
Article:
Hanchard, M. orcid.org/0000-0003-2460-8638 (2022) The construction of rare disease discourse on YouTube: highlighting a disparity between policy rhetoric and patient practices around public engagement [version 2; peer review: 1 approved with reservations]. Wellcome Open Research, 6. 361. ISSN 2398-502X

https://doi.org/10.12688/wellcomeopenres.17324.2

RESEARCH ARTICLE

REVISED The construction of rare disease discourse on YouTube: highlighting a disparity between policy rhetoric and patient practices around public engagement [version 2; peer review: 1 approved with reservations]

Matthew Hanchard

iHuman institute, University of Sheffield, Sheffield, Yorkshire, S1 4DP, UK

v2 **First published:** 22 Dec 2021, **6**:361
https://doi.org/10.12688/wellcomeopenres.17324.1
Latest published: 01 Apr 2022, **6**:361
https://doi.org/10.12688/wellcomeopenres.17324.2

Abstract

Background: Policy rhetoric around the 6,000-8,000 rare diseases affecting 300 million people worldwide often focuses on public engagement. Meanwhile, medical authorities tend either to treat patients with rare diseases as pre-categorised data sources, proffer to them notions of technological self-care as empowerment, or recruit them as advocacy allies. Conversely, people living with rare diseases often mobilise and engage with one another in self-organised communities via social media to share discussion, information, and resources. How rare disease discourse forms on specific social media platforms, the role of different actors (including medical authorities and algorithms), and its relation to public engagement policy are poorly understood.

Methods: This paper examines data on YouTube video watching/sharing (gathered from YouTube's API via DMI's 'Data Tools for YouTube') through social network analysis (read through a controversy analysis lens).

Results: The paper identifies eight patterns – each revolving around different levels of: *focus* on rare disease content; *engagement* between content and viewers, i.e. through likes, dislikes, and surrounding particular videos*; permeability* of videos between categories; and *repetition* in viewers watching the same video. Across six of the patterns, the paper finds a rare disease issue-network forming, where discourse is constructed through three distinct communication strategies, each garnering a different form of engagement.

Conclusions: Overall, the paper highlights a disconnect between how rare disease discourse is enacted on YouTube and policy promises of public engagement, with potential spaces for dialogue often closed off by medical authorities. To close, the paper provides recommendations for how policymakers might engage with and facilitate more inclusive forms of social media interaction between specific rare disease related

Open Peer Review

Approval Status ?

1

version 2 (revision) 01 Apr 2022

version 1 22 Dec 2021 ? view

1. **Narelle Warren**, Monash University, Melbourne, Australia

Any reports and responses or comments on the article can be found at the end of the article.

communities and clinicians to develop more meaningful forms of knowledge exchange.

Keywords
Controversy analysis, Issue-mapping, Public engagement, Rare disease, Rare disease day, Social network analysis, YouTube

Corresponding author: Matthew Hanchard (m.s.hanchard@sheffield.ac.uk)

Author roles: Hanchard M: Conceptualization, Data Curation, Methodology, Visualization, Writing – Original Draft Preparation, Writing – Review & Editing

Competing interests: No competing interests were disclosed.

Grant information: This research was supported by Wellcome [219875]. For the purpose of open access, the author has applied a CC BY public copyright licence to any Author Accepted Manuscript version arising from this submission.

How to cite this article: Hanchard M. **The construction of rare disease discourse on YouTube: highlighting a disparity between policy rhetoric and patient practices around public engagement [version 2; peer review: 1 approved with reservations]** Wellcome Open Research 2022, **6**:361 https://doi.org/10.12688/wellcomeopenres.17324.2

First published: 22 Dec 2021, **6**:361 https://doi.org/10.12688/wellcomeopenres.17324.1

REVISED **Amendments from Version 1**

I have now added a sentence to the end of Introduction section to clarify these as two separate goals. The text reads: "As such, the paper examines both the construction of discourse and the role of an algorithm in shaping it, with more weight given to examining the latter.

I have redrafted the Framework section, adding more detail in the second and third paragraphs.

I have now added a sentence to clarify what crawl depth means to t methods section. T reads: "Crawl depth depicts the depth to which search is conducted, and thus provides an incremental set of results. For this query, a crawl depth of 0 returned 50 videos, while a crawl depth of 1 returned 250"

I have now added further detail on the qualitative research into the Methods section (in the last paragraph).

I have now included producers within the ethical considerations section. They should have been included in the first version. The suggested paper will also be useful for an upcoming one I am writing based on Twitter data around rare diseases. My thanks to the reviewer.

I have created a table (Table 2) and submitted it as part of the revised paper to aid comparison of clusters and patterns

Any further responses from the reviewers can be found at the end of the article

Introduction

Over 300,000,000 people worldwide live with a rare disease (Yáñez-Muñoz, 2017), defined in America, Europe and many other territories as long-term health conditions affecting fewer than 1 in 20,000 citizens (Côté & Keating, 2012; Mikami, 2019). To understand the experiences of patients with rare diseases, medical authorities (i.e. patient organisations, pharmaceutical companies, policymakers, and researchers) have recently turned to patient engagement/empowerment (Bauer, 2017; EMA, 2020; HM Govt, 2020) following a wider turn to inclusive governance (de Saille, 2015a) that 'actively involve[es] and support[s] patients in health care and treatment decision making activities' (Grande *et al.*, 2014, p. 281). The benefits of doing so, as the National Health Service (NHS) England (2017, p. 8) note, are that:

> [Patients] can bring unique perspectives and insights into its work, perhaps through their lived experience as a patient/carer or as a member of a community with particular health and care needs. They can challenge thinking, [and] help innovate and improve…

In conceptual terms, '[a]longside evidence-based medicine (EBM), "patient-centredness" may represent one of the major transformative trends within health care in recent times' (Gardner, 2017, p. 240) as a non-patriarchal and inclusive approach. However, beyond policy rhetoric and theory, in practice patients are often: (1) pre-categorised as data sources (Hess, 2015), i.e. for real-world evidence in biomedical research where small disease populations render full clinical trials infeasible (Annemans & Makady, 2020) – their input (via observational data) tabulated as data endpoints in patient registries (Wu *et al.*, 2020); and/or (2) they are recruited as allies for mission-orientated advocacy campaigns over pricing (Mazzucato, 2015) and/or faster drug approvals (Chapman *et al.*, 2020). Bringing together various actors into a single entity to claim legitimation through numbers offers strengthened voice (Rabeharisoa *et al.*, 2014), opening questions over who sits at the centre/periphery of such alliances. When such entities form around rare diseases, they often 'have stronger opportunities to democratise research than patient organisations for more common conditions' (Pinto *et al.*, 2018, p. 124) owing to the smaller size of their patient communities. Thus, questions about how public engagement is carried out around rare diseases, and the (un)evenness of social relations they cohere around become paramount. Elsewhere, medical authorities champion digital resources and technological self-care (Petrakaki *et al.*, 2018), effectively deferring engagement responsibilities onto patients themselves under the guise of patient empowerment.

