Person. (Quelle: [3])
- “Echo Chambers sind Umgebungen, in denen Personen ideologisch konsistenten Informationen ausgesetzt sind. Echo-Chambers auf YouTube bedeutet, dass die Nutzer vorwiegend Videos aus einer politischen Kategorie konsumieren.” (Quelle: [4])
- “Einzelpersonen können in ideologischen Echo Chambers leben, in denen sie selten mit alternativen Perspektiven konfrontiert werden. Ideologische Echo Chamber - eine Verteilung von Videos für einen bestimmten Nutzer, die ideologisch homogen ist und sich auf die eigene Ideologie des Nutzers konzentriert.” (Quelle: [5])

Da wir in unserem Projekt nicht das Verhalten einer Person und ihre persönliche Überzeugungen analysieren wollten, haben wir die Definitionen aus den Papers [1], [2], [3] und [5] ausgeschlossen und nicht weiter berücksichtigt.

Definition [4] war hingegen besser zutreffend. Diese Definition haben wir an unsere Projektziele angepasst und somit unsere Definition von Echo Chamber-Effekt erhalten:

Echo Chambers sind Umgebungen, in denen Personen überwiegend mit Informationen konfrontiert werden, die ideologisch einheitlich sind. *Echo Chambers auf YouTube* bedeutet, dass ein Algorithmus dem Nutzer Videos vorwiegend über ein Thema und/oder aus demselben YouTube Kanal vorschlägt.

2.5 Auswahl von Seed Videos

Die Auswahl der Seed Videos war von wichtiger Bedeutung im Hinblick auf die erstrebte Erkenntnis aus dem Projekt. Eine reine Auswahl an Videos aus verschiedenen politischen Spektren mit ähnlicher Popularität und ähnlicher Aktualität zum Beispiel, würde darauf abzielen Echo Chambers innerhalb der politischen Szene zu untersuchen und herauszufinden, ob eine politische Richtung eine stärkere Neigung dazu hat, Videos aus dem eigenen

Spektrum vorzuschlagen als eine Andere. Unser Ziel war es jedoch vorerst grundlegend das Phänomen der Echo Chambers und die damit verbundenen Algorithmen auf der Plattform YouTube besser zu verstehen und mögliche Unterschiede in verschiedenen Themenbereichen zu erkennen. Hierfür ist eine Repräsentation von verschiedensten Themen in den Seed Videos essentiell. Unsere Wahl fiel auf die Analyse von drei Videos mit Potential zum Bilden von Echo Chambers, sowie drei Vergleichsvideos ohne ersichtlichen ideologischen Denkweisen. Dabei war uns wichtig, sowohl verschiedene Themengebiete, als auch verschiedene große räumliche Kontexte zu betrachten.

Angefangen mit einem **Gitarren-Tutorial**, welches eine reine Hobby-Sparte im Musikbereich repräsentiert, in der sich jegliche Form von Polarität alleine auf den Musikgeschmack der Hörer beschränkt und somit keine gesellschaftlichen Uneinigkeiten aufwirft, wählten wir hiermit ein Video, welches wenig Potential an Kontroverse und eine niedrige emotionale Involvierungen besitzt.

Der gewählte Trailer des Videospiels **"It Takes Two"** wurde von uns als Video mit dem höchsten Potential, in andere Themengebiete geleitet zu werden, bewertet. Sich die niedrige Wahrscheinlichkeit von Kontroversen und emotionalen Diskussion mit dem **"Gitarren-Tutorial"** teilend, bringt ein Videospiel-Trailer den wichtigen Aspekt mit sich, dass das Anschauen des Video oft aus der Überlegung entspringt, das vorgestellte Spiel zu erwerben. Somit ist das gezielte Suchen des Videos denkbar. Interessant ist, ob der Youtube Algorithmus das weitere Vorschlagen von ähnlichen Videospiel-Trailern für sinnvoll hält oder sich unsere Vermutung bestätigt und die vorgeschlagenen Videos in andere Themengebiete leiten.

Mit dem **CSGO-Video** wird die Gaming-Kategorie repräsentiert. Als eines der größten und erfolgreichsten Themengebiete auf YouTube, fällt Gaming (in diesem Fall ein **"Let's Play"** des Spiels CSGO) ebenfalls in die Kategorie Unterhaltung. Im Unterschied zu dem Videospiel-Trailer handelt es sich hierbei allerdings um das aufgenommene Spielen eines Spielers mit gleichzeitigem Kommentar, während der Videospiel-Trailer mit cinematischen Effekten eher einem Kurzfilm ähnelt.

Außerdem haben wir ein Video über die Verschwörungstheorie der **"Flat-Earth"** ausgewählt. Hier ist die Ausprägung einer Echo Chamber möglich, da diese Theorie ideologische Ansätze bietet. Dieses Video repräsentiert sowohl das Themengebiet der Verschwörungstheorien, als auch einen internationalen Raum. Bei diesem Video erwarteten

wir eine starke Ausprägung von ähnlichen Themen und Kanalbetreibern mit ähnlichen Denkweisen.

Mit einer [Rede des aktuellen deutschen Bundeskanzlers](#) ist der räumliche Kontext auf die nationale Ebene gewechselt. Trotz des sachlichen und theoretisch objektiven Anlasses einer Regierungserklärung zur Entwicklung der deutschen Politik, hielten wir eine Bildung von Echo Chambers aus dieser Ausgangssituation in viele Richtungen für möglich.

Als letztes Video wählten wir eine [AfD Wahlwerbung](#) aus dem Jahr 2023 des AfD-eigenen Kanals AFD TV. Den noch kleineren räumlichen Kontext von Berlin repräsentierend, sprachen wir diesem Video eine hohe Wahrscheinlichkeit für die Bildung einer Echo Chamber zu. Durch eine bereits stark geprägte ideologische Ansicht zu vielen Themen ist es interessant zu sehen, in welchem Maß der Youtube Algorithmus diese Denkweise mit den vorgeschlagenen Videos weiter vertieft und ob eine differenzierte Bildung der eigenen Meinung aufgrund der vorgeschlagenen Videos möglich ist.

3 Methoden

3.1 Datensammlung

- *Mittel zur Datensammlung.*

Wir haben uns entschieden die „Youtube Data API“ zur Sammlung von Daten zu verwenden. Die Alternative hierzu wäre gewesen einen Scraper zu schreiben, was nach dem deutschen Gesetz in einer Grauzone liegt, weshalb wir lieber darauf verzichteten. Ein weiterer positiver Aspekt der verwendeten API ist, dass es möglich ist verwandte Videos anzufragen, was uns befähigte ein Netzwerk aus verwandten Videos zu erstellen. Hierzu standen wir außerdem vor der Entscheidung, ob wir nicht einen Weg finden könnten, den „YouTube Empfehlungsalgorithmus“ nachzuvollziehen, und somit eine akkurate Repräsentation der Videos zu erhalten, die einem Nutzer tatsächlich vorgeschlagen werden. Da dieser Algorithmus proprietär und somit nicht öffentlich einsehbar ist, mussten wir uns jedoch mit den „Verwandten Videos“ zufriedengeben, welche durch die „YouTube Data API“ verfügbar sind.

- *Umsetzung*

Nachdem wir uns einen abstrakten Plan erstellt hatten, wie wir die Daten sammeln und verarbeiten wollen, begannen wir mit der Implementierung unserer Idee. Dafür entschieden wir uns die Programmiersprache Python zu verwenden, da Python einerseits eine Sprache ist, welche auf einer hohen Ebene operiert, und wir somit einiges an Code aus vorhandenen Libraries entnehmen konnten. Andererseits bietet Python viele Werkzeuge, welche gerade für die Verarbeitung und Analyse von Daten sehr hilfreich sind, was uns in unserem Vorhaben äußerst gelegen kam. Bei der Implementierung sind wir auf einige Hürden gestoßen, welche im Zusammenhang mit dem täglichen Nutzungslimit der API, und dem korrekten Zusammenfügen der gesammelten Daten standen. Das erste Problem lösten wir, indem wir über unsere Gruppenmitglieder verteilt insgesamt 19 API-Keys erstellten. Weiterhin fanden wir einen Weg unsere gesammelten Daten in log Dateien zu speichern und das Sammeln eines Datensatzes über mehrere Tage hinweg fortzusetzen, wenn wir das tägliche Kontingent der API überschritten hatten.

- *Methode*

In diesem Abschnitt erläutern wir einmal näher die Arbeitsweise unseres Programmes zur Datensammlung. Die grundlegendste Funktion unseres Programmes besteht darin, von einem Video ausgehend, einen Baum verwandter Videos zu erstellen, dessen Tiefe und Breite anhand von Eingabeparametern bestimmt werden können.

Daraufhin wird eine abstrakte Darstellung des Baumes in Form einer Liste aus Listen von Hash Maps, welche die Ebenen des Baumes repräsentieren, in einer log Datei gespeichert. Da wir nun dem Problem gegenüberstanden, dass wir aufgrund des Nutzungslimits der YouTube API nur eine limitierte Menge an Daten pro Tag sammeln, und zu einem größeren Baum zusammenfügen konnten, mussten wir zusätzlich einen Mechanismus implementieren, damit das Programm weiß, an welcher Stelle es die Datensammlung fortsetzen muss. Somit konnten wir dann das Programm jeden Tag automatisch laufen lassen, bis wir alle nötigen Daten gesammelt hatten. Für nähere Implementierungsdetails verweisen wir hiermit noch auf das [Gitlab Repository](#) welches wir zum kollaborativen programmieren verwendeten.

3.2 Datenauswertung

3.2.1 Datenauswertung für Graphen

- *Erstellung der Graphen*

Es wurde bereits erwähnt, dass wir unsere gesammelten Daten in log Dateien speicherten, welche wir nun im weiteren Verlauf zu Graphen umwandelten. Dafür fügten wir in unserem Programm die Funktionalität hinzu, dass eine log Datei in eine graphml Datei umgewandelt werden kann. Hierbei sollte noch einmal erwähnt werden, dass die log Datei eine abstrakte Darstellung eines Baumes von verwandten Videos ist. Die Umwandlung dieser Dateien in Graphen geschah anhand der Richtlinien, dass erstens jede Kante zwischen zwei Knoten des Baumes, in den Graphen übernommen wird. Weiterhin wurde für jedes wiederholte vorkommen einer Kante, das Kantengewicht für diese Kante im Graphen erhöht. Zuletzt skalierten wir die Größe der Knoten des Graphen anhand der Häufigkeit der Knoten in dem Ausgangsbau.

- *Weitere Verarbeitung*

Nun verblieb uns nur noch die graphml Datei mithilfe des Netzwerkanalyse- und Netzwerkdarstellungs-Programms Gephi zu visualisieren, und zur weiteren Auswertung aufzubereiten. Dafür verwendeten wir eine Kombination aus dem „Fruchterman Reingold“ Algorithmus und dem „Force Atlas 2“ Algorithmus.

3.2.2 Datenauswertung für Topic Modelling

- *Eigenschaften des Datensatzes.*

Nachdem wir die Daten gesammelt hatten, erhielten wir unseren Datensatz in

Form von einer log-Datei für jedes Seed-Video. Der vollständige Datensatz für das Projekt ist im Anhang enthalten.

- *Wie wurde der Datensatz bereinigt?*

Um das Topic Modeling durchführen zu können, haben wir den erhaltenen Datensatz bereinigt. Dazu haben wir verschiedene Schritte mit Hilfe von Python-Bibliotheken (siehe 8.1) unternommen:

- Wir haben aus der log-Datei die Themen der Videos extrahiert, dadurch konnten wir einen Überblick über die verschiedenen Themen erhalten, die in den Videos behandelt wurden.
- Um die Daten übersichtlicher darzustellen, haben wir die log-Datei in eine CSV-Datei konvertiert, um sie als Tabelle anzeigen zu können.
- Um spätere Bereinigungsschritte zu erleichtern, haben wir alle Großbuchstaben in der csv-Datei in Kleinbuchstaben umgewandelt. Dadurch können wir einheitlich arbeiten und mögliche Inkonsistenzen aufgrund von Groß- und Kleinschreibung vermeiden.
- Um den Text für spätere Analysen besser handhabbar zu machen, haben wir alle Sonderzeichen, Symbole und Emojis aus der csv-Datei entfernt. Durch diese Bereinigung bleiben nur die reinen Textinhalte erhalten, was die Qualität und Verständlichkeit der Daten für das Topic Modeling verbessert.
- Mithilfe der Python-Bibliothek "nltk.corpus" haben wir eine Liste mit häufig vorkommenden Wörtern wie "und", "oder", "der", "die", "das" und "aber" erstellt, die wenig zur Bedeutung der Texte beitragen. Anschließend haben wir diese Wörter aus den Daten gelöscht. Dadurch konnten wir die Daten von irrelevanten und allgemeinen Wörtern befreien, um den Fokus auf die bedeutungsvollen Inhalte zu legen.

- *Wie wurden Topics erstellt?*

Wir haben uns für die Verwendung von BERT zur Erstellung der Themen entschieden. Hierbei haben wir BERT mit unseren Eingaben trainiert, um die Themen zu erstellen. Nach dem Training des Modells haben wir die Themenextraktion durchgeführt (Beispiel in Abbildung 1).

Topic	Count	Name
-1	86	-1_umgang_cranberries_steve_zombie
0	53	0_greifen_gitarren_einfach_tipps
1	48	1_schnelle_finger_teil_beste
2	41	2_wechseln_gitarren_einfach_anfnger
3	38	3_fingerpicking_zupfen_mehrere_gleichzeitig
4	35	4_tne_sinnvoll_finden_tonleitern
5	35	5_riffs_guitar_plan_master
6	34	6_hilft_sofort_unbewegliche_finger

Abbildung 1: Topic Modelling mit BERT. Beispiel aus dem Datensatz für Gitarrentutorial

Für weitere Details siehe bertopic.py auf GitLab 8.1.

- *Wie wurden Topics visualisiert?*

Bei der Visualisierung von Themen haben wir folgende Ziele verfolgt:

- die Struktur von Themen zu erkennen;
- Ähnlichkeit oder Unterschiedlichkeit zwischen den Themen festzustellen;
- das Hauptthema zu identifizieren, sofern möglich;
- die Existenz von Echo Chambers zu bestätigen oder zu widerlegen.

Dementsprechend haben wir uns für zwei Methoden entschieden: *Hierarchische Labels* und *Similarity Matrix*.

Wir nutzen hierarchische Labels, um die Themen in eine Hierarchie zu bringen und dadurch ihre Struktur besser zu verstehen. Zusätzlich ermöglicht uns die Zusammenführung von Themen, ihre Ähnlichkeiten zu erkennen. Um die Ähnlichkeiten und Unterschiede noch deutlicher darzustellen, verwenden wir auch die Similarity Matrix Methode. Diese beiden Methoden helfen uns ebenfalls dabei, das Hauptthema zu identifizieren. Dadurch können wir die Existenz von Echo Chambers bestätigen oder widerlegen.