In contrast to policy rhetoric about public engagement and the ensuing practices of medical authorities, people living with rare diseases (patients) often mobilise and interact with one another by participating in social media community support groups (Ainsworth, 2020; Milne & Ni, 2017; Young & Fujimoto, 2021) to share information or misinformation (Chiang, 2020), resources (Mazanderani *et al.*, 2018), and/or to participate in shared discussion (McKee & Richardson, 2021). Here, social media affords inclusion of 'unruly' publics typically held outside the purview of medical authority including activists and actors (de Saille, 2015b). It enables patients to collaborate and interact directly with advocates, clinicians, researchers, technologists and many others to co-construct and exchange knowledge via relatively horizontally structured networks – albeit often steeped within an uneven set of relations (Tempini & Del Savio, 2019). How discourse forms around rare diseases on particular social media platforms and the role of different actors in its construction is not well understood – in part due to a scarcity of literature and research on the topic.

This paper draws on social network and controversy analyses to examine data on a selection of YouTube videos relevant to 'rare disease' - and those watched immediately before or after. It includes videos users' have purposively selected and ones recommended by YouTube's 'related videos' feature. It addresses questions about: (1) what groups form around rare disease related videos on YouTube - and whether there are any discernible patterns; (2) to what extent YouTube's algorithmic recommendations are formulative of those groups; and (3) what discourses circulate within and between them - and within what sets of relations.

The social media platform specifics of YouTube

Social media communication about health conditions often involves communities forming around influential actors and/or content (Vicari, 2017; Vicari & Cappai, 2016). The affordances of specific social media platforms shape communication within/between those communities (Struck *et al.*, 2018) offering

variable data. Twitter offers publicly open and conversational interaction through 280-character limited posts and private messaging (Bruns, 2012; Twitter, 2021). Data on pre-set fields via an application programming interface (API), enables examination of connections between users and hashtags (Giglietto & Lee, 2017). Facebook users can confine posts of up to 63,208 characters (Bossetta, 2018) to a page, group, pre-approved list of friends, or make it publicly open (Facebook, 2021). Meanwhile, tools like CrowdTangle (Shiffman, 2021) interface with Facebook's API to return various data on those posts. Nuance between different social media platforms' affordances spawns unique communication etiquettes too. Users are 'more uncivil and impolite and less deliberative among strangers on Twitter…than on Facebook' (Oz *et al.*, 2018, p. 3414), raising questions about the specificity of YouTube regarding the types of interaction and data it affords.

As a highly popular video-sharing platform (Covington *et al.*, 2016), YouTube is steeped in cultural participation (Burgess & Green, 2009; Carpentier, 2014; Pires *et al.*, 2021) with users uploading content, sharing others', and/or simply watching videos. It is often engaged within a medical context for information and/or education about particular health conditions (Struck *et al.*, 2018). However, despite being second only to Facebook in popularity - with 2.3 billion active users worldwide as of January 2021 (Statista, 2021) - there has been little research about how rare disease discourse is generated on YouTube, what forms of engagement it fosters, what etiquette(s) it encompasses, and/or the extent to which users' choice of videos is algorithmically shaped.

YouTube users (called subscribers and/or channels) typically 'watch multiple videos during sessions that last about 40 minutes… [where] a viewer might conduct one search, watch a video, and then go on to watch a suggested video' (Jarboe, 2020) and/or one of their own choice. An average video length of 11.7 minutes (Statista, 2019) suggests three or four videos being watched in each sitting. Studying health-related videos watched on YouTube in the UK and USA, Godskesen *et al.* (2021) find this average drops to a mean of 5.7 minutes, suggesting a potential for more videos to be watched in each sitting when it comes to health-based content. Alongside length and number videos watched, the specific videos people actively search for, *and* the ones received as recommendations are important for understanding communication around rare diseases on YouTube. Here, the paper treats videos as the central actor, *not* users, mindful that interactions around each video may take many forms involving both human and non-human actors, e.g., algorithms, hashtags, hyperlinks, and/or bots. On YouTube, comments and replies on videos (between subscribers) can be conversational – albeit through quasi-anonymous and/or private accounts more frequently than on Facebook or Twitter (Park *et al.*, 2015) and limited to few interactions owing to YouTube's cumbersome interface (Murthy & Sharma, 2018). Channels can allow/disallow comments on videos too, providing nuanced levels of control. YouTube also offers less conversational forms of interaction, i.e., liking/disliking, marking a video as favourite, sharing it, and following/unfollowing a channel. As such, the paper examines both the construction of discourse and the role of an algorithm in shaping it, with more weight given to examining the latter.

Understanding YouTube's 'related videos' feature and algorithm

When users search/watch YouTube videos, they receive recommendations on what to watch next via the platform's 'related videos' feature - powered by its recommendation system algorithm. The latter offers a 'codified step-by-step processes implemented by YouTube to afford or restrict visibility' (Bishop, 2019, p. 2589) of videos to particular users. The structure of this proprietary feature is not publicly documented (Airoldi *et al.*, 2016), however, in its technical development Covington *et al.* (2016) recall facing several challenges. The size and scale of YouTube's data rendered some types of algorithms unfeasibly slow. Meanwhile, the speed and frequency of content uploaded to YouTube meant traditional predictive models would be unable to keep pace, posing problems in balancing new and old content. Covington *et al.* note '[h]istorical user behavior on YouTube is inherently difficult to predict due to sparsity and a variety of unobservable external factors' (2016, p. 191), making it difficult to recommend videos reliably from past user activity alone. Turning to deep learning and Google's 'TensorFlow' library, they adopted an approach whereby 'algorithms no longer explicitly specify a decision model, but draw on user feedback to inductively generate such a model' (Rieder *et al.*, 2018, p. 53). Here, the recommendations users encounter as 'related videos' on YouTube are iteratively and continually updated based on their own 'personal activity (watched, favourited, liked videos) as seeds and expanding the set of videos by traversing a co-visitation-based graph of videos' (Davidson *et al.*, 2010, p. 294). This draws on web browser and/or Google account histories, alongside aggregated viewing histories of other users, i.e., what was watched after a particular video, weighted by number of users (Yang *et al.*, 2017). As such, YouTube's related videos feature offers users predictive recommendations that mirror their own past choices (personal and collective).

Framework: Issue-mapping, controversy analysis, and YouTube videos as actors

As a theoretical lens, the paper uses Marres & Moats' (2015) 'issue-networks' - borne from controversy analysis. The former involved mapping hyperlinks between climate change websites to examine how different 'publics engage with debates about technologies…to analyse interactions between the problems and social dynamics shaping [particular] technological controversies' (Waller & Gugganig, 2021, p. 589). Using social network analysis (SNA), Marres and Rogers mapped 'networks composed of heterogeneous set[s] of entities (actors, documents, slogans, imagery) that have [been] configured into [a] hyperlink-network around a common problematic' (2005, pp. 6–7). This involved 'identifying and tracing the associations between actors involved with an issue, and to render them both in narrative and visual form' (Rogers *et al.*, 2015, pp. 9–10). Marres & Moats (2015) later suggested this might resolve a balance between science and technology studies' focus on historical developments of artefacts and media studies' focus

on content and reception, labelling their approach 'controversy analysis'. Here, they combined: the radically constructivist notion of generalised symmetry in actor-network theory (ANT), which treats all actors (human and non-human) as equal and all claims to knowledge as equally valid (Moats, 2019); and social construction of technology (SCOT) studies' focus on the positions of different actors in shaping the trajectory of a technologies' development over time with Marres and Moats (Marres & Moats, 2015). This paper appropriates their approach to focus on the development of public engagement around rare diseases on a social media platform and subsequent construction of discourse rather than focussing solely a specific (material) technology. At the same time, this paper also examines the role of YouTube's related videos feature (algorithm) in framing the construction of discourse, and therefore find inequality amongst actors. Here, the focus on centrality measures and prominence social network analysis (discussed below) marks a shift away from the symmetry found in Marres and Moats' work.