4 Ergebnisse der Analyse

4.1 Analyse von Graphen

- Gitarren-Tutorial

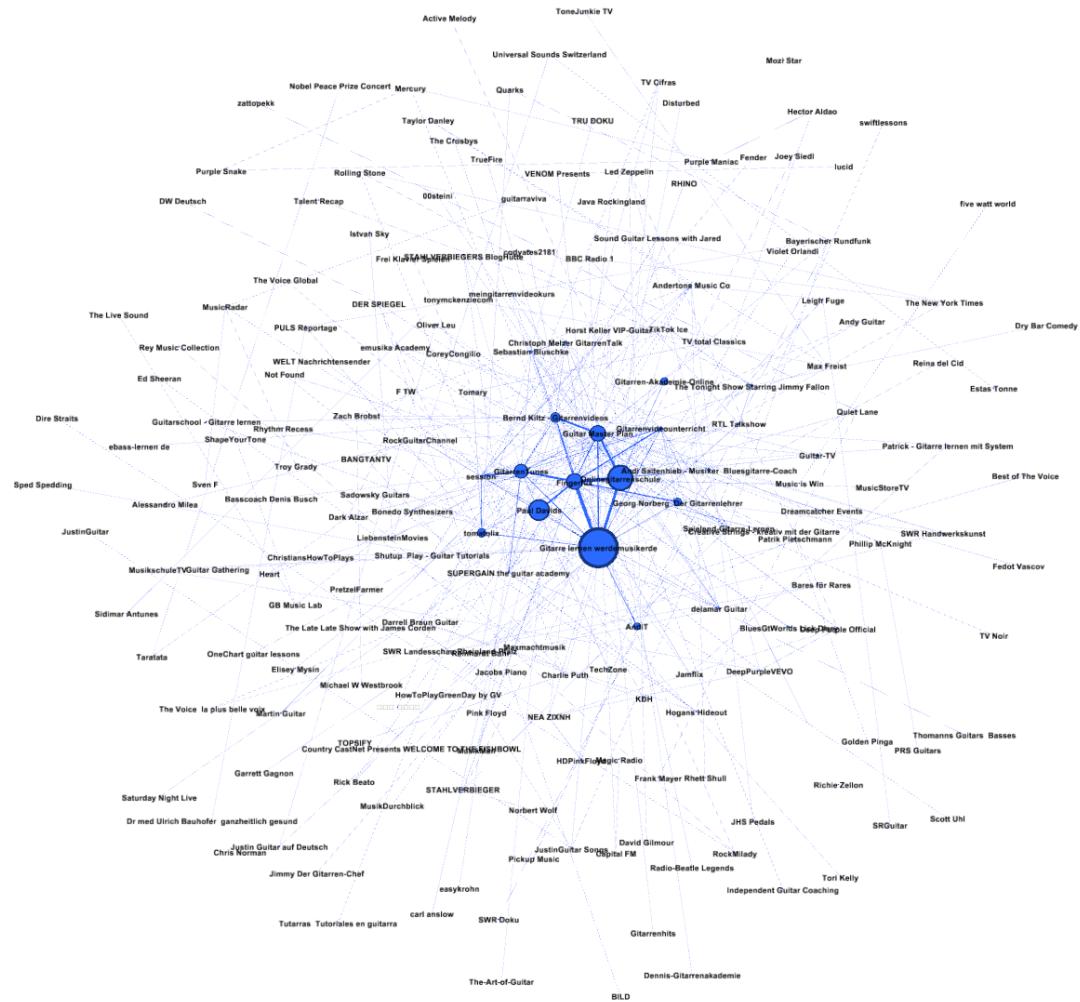


Abbildung 2: Netzwerkgraph für "Gitarren-Tutorial" Video

Die Kanäle sind vergleichsweise stark zentriert mit nur wenigen großen Knoten, welche starke Verbindungen untereinander aufweisen. Das Zentrum besteht weiterhin aus Knoten, welche thematisch sehr Gitarren-fokussiert sind.

- CS:GO

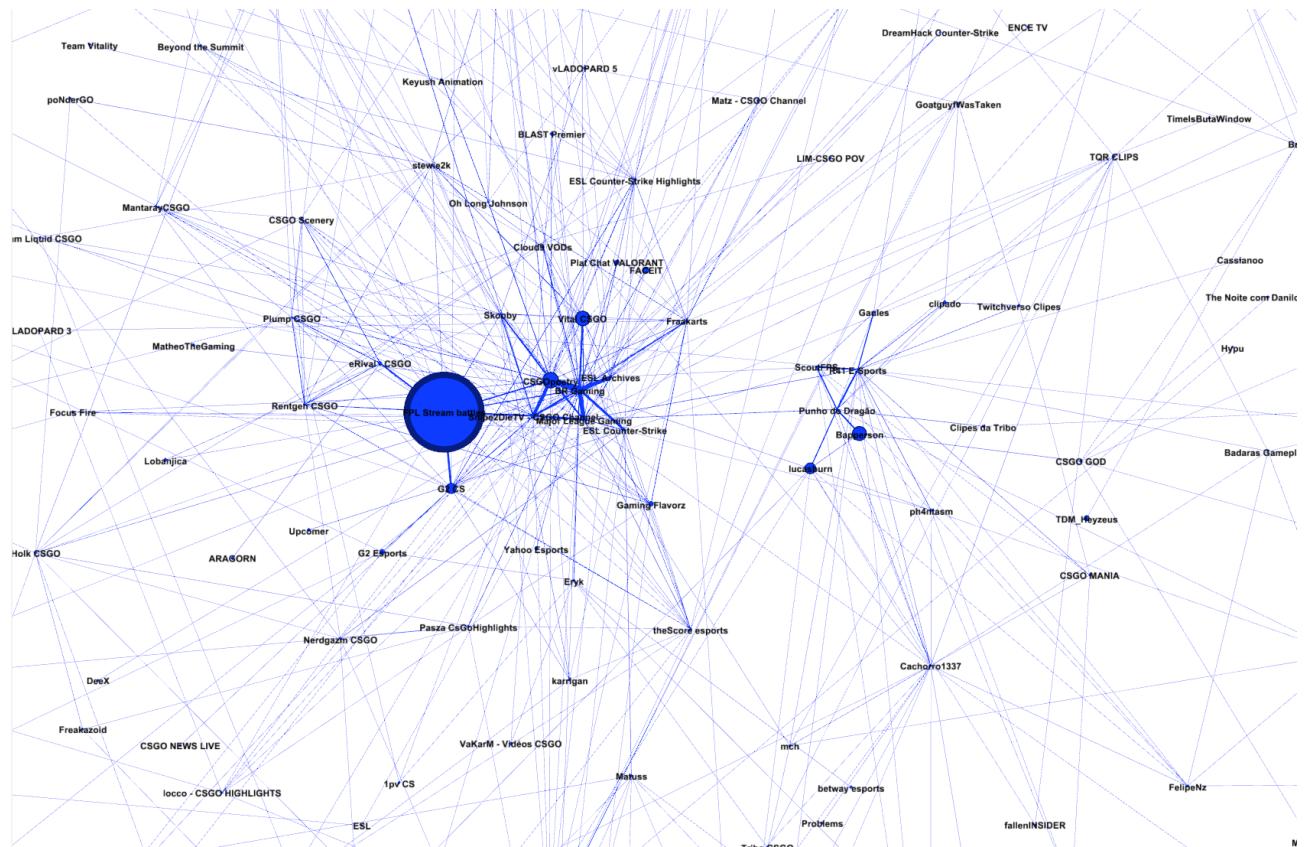


Abbildung 3: Netzwerkgraph für "CS:GO" Video

In diesem Graph haben sich vorrangig zwei Cluster gebildet, welche sich in einerseits englische und andererseits brasilianische Kanäle unterteilt. Während der Großteil der Kanäle nur eine geringe Anzahl an Vorschlägen aufweist, ist der Kanal "FPL Stream Battle" deutlich häufiger vertreten.

- It Takes Two Trailer

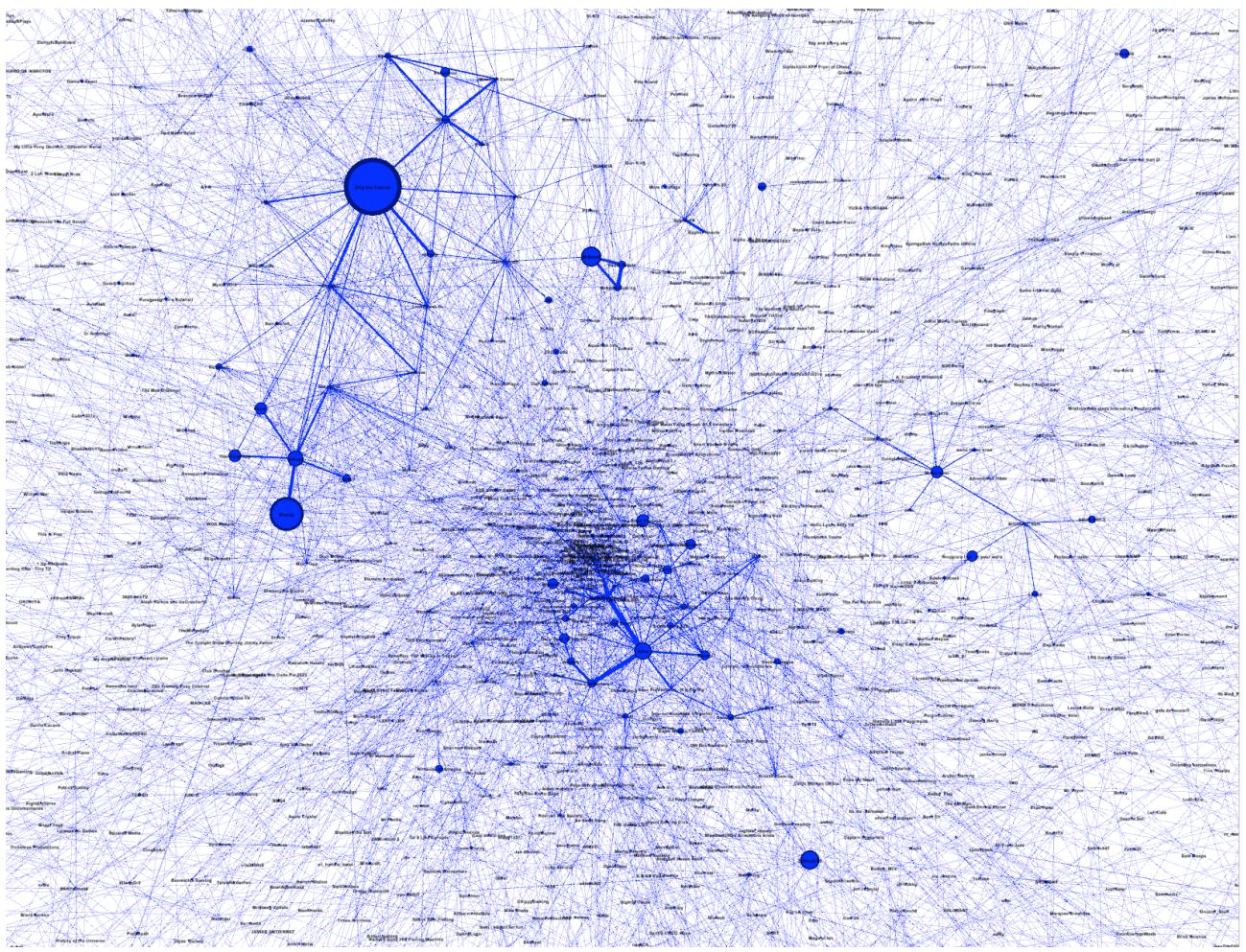


Abbildung 4: Netzwerkgraph für "It Takes Two Trailer" Video

Der Graph zeigt Kanäle mit sehr diversen Themenbezügen und stark variierenden Distanzen untereinander. Die zwei größten Knoten (Skip the Tutorial und Branzy) repräsentieren Minecraft-Kanäle, was die vielen Minecraft-Themen im Topic Modeling erklärt. Der EA-Kanal, von dem auch das Seed-Video gewählt wurde, bildet das Zentrum, jedoch sind Videos dieses Kanals nicht sehr häufig vertreten. Generell ist das Thema des Ausgangsvideos nicht signifikant relevant für die Themenbereiche des finalen Graphens.

• Flat-Earth

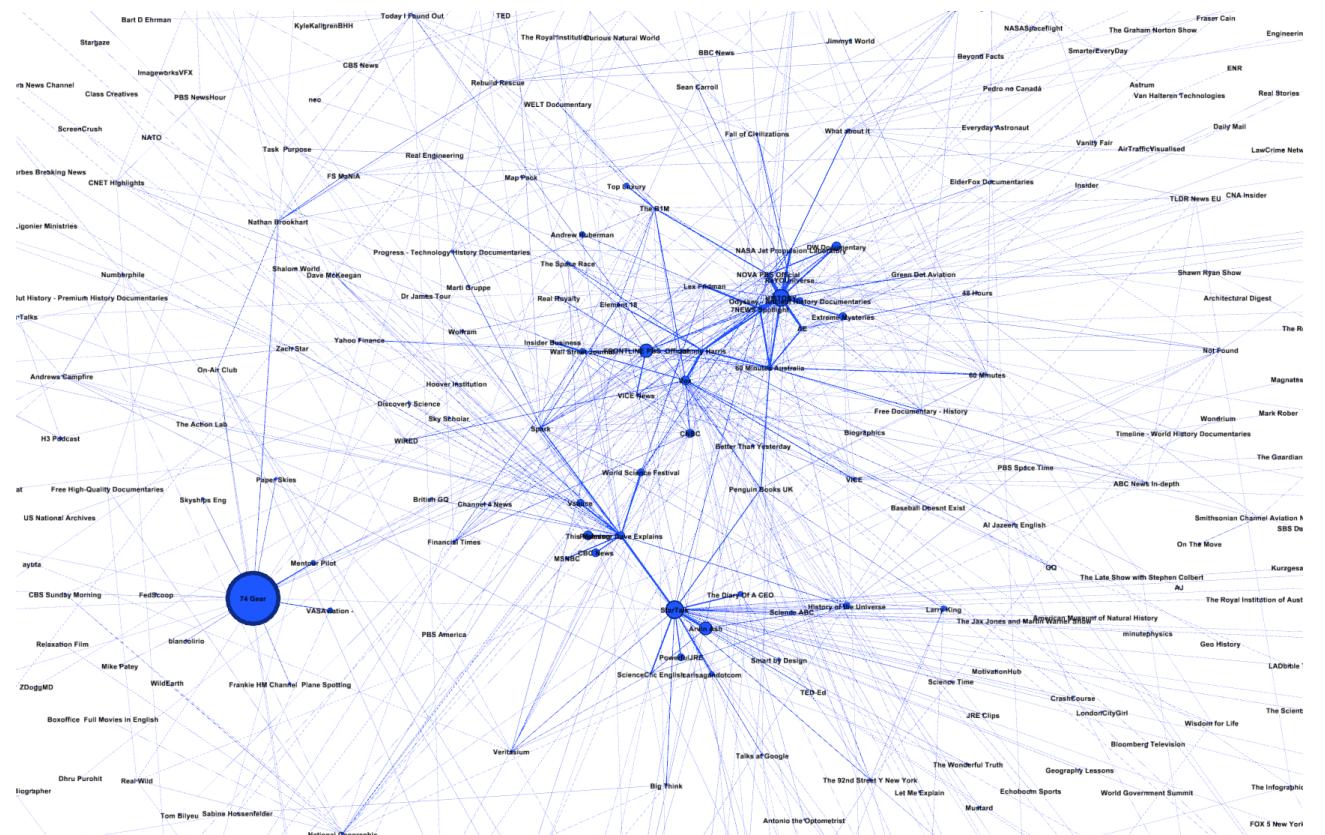


Abbildung 5: Netzwerkgraph für "Flat-Earth" Video

Mehrere erkennbare Cluster haben sich gebildet. Die Zentren dieser Cluster besitzen eine hohe Distanz zu den Zentren anderer Cluster. Auffällig ist der Kanal 74 Gear, von welchem deutlich häufiger Videos vorgeschlagen wurden, wobei diese Vorschläge vorrangig von eigenen Videos waren, was man an der hohen Distanz zu allen anderen Knoten erkennen kann. Dies wären mögliche Indizien für eine Echo-Chamber.