In later works, Marres (2015) classified controversy analyses as those that: (1) typify knowledge claims as legitimate or illegitimate (mapping ontologies); (2) uncover discursive 'relations between substantive arguments and socially and politically located actors...by analyzing which claims and issue terms have support from which actors' (Marres, 2015, p. 663) to see which groups of actors support one another; or (3) start from a (radically) constructivist position of '...making no decisions on the site of study upfront' (Ibid.). This paper takes a discursive approach, treating YouTube videos, users (channels and subscribers), playlists, hashtags, and comments as actors. However, rather than speaking to rare disease debates as a topic that can be reconciled, the paper uses the notion of 'issue-networks' to signify its focus on mapping connections between users and groups (as networks) and the issue(s) constructed through their dialogue. To do so, the paper draws on social network analysis, and measures of centrality to identify particular issue-networks and their influence, as well as the role of specific actors such as YouTube's algorithm. As such, this framework enables the paper to look at which videos are foregrounded (most viewed/most commented), which actors' knowledge claims are dominant in each, and how they connect to one another to construct discourse around rare diseases on YouTube.

Methods

Gathering data for social network and applied thematic analyses

The paper discusses data gathered through 'Data Tools for YouTube' (DTFY) on 19-May-2021. DTFY is a SQL-based tool developed by the University of Amsterdam's Digital Methods Initiative which interacts with YouTube's API (v3.0) through various modules (Reider, 2015b), starting from either a set channel/video or text query. As inclusion criteria, this paper uses DTFY's 'video network' module to gather videos via a text query using "rare disease" as a search term, with no pre-set date range or geolocative parameters. All returned videos have been included in this research. The module 'creates a network of relations between videos, starting from a search… [and a] network of channels based on the same relations…[by] retriev[ing] "related videos" from the *search/list#relatedToVideoId* API endpoint' (Reider, 2015a). It combines two YouTube API elements: (1) the 'list' operator (by keyword) retrieves 'the first 25 search results associated with the keyword… includ[ing] videos, playlists, and channels' (Google, 2021). Setting DTFY to the maximum 10 query iterations increases results to the 250 most-viewed videos with 'rare disease' in their title; and (2) the 'relatedToVideoId' parameter. Setting a crawl depth of 1 returned the 250 most-viewed videos about rare disease (seeds) *and* any associated with them within one generation by crawling from those seeds. Crawl depth depicts the depth to which search is conducted, and thus provides an incremental set of results. For this query, a crawl depth of 0 returned 50 videos, while a crawl depth of 1 returned 250. Results include videos recommended by the 'related videos' feature, within the same playlist, or frequently watched by users immediately before/ after one of the 250 most popular ones. Here, YouTube's algorithm offers a potential source of bias in as far as the process by which it accounts for videos being within the 25 most watch is not documented. Likewise, the use of an English language search term could potentially limit the scope of the research linguistically and/or geographically. As an overall sample, DTFY returned 7,469 nodes (individual videos with unique URLs posted between 28-Jun-2006 and 19-May-2021), 7,167 of which contain "rare disease" in their title. These represent the most watched videos when users search YouTube for rare disease related content. Within the sample, only 396 are 'related videos' recommendations, highlighting that a narrow selection of videos are repeatedly encountered as recommendations by multiple users. As eligibility criteria, all videos returned by DTFY have been included in this research. As an overall sample, DTFY returned 7,469 nodes (individual videos with unique URLs posted between 28-Jun-2006 and 19-May-2021), 7,167 of which contain "rare disease" in their title. These represent the most watched videos when users search YouTube for rare disease related content. Within the sample, only 396 are 'related videos' recommendations, highlighting that a narrow selection of videos are repeatedly encountered as recommendations by multiple users.

DTFY returns data in GDF, a file format suitable for SNA. It includes the video title, URL, channel, date of upload, and video category (assigned by channel owners). It also includes a count of comments, dislikes, favourites, likes, and views for each video. The gathered dataset has 72,927 edges, each representing a connection between two nodes, i.e., where a user watches a video immediately before or after another one. These are directed, meaning one video may be recommended to users (or viewed) more often than others. Nodes with more edges are potentially more influential in shaping discourse. Analysing the data in Gephi (0.9.2), an open-source data visualisation software, enables SNA and representation via social graphs (see below) alongside generation of various statistics on the network. In the SNA, each video is a node while clusters are sets of nodes with more connecting edges than the network average.

Examining content and interactions surrounding videos provides useful insights about how/why particular videos circulate within each cluster. Here, the research incorporates applied

thematic analysis (ATA), a three-stage process of: (1) qualitatively open-coding data manually; (2) then iteratively amalgamating and whittling down codes to a narrowed set; before (3) arriving at a set of conceptual themes (Guest *et al.*, 2014) aligned with the SNA. This involved using NVivo (12) to analyse video content, as well as the textual comments around videos. For the latter, a training set of codes were applied manually before using the autocoding feature to complete the open coding of all 7,167 rare disease related videos. Watching specific videos helped in whittling and narrowing the open codes towards conceptual themes. As a research design, conducting SNA before ATA enables connections between various subclusters to be established as an exploratory issue-mapping exercise before delving into their narratives and the constitution of rare disease discourse across it.

Ethical considerations

The research presented in this paper received approval from the University of Sheffield research ethics committee (reference: 040659) on 14-Jun-2021. The open access dataset this paper draws on uses pseudonyms to preserve YouTube channels, users (subscribers), and video producers - with the exception of public figures when acting in public capacity. These alterations have not distorted scientific meaning.

Results

As noted in the methods section above, this paper draws on sample of 7,469 videos (as nodes) posted on YouTube between 28-Jun-2006 and 19-May-2021, gathered using DTFY on 19-May-2021 (Hanchard, 2021). Within this, 7,167 videos contain the term "rare disease" in their title and 396 are 'related videos', i.e. algorithmically generated recommendations based videos with "rare disease" in their title having been viewed. There are 72,927 edges connecting the nodes. This section examines the data by using all the nodes and edges returned by DTFY (rather than focussing on any subset). Through modularity it finds that there are 54 clusters, within which it identifies eight distinct patterns in the pre-set categories assigned to YouTube videos at upload (see Figure 1). These are: activist, current affairs, educational, follower, entertainment, infotainment, socially concerned, and specific interest.

The section shows that each pattern comprises a particular level of: *focus* on rare disease content; *engagement* between content and viewers, i.e., likes/dislikes and comments surrounding videos*; permeability* of videos between categories; and *repetition* in viewers watching the same video. Across six of the eight patterns, the paper highlights an issue-network that connects clusters/subclusters via specific bridging videos. Some clusters/subclusters revolve around a particular rare disease, others around relevant topics crossing over into separate domains. By looking at content and surrounding interactions within the issue-network, the section reveals three communication strategies at play, each of which fosters particular forms of engagement. Medical terminology and reference to clinical processes provides a 'professional' audience with practical advice for improving practices but lacks space for dialogue. A 'general' audience are presented information on rare diseases for entertainment and/or education, garnering little discussion/engagement - despite the platform affording ample opportunities. Elsewhere, content is used persuasively to market specific drugs, services, or treatments to an 'insider' audience of patients living with a particular rare disease whilst providing useful information. This garners complementary and contradictory comments, through which users form community. As such, the results set out in this section lead to an

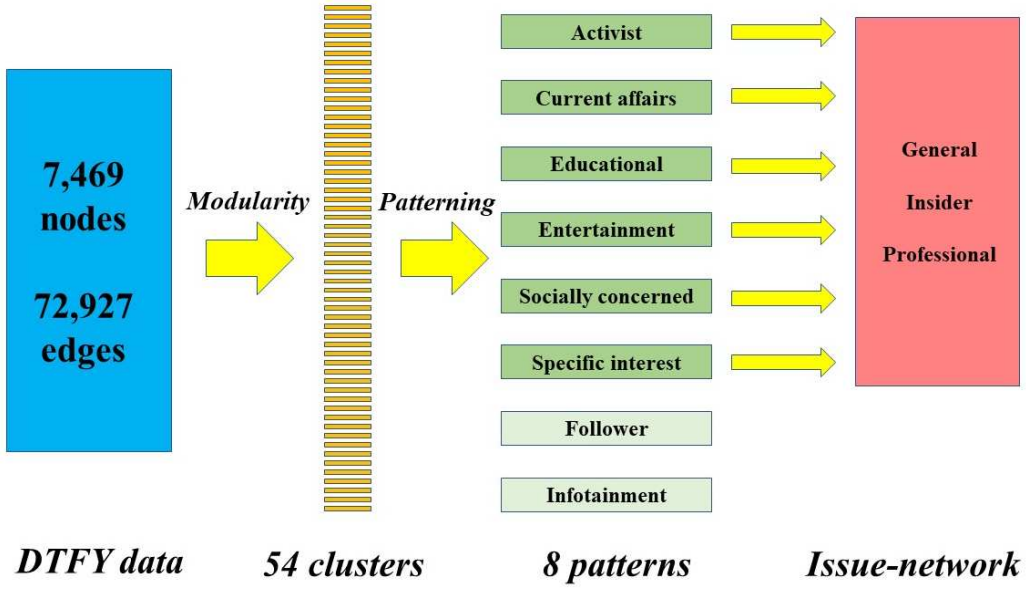

Figure 1. Flow diagram of analysis. DTFY=Data Tools for YouTube.

argument that medical authorities could offer more meaningful forms of public engagement and in turn gain input on specific rare diseases by exploiting the existing potential of social media platforms like YouTube as open spaces for dialogue.

Eight patterns of interaction amongst rare disease related video categories

The overall network is tightly knit (Figure 2) with all videos connected through eight neighbours (average path length 7.99) up to a maximum of 26 (network diameter). The average weighted degree of 9.64 also depicts ~9-10 edges per node. In short, rare disease content viewers watch almost twice as many videos than is typical on YouTube – consonant with (Godskesen *et al*., 2021). Within a 0.76 modularity there are 54 clusters (0 to 53 below) based on video URLs watched together. These range from a giant 843-video cluster to one with 12. Thus, not only do people receive (and watch) recommendations for multiple videos when searching for "rare disease" content, there are similarities (homophily) between the videos they watch. This opens questions about the importance of recommendations and video categories assigned at upload.

Within the 20 most viewed videos, 19 are in the *Music* category (Figure 2; Table 1); a unsurprising predominance given YouTube's status as highly popular platform for music video sharing (Airoldi *et al*., 2016; Allgaier, 2013; Yu & Schroeder, 2018). Whilst this indicates categories are important. *Music* videos also hold a low average clustering coefficient of 0.32 (1 would indicate all are from the same category, 0 that none are).

Likewise, the most and least likely category for videos to be watched together are *Autos & Vehicles* and *Science & Technology,* respectively, holding cluster coefficients of 0.25 and 0.50. In short, although *Music* videos are popular when searching for rare disease content, users often go on to watch videos from other categories. So, although YouTube's pre-set categories are important, they have limited impact on the 54 clusters' formation and/or users' viewing choices. This opens questions about whether there are any patterns in the categories of videos people watch surrounding rare disease, an examination of which (which also attends to crossover between categories) reveals eight distinct patterns (also see Table 2):

> ***(1) Activist*** clusters (16, 18 and 28) have ~1.5 times the network average videos per cluster, with *Nonprofit & Activism* categories featuring strongly - often crossed with *Education* and *People & Blogs*. They hold divergent foci too. Cluster 18, for instance, attracted >6.5 million views of just 796 videos surrounding 'Rare Disease Day' (of various years), accruing a mean of 72,456 likes, 3,273 dislikes, and 3,216 comments per video. Cluster 16, by contrast, garnered only 201,884 views of 51 videos, centring around the charitable work of public figures. Its high cluster coefficient of 0.66 suggests greater homophily amongst videos watched/recommended. However, it attracted only 55 comments per video (mean average) and >20 times more likes than dislikes, pointing at a relatively shallow level of engagement. Overall, 'activist' clusters hold a narrow remit, focussing on videos about famous people and/or educational content, with some category crossover. The latter includes a focus on rare disease slanted towards particular politics. For example, 65 of cluster 28's videos are from the official 'National Organization for Rare Disorders (NORD)' channel (a US-based patient organisation with a strong advocacy focus), holding a mean outdegree of 32.76. Thus, it shows that activism on social media tends to revolve around a 'committed minority amplifying the group position' (Recuero *et al*., 2019, p. 9), marking the pattern as a potential site for advocacy, with clusters 18 and 28 particularly relevant for rare disease.

> ***(2)*** **Current affairs** clusters (0, 4, 5, and 9) contain 551 videos focussed on *News & Politics* and *Entertainment* categories, with minor crossover into others ranging from *Nonprofits & Activism* to *Science and Technology*. Together they received >7 million views with a mean 125,943 likes to 3,100 dislikes per video, with similarly weighted in/out degrees of 6.08 and 6.41 show content to be relatively evenly dispersed between being watched and shared. This suggests a reasonably well-focussed level of engagement, equally notable in the fairly high cluster coefficient of 0.44 and mean 4,257 comments per video, marking the pattern clusters as potential sites for dialogue and exchange. In terms of content, however, the focus sits on news stories and political events worldwide, not rare diseases specifically. Videos on the latter are, instead, interspersed with general health videos alongside personal testimony and/or news on a broad spectrum of topics.

> ***(3) Educational*** clusters (6, 17, 19, 21, 23, 30, 35, 39, 44, 46, 51, and 53) hone-in on informative and educational content, generating a mean average cluster coefficient of 0.47. The 12 pattern clusters collectively hold only 1,757 videos - yet garner over 3.3 billion views, with the same set of videos watched and recommended repeatedly. This is notable in the high mean of views per video (1,883,985.17 for seeds and 1,985,631.76 or non seeds). Here, no particular videos dominate, with educational clusters collectively holding fairly equal weighted in/out degrees (9.94 and 9.93). In terms of engagement with content, the pattern holds 22,973 mean likes to 1,017 dislikes per video, but fewer than 1,624 comments. This varies between clusters, with a general focus on health providing cluster 17 with almost 2.4 times the comments of cluster 23. Whilst the latter focuses narrowly on rare diseases, it includes 32 videos (>10% of the cluster total) from Belgium-based biopharmaceutical company 'UCB'. These range from individual researchers' presentations at Rare Disease Day 2020 to informative clinical studies about how medicines are made. Likewise, cluster 44 is composed entirely of 77 videos on the US 'Food and Drug Administration' (FDA) channel, covering general health and rare disease information alike. In addition, particular channels are central to communication flow between educational clusters as influencers

Figure 2. **Social graph of the full network.**

Table 1. **The 10 most viewed 'rare disease' related YouTube videos.**

Video label	Channel	Views million	Category
Christina Perri - A Thousand Years [Official Music Video]	Christina Perri	1847	Music
Lukas Graham - 7 Years [Official Music Video]	Lukas Graham	1133	Music
Bebe Rexha - Meant to Be (feat. Florida Georgia Line) [Official Music Video]	Bebe Rexha	993	Music
Rick Astley - Never Gonna Give You Up (Video)	RickAstleyVEVO	946	Music
Eagles - Hotel California	Chili World	609	Music
Galantis - No Money (Official Video)	Galantis	590	Music
James Blunt - You re Beautiful (Video)	James Blunt	546	Music
Dua Lipa - Be The One [Official Music Video]	Dua Lipa	507	Music
His Voice Is So Emotional That Even Simon Started To Cry!	Viral Feed	401	Entertainment
Marshmello - Stars [Official Music Video]	Marshmello	396	Music

Table 2. Comparison of clusters and patterns.