• Scholz-Rede

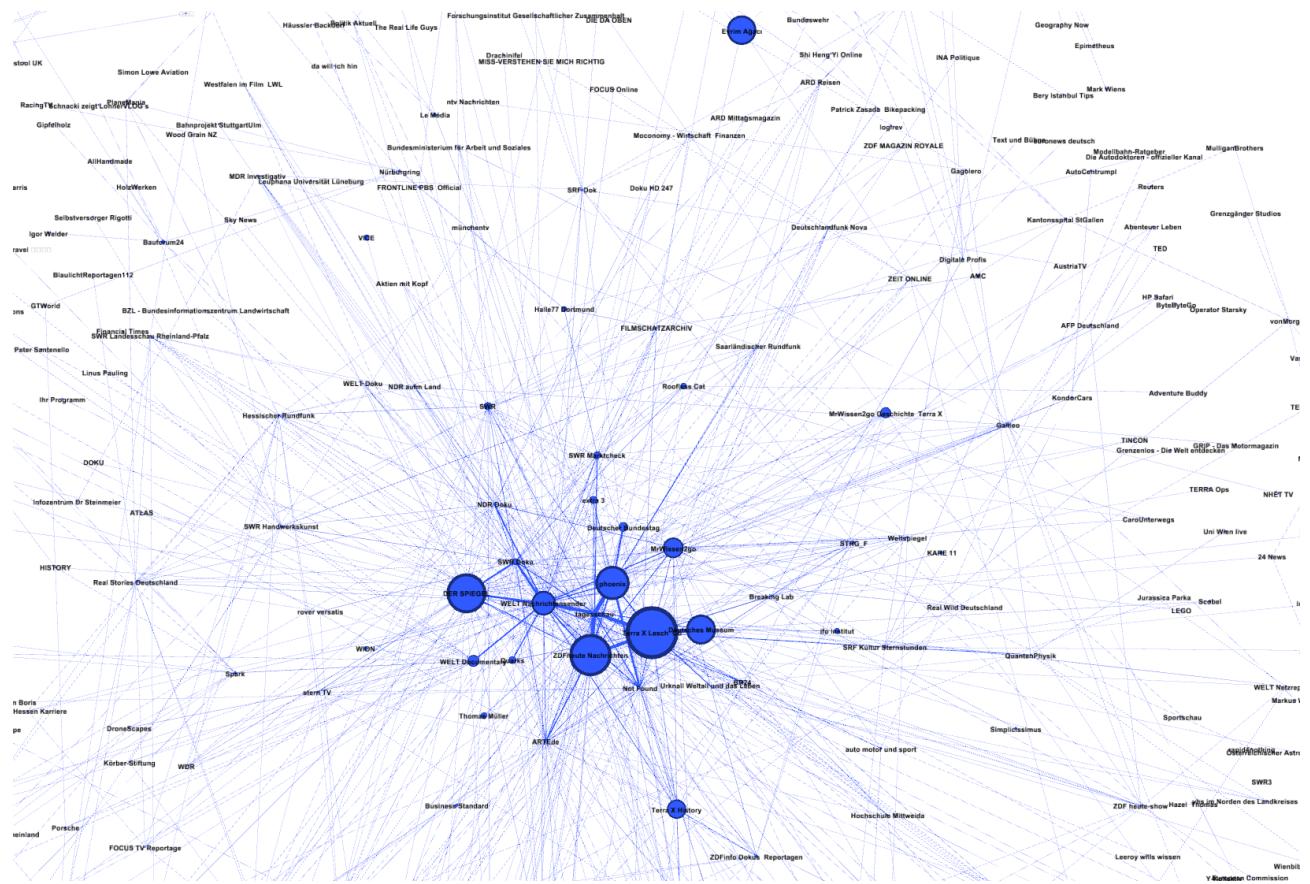


Abbildung 6: Netzwerkgraph für "Scholz-Rede" Video

Im Gegensatz zu dem vorherigen Graphen ist hier ein deutliches Zentrum erkennbar, in welchem alle Kanäle einen gleichmäßigen Abstand zueinander besitzen. Die relevanten Kanäle sind in moderater Häufigkeit vertreten und es stechen keine großen Kanäle heraus, welche sich abgegrenzt haben.

- AFD Wahlwerbung 2023

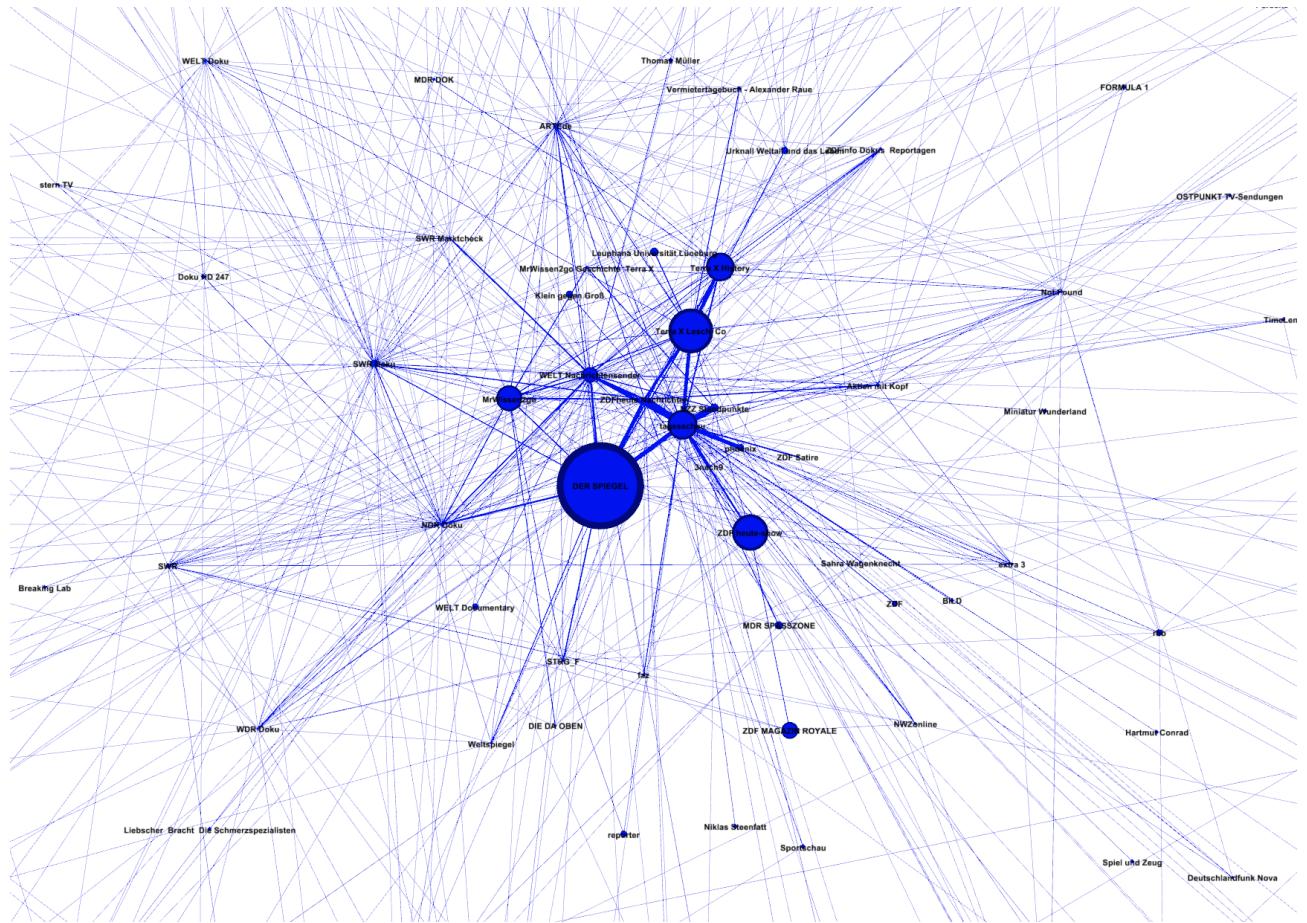


Abbildung 7: Netzwerkgraph für ”AfD” Video

Es ist eine starke Konzentration der Knoten in der Mitte mit einigen großen Knoten und starker Vernetzung untereinander erkennbar. Die wenigen großen Knoten im Zentrum sind zwar politisch orientierte Nachrichtenkanäle (Spiegel, Mr. Wissen2Go, TerraX, Heute Show etc.), vertreten jedoch nicht die selben ideologischen Werte wie die AfD Partei. Im Gegenteil werden sogar Kanäle vorgeschlagen, die sich teilweise äußerst kritisch gegenüber der AFD positionieren (Heute Show, ZDF Magazin Royale, TerraX) und es wurden vergleichsweise wenige Kanäle aus dem konservativen Spektrum wie Welt, Sarah Wagenknecht etc. vorgeschlagen.

4.2 Topic Modelling mit BERT

Zur besseren Lesbarkeit aller Abbildungen siehe HTML-Dateien im Anhang.