Cluster	*Cluster composition % YouTube categories*	*Pattern*	*No. of Videos*	*Avg. weighted In-degree*	*Avg. weighted Out-degree*	*Cluster*
2	Education (10%) Entertainment (17.50%) News & Politics (20%) People & Blogs' (22.50%) Science & Technology (20.20%)	Infotainment	40	2.58	2.58	2
6	Science and technology (50%) Education (21.88%)	Educational	66	5.91	6.18	6
7	Education (19.57%) Entertainment (32.61%) News & Politics (20%) People & Blogs (23.91%)	Infotainment	72	6.81	7.83	7
8	Education' (21.88%) Entertainment (50%) People & Blogs' (14.06%)	Infotainment	64	8.42	8.28	8
10	Education (28.07%) Entertainment (33.33%) People & Blogs' (17.04%)	Infotainment	57	4.89	4.89	10
12	Entertainment (36.36%) Music (25.45%) News & Politics (21.82%)	Entertainment	55	14.49	14.13	12
15	Entertainment' (24.74%) Music (27.84%) Science & Technology (22.68%)	Entertainment	97	8.99	7.4	15
17	Education' (26.71%) Non-profits & Activism (20.28%) Science & Technology (36.75%)	Educational	498	9.82	9.28	17
19	Education (24.92%) Non-profits & Activism (17.51%) Science & Technology (36.36%)	Educational	297	8.3	8.61	19
29	News & Politics (100%)	Specific interest	50	20.04	20.12	29
21	Education (53.33%) Entertainment (16.77%)	Educational	30	2.37	2.93	21
23	Education (29.28%) Non-profits & Activism (12.88%) People & Blogs (20.72%) Science & Technology (23.03%)	Educational	304	9.24	9.59	23
25	Entertainment (77.55%)	Entertainment	49	10.78	10.86	25
27	Entertainment' (66.67%)	Entertainment	51	3.94	3.86	27
30	Education (31.44%) People & Blogs (40.55%)	Educational	217	11	11.09	30
35	Education (16.97%) People & Blogs (40.98%)	Educational	61	7.87	8.18	35
39	Education (72.00%) Entertainment (12.00%)	Educational	25	14	14.12	39
44	Education (37.66%) Science & Technology (62.34%)	Educational	77	21.44	21.01	44
46	Science & Technology (75.56%)	Educational	45	17.4	16.87	46

Cluster	*Cluster composition % YouTube categories*	*Pattern*	*No. of Videos*	*Avg. weighted In-degree*	*Avg. weighted Out-degree*	*Cluster*
49	Education (16.85%) News & Politics (16.30%) Non-profits & Activism (14.13%) People & Blogs (27.17%)	Socially concerned	184	8.8	9.31	49
50	Education (24.63%) People & Blogs (15.67%)	Educational	134	13.72	14.49	50
51	Education (18.42%) Science & Technology (57.89%)	Educational	38	12.16	12.24	51
52	Education (48.31%) Non-profits & Activism (15.73%) People & Blogs (20.79%)	Educational	178	15.43	14.97	52
53	Education (35.35%) Film & Animation (19.19%) People & Blogs (14.14%) Science & Technology (13.13%)	Educational	99	7.28	7.76	53
Cluster	***Cluster composition % YouTube categories***	***Pattern***	***No. Videos***	***Avg. weighted In-degree***	***Avg. weighted Out-degree***	***Cluster***
2	Education (10%) Entertainment (17.50%) News & Politics (20%) People & Blogs' (22.50%) Science & Technology (20.20%)	Infotainment	40	2.58	2.58	2
6	Science and technology (50%) Education (21.88%)	Educational	66	5.91	6.18	6
7	Education (19.57%) Entertainment (32.61%) News & Politics (20%) People & Blogs (23.91%)	Infotainment	72	6.81	7.83	7
8	Education' (21.88%) Entertainment (50%) People & Blogs' (14.06%)	Infotainment	64	8.42	8.28	8
10	Education (28.07%) Entertainment (33.33%) People & Blogs' (17.04%)	Infotainment	57	4.89	4.89	10
12	Entertainment (36.36%) Music (25.45%) News & Politics (21.82%)	Entertainment	55	14.49	14.13	12
15	Entertainment' (24.74%) Music (27.84%) Science & Technology (22.68%)	Entertainment	97	8.99	7.4	15
17	Education' (26.71%) Non-profits & Activism (20.28%) Science & Technology (36.75%)	Educational	498	9.82	9.28	17
19	Education (24.92%) Non-profits & Activism (17.51%) Science & Technology (36.36%)	Educational	297	8.3	8.61	19
29	News & Politics (100%)	Specific interest	50	20.04	20.12	29
21	Education (53.33%) Entertainment (16.77%)	Educational	30	2.37	2.93	21
23	Education (29.28%) Non-profits & Activism (12.88%) People & Blogs (20.72%) Science & Technology (23.03%)	Educational	304	9.24	9.59	23

Cluster	*Cluster composition % YouTube categories*	*Pattern*	*No. of Videos*	*Avg. weighted In-degree*	*Avg. weighted Out-degree*	*Cluster*
25	Entertainment (77.55%)	Entertainment	49	10.78	10.86	25
27	Entertainment' (66.67%)	Entertainment	51	3.94	3.86	27
30	Education (31.44%) People & Blogs (40.55%)	Educational	217	11	11.09	30
35	Education (16.97%) People & Blogs (40.98%)	Educational	61	7.87	8.18	35
39	Education (72.00%) Entertainment (12.00%)	Educational	25	14	14.12	39
44	Education (37.66%) Science & Technology (62.34%)	Educational	77	21.44	21.01	44
46	Science & Technology (75.56%)	Educational	45	17.4	16.87	46
49	Education (16.85%) News & Politics (16.30%) Non-profits & Activism (14.13%) People & Blogs (27.17%)	Socially concerned	184	8.8	9.31	49
50	Education (24.63%) People & Blogs (15.67%)	Educational	134	13.72	14.49	50
51	Education (18.42%) Science & Technology (57.89%)	Educational	38	12.16	12.24	51
52	Education (48.31%) Non-profits & Activism (15.73%) People & Blogs (20.79%)	Educational	178	15.43	14.97	52
53	Education (35.35%) Film & Animation (19.19%) People & Blogs (14.14%) Science & Technology (13.13%)	Educational	99	7.28	7.76	53
Cluster	***Cluster composition % YouTube categories***	***Pattern***	***No Videos***	***Avg. weighted In-degree***	***Avg. weighted Out-degree***	***Cluster***
2	Education (10%) Entertainment (17.50%) News & Politics (20%) People & Blogs' (22.50%) Science & Technology (20.20%)	Infotainment	40	2.58	2.58	2
6	Science and technology (50%) Education (21.88%)	Educational	66	5.91	6.18	6
7	Education (19.57%) Entertainment (32.61%) News & Politics (20%) People & Blogs (23.91%)	Infotainment	72	6.81	7.83	7
8	Education' (21.88%) Entertainment (50%) People & Blogs' (14.06%)	Infotainment	64	8.42	8.28	8

(cf. Vicari, 2017). This is pronounced in cluster 30, where a vlog by 'Chronically Jenni' with 26 videos (>12% of the cluster total) covers individual chronic health conditions in an educational manner. Overall, the cluster pattern shows that rare disease discourse is not always advocacy-orientated or found only within the most popular clusters.

(4) Entertainment clusters (12, 15, 25, and 27) revolve around YouTube's *Entertainment* and *Music* categories, with *Relaxing Sleep Music = Deep Sleeping Music Relaxing Music Stress Relief Meditation Music (Flying)* on the 'Soothing Relaxation' channel its most watched video (272 million views since 03-Jul-2016). There is crossover between categories, despite the high average cluster coefficient of 0.53. In part, this is due to repeat viewing - where 252 videos have been collectively watched ~7.7 million times. In engagement terms, entertainment pattern clusters garner 53,304 mean likes to 3,516 dislikes per video and 3,823 comments. This suggests people tend not to generate participate in discussion around its videos, watching and occasionally 'liking' them instead – as reflected in the high weighted average in-degree than out-degree (9.55 to 9.06). By extension, in cluster 15 'EURORDIS's channel (a major figure in European rare disease advocacy) has a video labelled *Rare Diseases: we are 30 million in Europe. Each of us is a real person*. Despite receiving 17,259 views since 12-Nov-2012, it has only 4 comments, 82 likes, and 1 dislike - highlighting a very shallow level of engagement. This is typical across entertainment pattern clusters where an average eigenvector centrality of 0.04 depicts only moderate potential for its videos to shape the construction of discourse. Channels within entertainment clusters often blur boundaries with others too when dealing with rare disease related content. 'Sanofi Genzyme Europe' (the European wing of a large US-based biotechnology company), for example, posted a video of their awareness-raising flashmob in Amsterdam at the end of Rare Disease Day 2011. Although the video sits within an entertainment cluster, it overlaps with information and infotainment pattern ones through recommendations and viewing choices. Thus, it highlights a fluidity in videos and channels moving between clusters - and with it, permeability between pattern boundaries.

(5) Follower clusters (13 and 22) revolve around influencers and/or public figures, with little crossover to other types; notable in high average cluster coefficients of 0.67 for both clusters. This follows across all 101 videos, gathering a mean average of 5,346 likes to 91 dislikes per video, and 434 comments. However, when combined with fairly equal weighted in and out degrees (19.25 and 19.94) these figures depict a very shallow level of engagement. The pattern also contains only one video directly related to rare disease - *A Trip to DC for RARE Disease Day | RAGE REGARDLESS RY !!!* on 'The Wharton family' channel (a US-based social media influencer), gathering 94,760 views but only 274 comments. As such, the pattern revolves around people following particularly influential channels, encompassing videos on a myriad of topics, albeit with little or no direct focus on rare disease content.

(6) Infotainment clusters (2, 7, 8, and 10) sit between educational and entertainment patterns in terms of content, with a middling-to-average cluster coefficient of 0.48 and 2.9 million views of 223 videos. Despite the high viewing figures, infotainment videos only gather a mean average of 65,323 likes to 3,544 dislikes and 4,157 comments, meaning repeat viewing features strongly and discussion with others is limited. Likewise, the relatively even average weighted in and out degrees (means 6.06 and 6.33) show no single type of video dominating. Popular videos like *Relaxing Jazz Music - Background Chill Out Music - Music For Relax Study Work*' on the 'Cafe Music BGM' channel and *NBA RARE Moves Part 1* on 'Kawhi Not' respectively gained ~58 million and >16 million views (since 18-Feb-2017 and 30-Oct-2017), highlighting the popularity of *Sports* and *Music* videos - and their crossover. As an outlier, cluster 8 holds some rare disease related videos. These, however, tend to be interspersed with general health-related ones, garnering little direct interaction - with almost half on the 'Autoimmune Hepatitis' channel, covering a broad range of related concerns. One video (posted 08-Mar-2021), for instance, questioningly titled *Is the Johnson & Johnson Vaccine Safe for AIH Patients?* gathers only 81 likes, 3 dislikes, and no comments. Similarly, another explanatory expert presentation (posted 19-Aug-2019), asks rhetorically *Who Done It? How Did I Catch Autoimmune Hepatitis?*, gaining only 2 comments over 1,482 views. Both highlight limited dialogue around infotainment videos.

(7) Socially concerned clusters (1, 3, 26, 31, 32, 38, 41, 45, and 49) had >3.6 million views of 2,281 videos, showing high repeat viewing. Meanwhile, means of 24,564 to 916 dislikes, and 2,787 comments per video, pointed to low engagement further supported by fairly equal yet relatively high average in and out degrees (10.03 and 9.09). Paired with a mildly low average cluster coefficient of 0.39, these figures depict socially concerned clusters as less focussed on video categories than other patterns. Although social concerned clusters had little impact on rare disease discourse, cluster 1 and 26 were outliers. With an average cluster coefficient of 0.28 and thus greater crossover with other categories, cluster 1 covered variable content including 19 videos on the 'VASCERN ERN Rare Vascular Diseases' channel and 10 on 'EveryLife Foundation for Rare Diseases', making it a hub for rare disease content albeit interspersed with children's cartoons and music videos. Meanwhile, cluster 26 held content on specific rare diseases, including 32 videos on 'Sanofi Genzyme' (a biotech company) with translated counterparts on 'Sanofi Genzyme Europe' e.g., *Celebrating Pompe disease* translated into Ukrainian and Russian. As such, they

highlight potential for a subset of clusters within the socially concerned pattern to influence rare disease discourse, even if only to a limited extent. As such, cluster patterns are shown to have variable levels of internal consistency.

(8) Specific interest clusters (11, 14, 20, 24, 29, 33, 34, 36, 37, 40, 42, 43, 47, 48, 50 and 52) focus tightly on particular topics - reflected by a high average cluster coefficient of 0.57. This rises to 0.67 and 0.70 for clusters 29 and 34, which both revolve primarily around a single video category (*News & Politics* and *Science & Technology*). Overall, specific interest cluster videos average >6.9 million mean views per video, accruing 106,192 likes to 4,341 dislikes, and 5,517 comments (<20 per video), suggesting shallow engagement. This follows into category crossover and video content too; a limited range of specific interest videos cover rare diseases and tend to be interspersed with general health matters. As outliers, clusters 47, 50, and 52 focus specifically on rare diseases. Cluster 47 has 72 videos centred the *Nonprofits & Activism* category, with 42 on the 'EURORDIS' channel, depicting the cluster as potential locus for rare disease advocacy. Cluster 50 has 134 videos split evenly between *Entertainment, People & Blogs*, and *Science & Technology* categories. While 42 of its videos are on the 'Rare Disease Report' channel, others are on equally informative channels targeting patients and clinicians, i.e. 'The Mayo Clinic' and 'VJHemOnc - Video Journal of Hematological Oncology'. However, paired with 92 comments per video (mean), the cluster holds only limited potential to influence rare disease discourse. As such, specific interest clusters focus on narrow topic areas, with some holding a higher-than-average number of videos on the channels of medical authorities, industry bodies, or advocacy groups. While the latter often hold a focus on rare diseases, they remain limited as sites of discourse generation.