- Gitarren-Tutorial

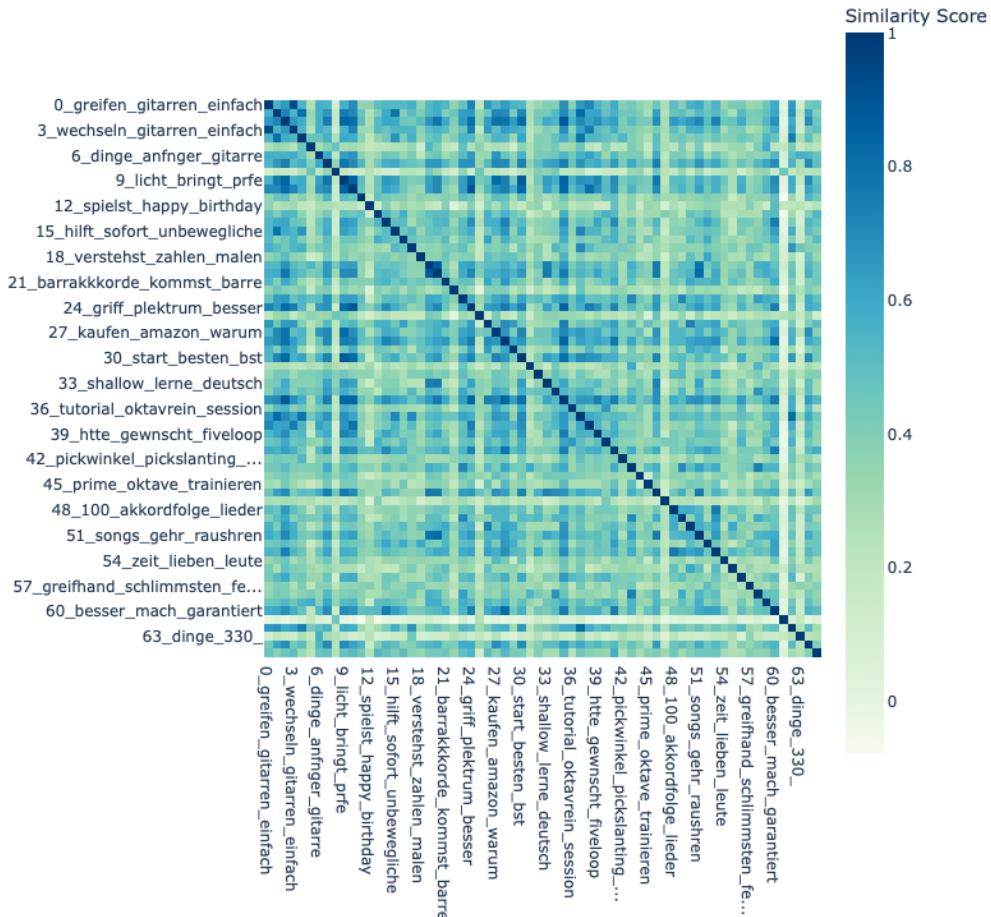


Abbildung 8: Similarity Matrix für "Gitarren-Tutorial" Video

Beim "Gitarren-Tutorial" Seed-Video zeigt die Similarity Matrix eine auffällige Dunkelblau-Färbung. Dies weist auf eine hohe Ähnlichkeit der Themen im Topic Modeling hin und deutet auf eine geringe Abweichung von einem zentralen Themengebiet hin. Bei genauerer Analyse bestätigt sich dieses Muster. Die Themen sind stark fokussiert auf Videos über Anfängerübungen auf der Gitarre, wie zum Beispiel "greifen Gitarren einfach" oder "wechseln Gittare einfach".

- CS:GO

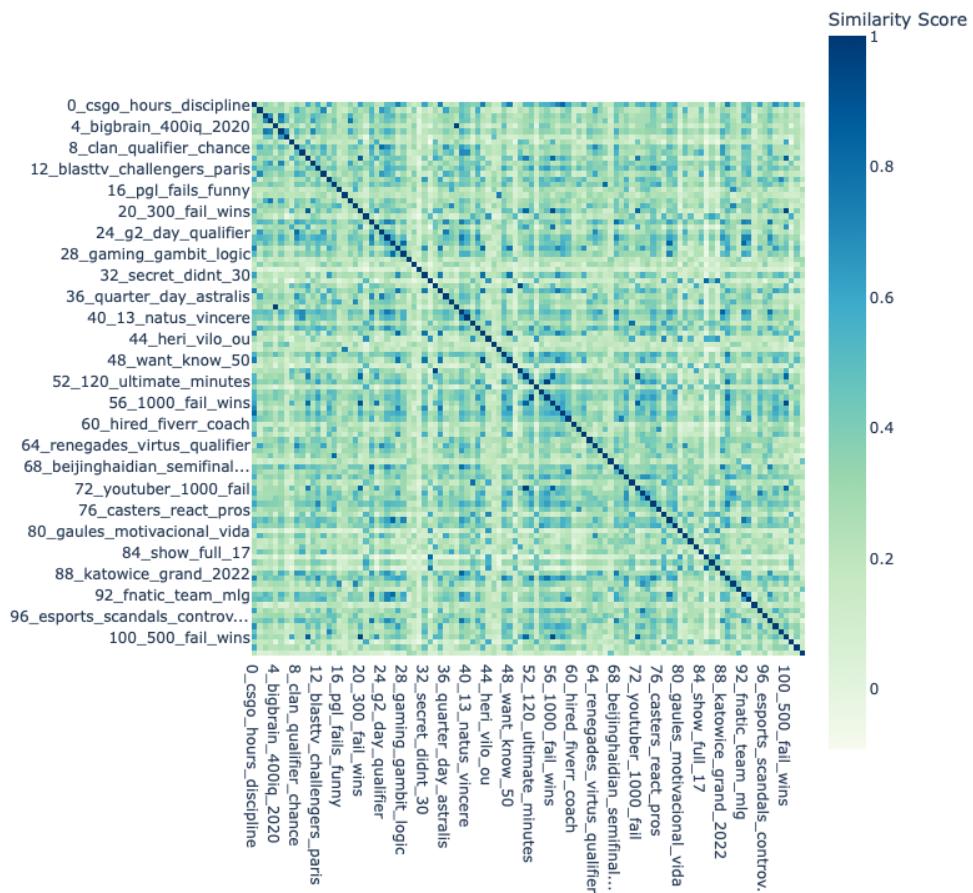


Abbildung 9: Similarity Matrix für "CS:GO" Video

Im Fall des CS:GO Seed-Videos ist die Verteilung der Themen ungleichmäßig. Es gibt sowohl Themen mit starken Häufigkeiten als auch solche mit geringen Häufigkeiten. Eine Vielzahl von Themen befindet sich im Bereich des Grünen, während nur wenige im dunkelblauen Bereich liegen. Die Themen variieren, jedoch bleiben sie im Rahmen von CS:GO. Einige beziehen sich auf Influencer-Inhalte ("Engagierte einen Fiverr-Coach"), andere sind Zusammenstellungen von "epischen Fehlern", während wieder andere den E-Sport behandeln. Die Vielfalt innerhalb des CS:GO-Universums ist groß, jedoch mag sie aus externer Sicht weniger unterschiedlich erscheinen.

- It Takes Two Trailer

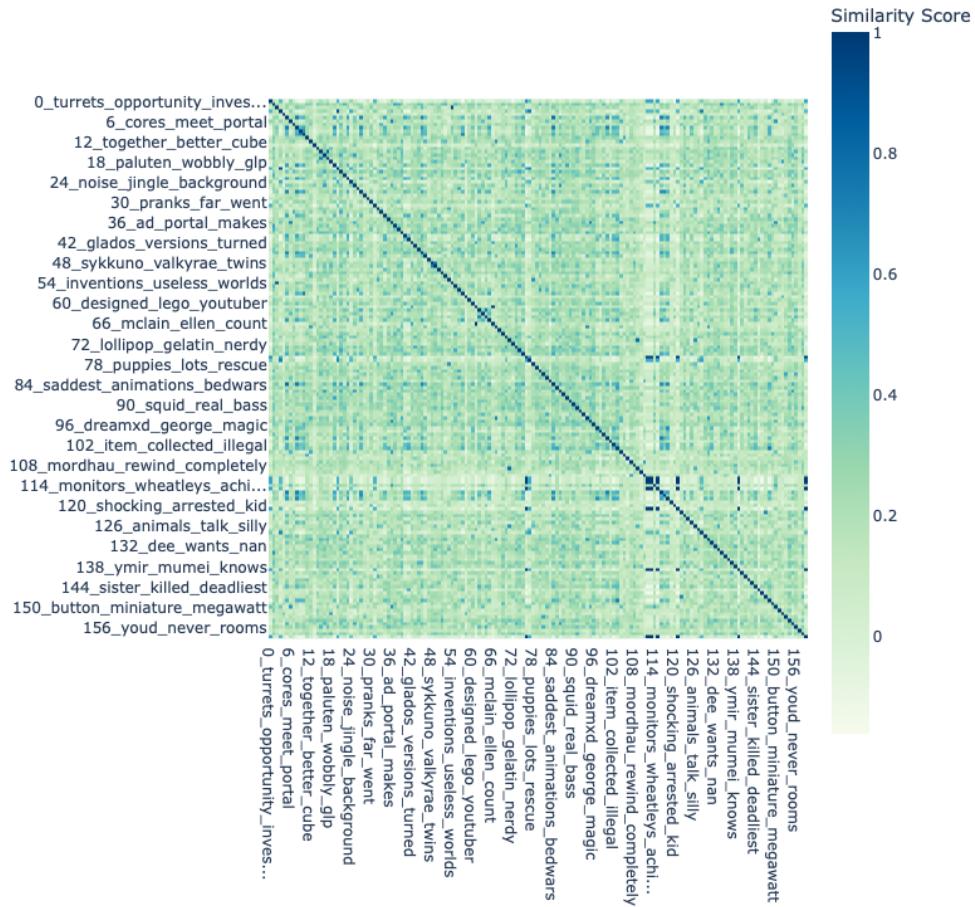


Abbildung 10: Similarity Matrix für "It Takes Two" Video

Beim "It Takes Two Trailer" Seed-Video ist deutlich erkennbar, dass nahezu alle Themen im hellgrünen Bereich liegen. Dies lässt darauf schließen, dass die Themen vielfältig sind und nur wenig miteinander gemeinsam haben.