In summary, rare disease related YouTube videos cluster around pre-set categories in eight patterns. Each pattern revolves around different levels of: (1) *focus* - on one or more categories (from tight to loose) - measured by averaged cluster coefficients; (2) *engagement* - with content and discussion, measured as counts of likes/dislikes and comments (from shallow to in-depth); (3) *permeability* - between video category boundaries and between clusters; and (4) *repetition* – (high to low) measured as a count of users repeatedly viewing the same videos. Here, YouTube's 'related videos' feature is shown to have only a limited impact video choice. Instead, specific clusters contribute towards the construction of rare disease discourse in a particular way, raising questions about how different actor engage which other around rare disease content and the role of medical authorities.

Three rare disease audiences constructed through communication strategies

This section examines video content and surrounding comments in rare disease relevant clusters (identified above) to identify a YouTube rare disease issue-network (Figure 3). It finds three communication strategies at play, each aimed at a particular set of audiences and associated with a specific form of engagement (general, insider, and professional).

Videos from rare disease relevant clusters attract comparatively little interaction compared to those elsewhere and are often interspersed with videos from other topics with variable levels of focus. In cluster 18, for instance, the ten most viewed and commented videos (averaging 326 million and 115,457 per video respectively) both primarily comprise *Music* category ones (Figure 4a and Figure 4b). Meanwhile, the ten least viewed and commented largely comprise rare disease related content, gathering <20 mean views per video - often paired with few comments (six received none).

Despite their overall unpopularity, rare disease videos garner high levels of focus and engagement within some clusters – bound together through particular videos. For example, in cluster 1, the video *1. Europe and Rare Disease by Prof Germano* acts as a bridging node. It is part of subcluster catering to a 'professional' audience of clinicians (Figure 5a). Other videos in the subcluster offer information about dealing medically with specific conditions (i.e., *What is aortic root replacement and when is it indicated?)* and technical guidance (i.e., *Downloading Dicom studies into Desktop*). While these three videos are all on the 'VASCERN ERN – Rare Vascular Diseases' channel, the same bridging node connects these videos to a French-language subcluster (also within cluster 1) with various videos and channels surrounding a Canadian conference about the evolutionary psychology of Bernard de Montréal. It also ties both subclusters to another containing informative videos about various rare diseases, i.e., Marfan Syndrome and Klippel-Trenaunay syndrome. As such, it illustrates a permeability of bridging nodes (videos) in moving between categories to connect different subclusters into a network.

This permeability is not only found within clusters, but across them too. Here, one cluster 18 video on the 'National Organization for Rare Disorders (NORD)' channel labelled *Living with Cornelia de Lange Syndrome (CdLS)* has 12,347 views but only 5 comments. With a betweenness centrality of 232,403 (versus mean averages of 13,648 and 18,815 across all rare disease related clusters) it bridges to various clusters/subclusters. For example, it connects with a specific rare disease related subcluster of informative CdLS-based videos such as *Cornelia de Lange Syndrome (CdLS) Awareness Video* (on the 'Cornelia de Lange Syndrome' channel) and *CDLS The Rollercoaster Ride* on 'Andrew Borge' (channel of former CdLS Foundation UK and Ireland board member and trustee) - both in cluster 18 (Figure 5b). It also ties these to a cluster 28 subcluster aimed at patients and patient groups, such as one covering the 2013 NORD webinar labelled *Patient Registries: What They Are and How to Start One* and another labelled *Advocating For A Rare Disease Advisory Council In Your State* (in cluster 28). These connections between subclusters pass via *Living with Cornelia de Lange Syndrome* where a single video bridges between subclusters across two activist clusters (18 and 28). Thus, bridging node videos connect the issue-network by

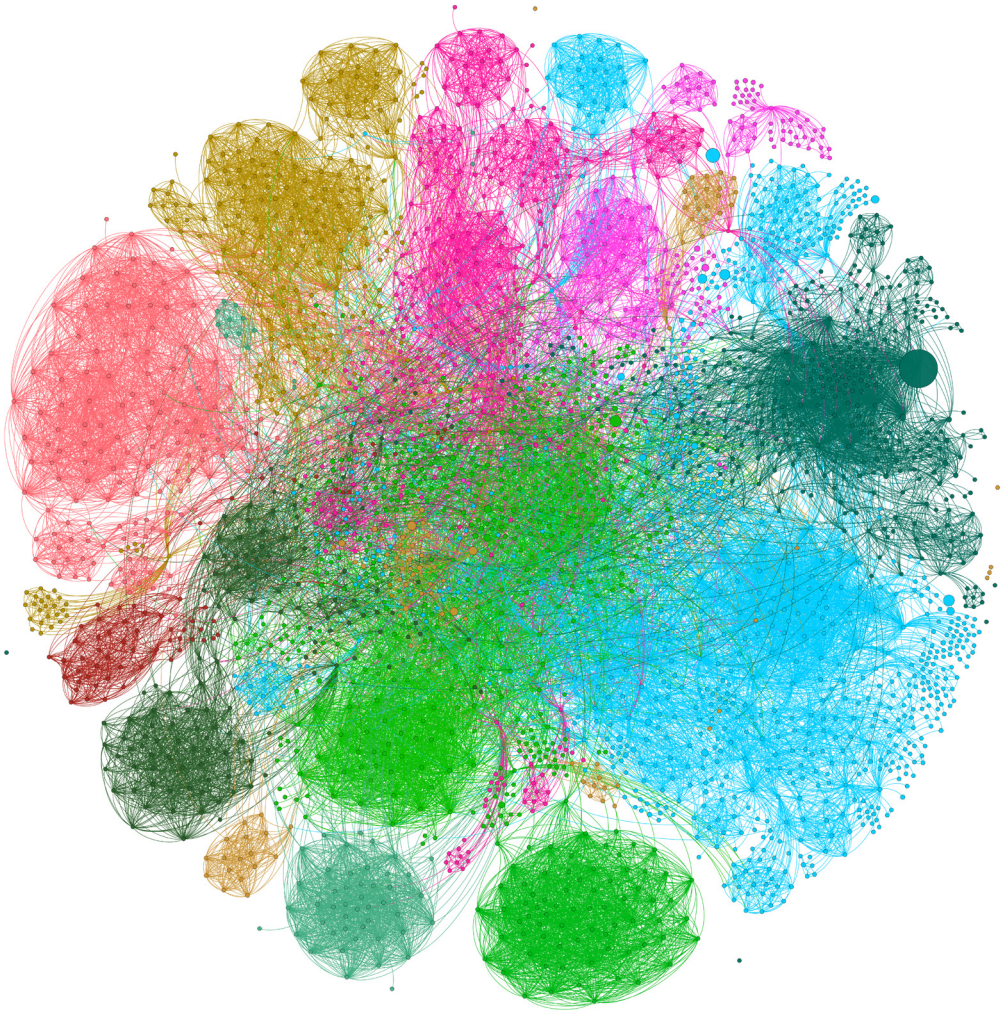

Figure 3. Social graph filtered to rare disease relevant clusters.