- Flat-Earth

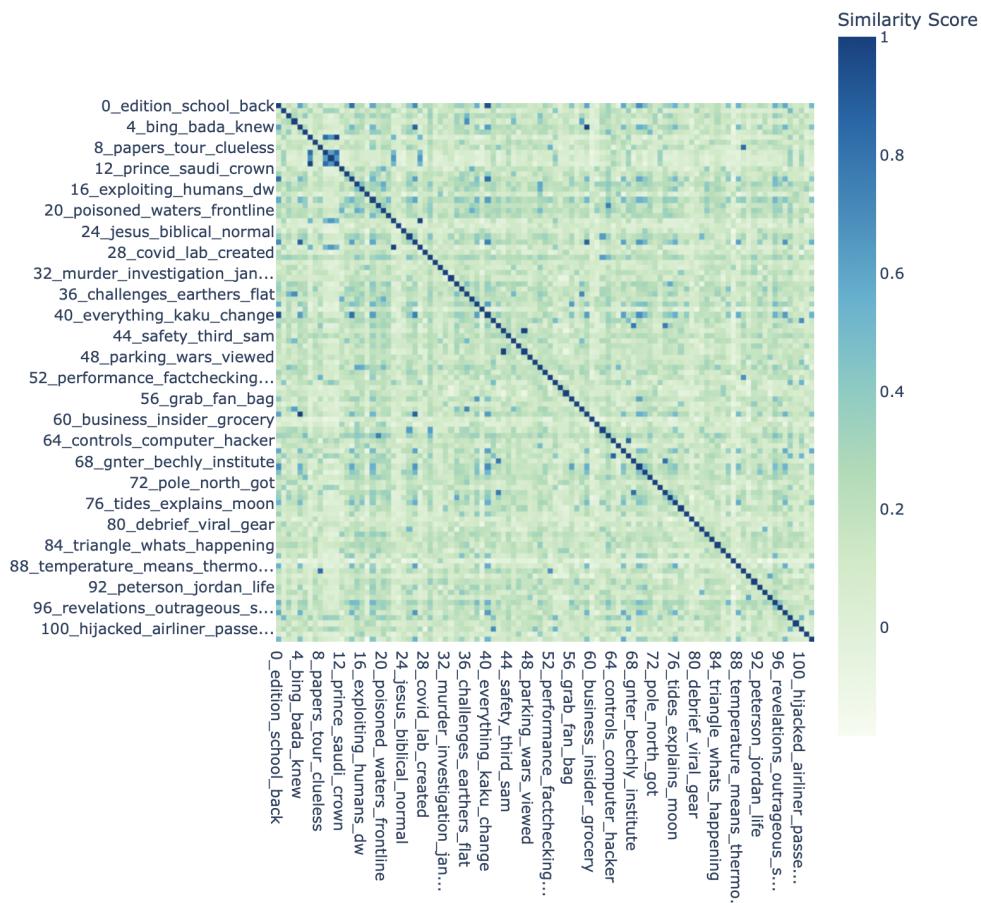


Abbildung 11: Similarity Matrix für "Flat Earth" Video

Ähnlich wie im "It Takes Two Trailer" ist auch die Verteilung der Themen im "Flat Earth" Seed-Video gestaltet. Die überwiegende Mehrheit der Themen liegt im hellgrünen Bereich und hat daher einen niedrigen Ähnlichkeitswert.

- Scholz-Rede

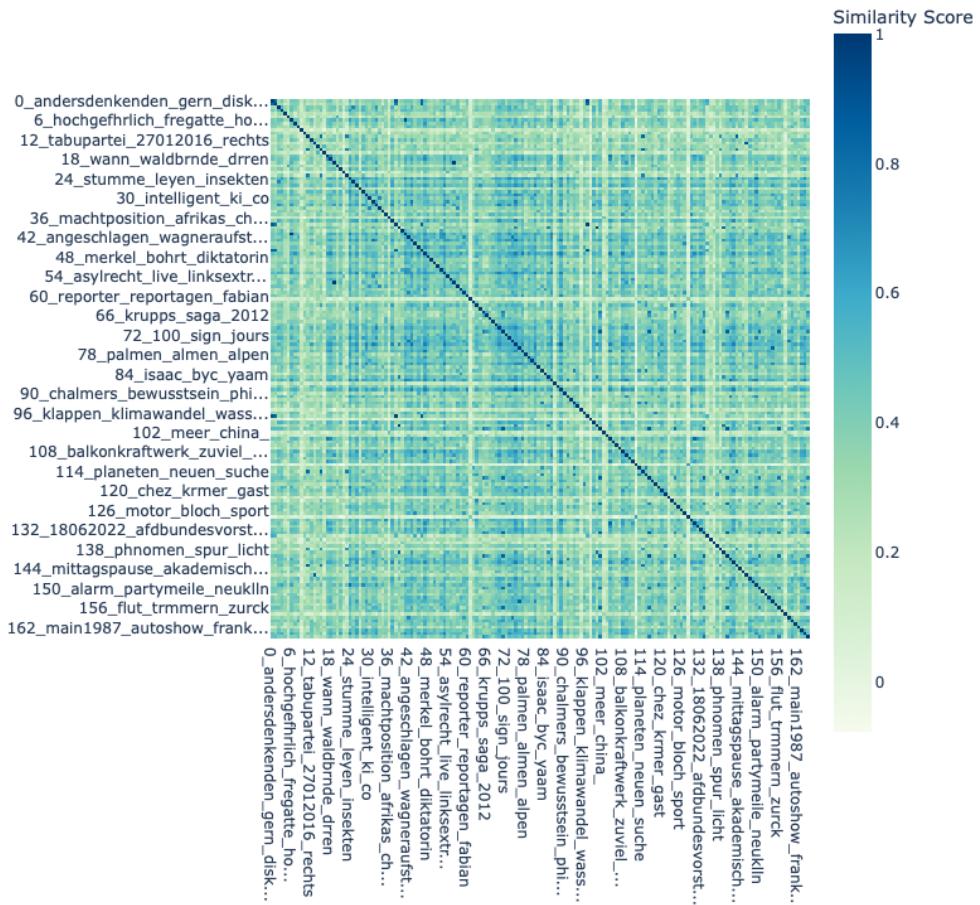


Abbildung 12: Similarity Matrix für "Scholz Rede" Video

Im "Scholz-Rede" Seed-Video sind die Themen stärker miteinander verknüpft als in den beiden vorherigen Videos. Dies zeigt sich daran, dass mehrere Themen im dunkelblauen Bereich liegen. Die vorgeschlagenen Videos behandeln hauptsächlich politische Themen, die in Form von dokumentarischen Filmen oder Talkshows präsentiert wurden.

- AFD Wahlwerbung 2023

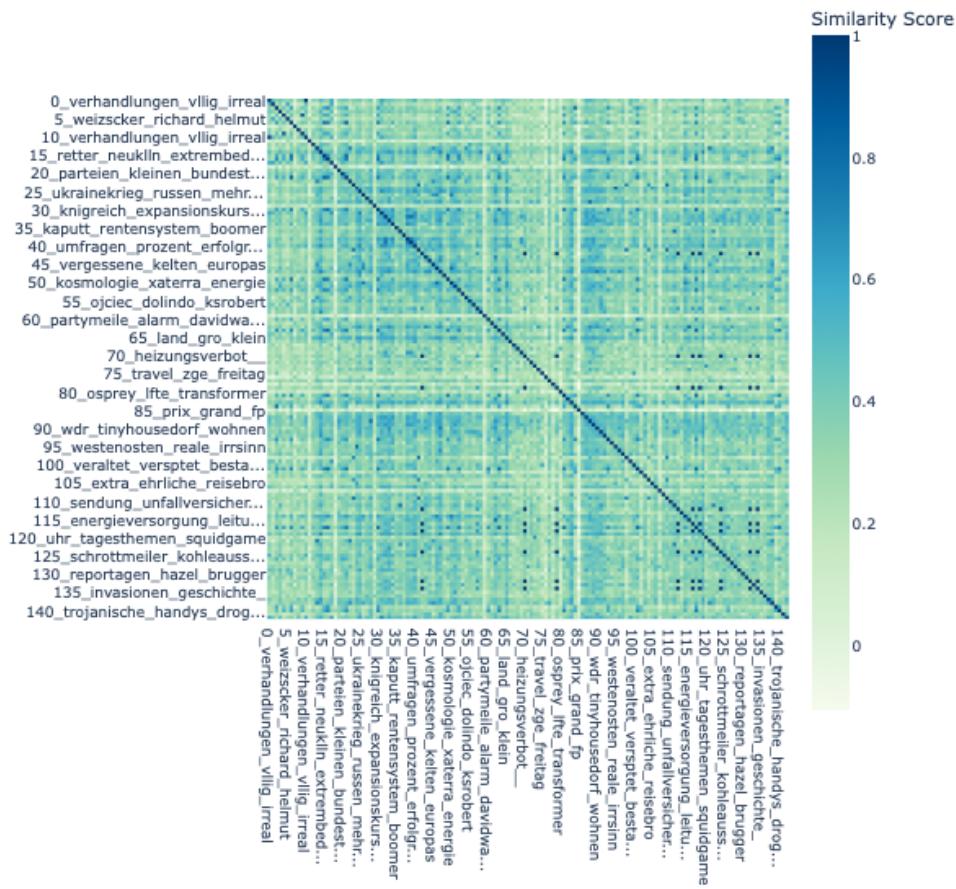


Abbildung 13: Similarity Matrix für "AFD Wahlwerbung" Video

Beim "AFD Wahlwerbung 2023" Seed-Video handelt es sich hauptsächlich um vorgeschlagene Nachrichtenvideos, die verschiedene politische Themen behandeln. Anhand der aufgeführten Themen wird deutlich, dass sie zwar politischen Inhalt haben, jedoch nicht die gleichen ideologischen Werte wie die AfD-Partei vertreten.

5 Diskussion der Ergebnisse

Insgesamt haben wir nur für sechs Seedvideos Daten erhoben und vor allem entstammt jedes dieser Videos aus einer eigenen Kategorie. Es ist also nicht möglich aus beispielsweise den Daten der AFD Wahlwerbung eine definitive Aussage über alle Videos der AFD oder anderer rechter Parteien zu ziehen. Dennoch bieten die Daten einen kleinen Einblick in die nähere Funktionsweise des Algorithmus von Youtube.

6 Konklusion

Es gibt ein breites Spektrum bei den Videos, wie stark die Themen und auch die Kanäle untereinander verknüpft sind. Bei dem einzigen Beispiel bei dem die Kriterien für eine Echo-Chamber am ehesten erfüllt sind, das Gitarren Tutorial, kann nicht von einer Echo-Chamber gesprochen werden, da die Inhalte gar nicht bis kaum ideologisch gestaltet sind, sondern lediglich Informationen über das Gitarre spielen vermitteln.

Die Videos bei denen Ideologie und politische Positionen klar vertreten sind (Scholz Rede, AFD Wahlwerbung, Flat-Earth) ergaben zwar sehr eng verbundene Kanalgraphen, jedoch waren die Themen relativ weit gefächert. Außerdem waren viele der Kanäle Nachrichtensender wie DER SPIEGEL, ZDFheute Nachrichten oder WELT Nachrichtensender, die zwar ihre eigene politische Richtung haben können, aber auch immer zusammen auftraten. Es kann hier also in keinem der Fälle von einer Echo-Chamber gesprochen werden.

Stattdessen haben wir eine andere Beobachtung gemacht. Der Algorithmus, der die "related videos" auswählt, bevorzugt scheinbar einige Kanäle/Themen mehr als andere. So waren in beiden deutschen Videos auch immer die ZDF Heute Show und Terra X vertreten oder beim "It Takes Two Trailer" hat sich ein großes Minecraft Cluster gebildet, obwohl es ansonsten eine eher heterogene Umgebung hatte.

7 Limitationen und Anknüpfungspunkte für zukünftige Arbeiten

Unsere Arbeit offenbart folgende Limitationen und mögliche Ansatzpunkte für zukünftige Analysen:

- Die erzielten Ergebnisse spiegeln lediglich einen Augenblick wider, da der Algorithmus einem ständigen Wandel unterliegt.
- Unsere Untersuchung konzentrierte sich ausschließlich auf einen unpersonalisierten Algorithmus und berücksichtigte nicht das Verhalten der Nutzer.
- Die eingeschränkte Anzahl der analysierten Videos beeinträchtigt die Aussagekraft in negativer Weise.
- Ein umfangreicheres Videomaterial würde zu deutlicheren Ergebnissen führen.

8 Quellen

8.1 Papers

Literatur

- [1] Reliability of Content and Echo Chambers on YouTube during the COVID-19 Debate
Niccolò Di Marco, Matteo Cinelli, Walter Quattrociocchi (July 20, 2021)
https://workshop-proceedings.icwsm.org/pdf/2022_64.pdf
- [2] Algorithmic Recommendations and Earned Media: Investigating Product Echo Chambers on YouTube
Lambrecht, Anja and Sen, Ananya and Tucker, Catherine E. and Wiertz, Caroline (October 27, 2021)
https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3951425
- [3] Examining Political Bias within YouTube Search and Recommendation Algorithms
Lutz, M., Gadaginmath, S., Vairavan, N., Mui, P. (2021, December)
<https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=9660012>
- [4] Evaluating the scale, growth, and origins of right-wing echo chambers on YouTube
HosseiniMardi, H., Ghasemian, A., Clauset, A., Rothschild, D. M., Mobius, M., Watts, D. J. (2020)
<https://www.arxiv-vanity.com/papers/2011.12843/>
- [5] Echo Chambers, Rabbit Holes, and Algorithmic Bias: How YouTube Recommends Content to Real Users
Brown, M. A., Bisbee, J., Lai, A., Bonneau, R., Nagler, J., Tucker, J. A. (2022)
https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4114905
- [6] Algorithmic extremism: Examining youtube's rabbit hole of radicalization
Mark Ledwich and Anna Zaitsev (2019)
[http://arxiv.org/abs/1912.11211arXiv:1912.11211.](http://arxiv.org/abs/1912.11211arXiv:1912.11211)
- [7] Auditing radicalization pathways on youtube
Manoel Horta Ribeiro, Raphael Ottoni, Robert West, Virgilio A. F. Almeida, and Wagner Meira (2021)
<http://arxiv.org/abs/1908.08313arXiv:1908.08313>

- [8] Algorithmic Recommendations and Earned Media: Investigating Product Echo Chambers on YouTube
Lambrecht, Anja and Sen, Ananya and Tucker, Catherine E. and Wiertz, Caroline
(October 27, 2021)
<https://ssrn.com/abstract=3951425>

8.1 Verwendete Software

- GitLab
https://git.tu-berlin.de/gunnar_scherdin/pposd-youtube-network
- YouTube API
 - google api python client == 2.86.0
- Python und Python Libraries:
 - matplotlib == 3.7.0
 - networkx == 3.1
 - nltk == 3.8.1
 - numpy == 1.24.2
 - pandas == 2.0.2
 - requests == 2.31.0
 - scikit learn == 1.2.2
 - torch == 2.0.1
 - transformers == 4.30.2
 - re == 3.11
 - emoji == 2.5.1
 - pandas == 1.3
 - bertopic == 0.15.0
- Gephi
- LaTeX Editor Overleaf
- tubCloud

9 Anhang

- Python Skript
(https://git.tu-berlin.de/gunnar_scherdin/pposd-youtube-network/-/tree/main/src)
- tubCloud mit allen Dateien für Topic Modelling
(<https://tubcloud.tu-berlin.de/s/Pfng7YNWqc3sMF6?path=%2F>)
- Datensätze für die einzelnen Videos
(<https://tubcloud.tu-berlin.de/s/Pfng7YNWqc3sMF6?path=%2FDatenstze>)
- Alle Abbildungen aus dem Abschnitt 4.2 (Topic Modelling mit BERT) als HTML-Dateien
(https://tubcloud.tu-berlin.de/apps/files/?dir=/YouTube_Projekt/BERT%20Visualisierung&fileid=3673234552)