bringing together various rare-disease related subclusters and clusters

In terms of discourse, comments surrounding rare disease videos involve three discernible communication strategies. For instance, in cluster 50, one subcluster revolves around amyloidosis (AL), a condition where amyloid proteins build up in the body, causing organs and tissues not to work properly (NHS, 2020). Its videos are informative and aimed towards a general audience with an interest in the disease. However, as is typical of health videos on YouTube (cf. Gardner *et al.*, 2019) it includes persuasive marketing. For example, one video on 'The Mayo Clinic' channel labelled *Amyloidosis: What you need to know - Mayo Clinic* offers a 4.5-minute video of Dr. Morie Gertz giving a simplified expert explanation of AL; what it is, the main symptoms, effects, and current types of treatment. The video says getting AL diagnosed can be difficult as a patient, before extolling the professional expertise abut AL at the Mayo Clinic. Thus, the video markets a commercial service under the guise of providing public information. Comments surrounding the video express variable levels of engagement with its content. At times, one-way feedback and appraisal of the videos' informational value sit at the fore. For example one users remarks that the video provides a 'clear introduction to amyloidosis [in] less than 4.5 min[utes] given by an expert with more than 30 years of experience.' The same video also connects patients by acting as forum for sharing advice and experiences. For example, *Bushra K* asks: 'Has anyone found [a] cure for amyloidosis on the skin?'. To which, the following replies ranged from expressions of support and empathy (practices of community-building) to suggestions for a specific medicine to use (information sharing), through to offering a specific clinician's details (as form of resource sharing):

> ***Franfran7904:*** Do you have it on your skin? I do. I don't know anyone else with it. It's so frustrating and depressing. If you have it maybe we can email and talk about our experiences and compare treatments etc

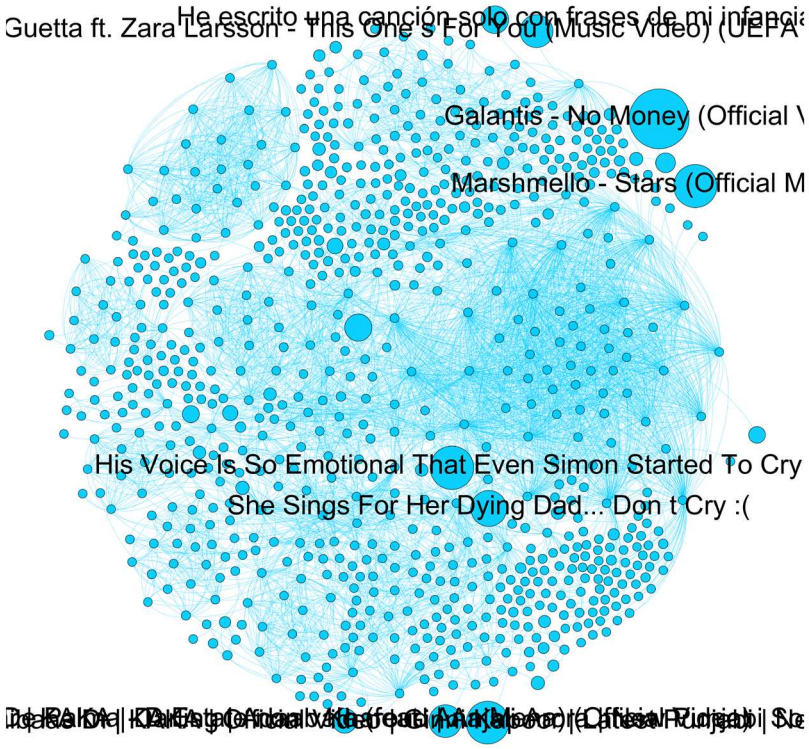

Figure 4a. **The ten most viewed videos around rare disease.**

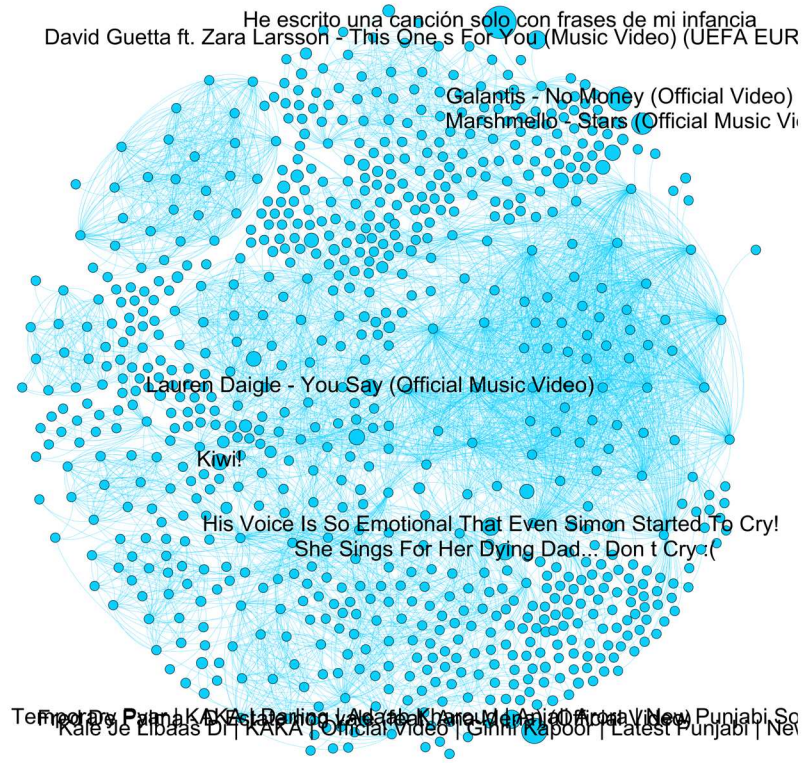

Figure 4b. **The ten most commented on videos around rare disease.**

Figure 5a. Example of nodes bridging within a subclusters.

> ***Wooof:*** i have on skin from eczema i hope one day its cured so depressing and upsetting
>
> ***Tracey Uberdown:*** +Franfran7902 I do. It really is terrible. sigh
>
> ***juan garcia:*** Bushra K.. a Proteolytic Enzyme: Which will Consume Amyloid Plaques!!best Brand: Serretia .. 👍
>
> ***Films Seek:*** i have same problem on my skin
>
> ***Aaradhya Paatni:*** Me too have the same problem on my skin..
>
> It's so irritating ☹
>
> ***Grace Driver - keow:*** That's so sad to hear you are still suffering from this same situation . Am feeling for you right now, I know of a doctor who can help you get rid of this. He also help me from this same situation , He can also help cure yours permanently

As their conversation illustrates, the comment-space surrounding YouTube videos provides opportunity for people to share experiences, information, and resources as part of an 'insider' communication strategy of patients co-constructing a community around shared knowledges of living with a rare disease.

Rather than the accounts of living a rare disease being isolated to an 'insider' strategy, it occurs frequently across other clusters with less interactive audiences too. In the 'educational' pattern, for instance, cluster 44 is composed entirely of 'FDA' channel videos (Figure 5c). One subcluster presents patient testimony of day-to-day life with specific rare diseases, e.g., *Chris Carroll s Rare Disease Story* of living with type 2D Limb-girdle Muscular Dystrophy (causing limb and muscle deterioration throughout the body) and/or *Nancy Rose Spector's Rare Disease Story* of living with Von Hippel Lindau (VHL) syndrome (causing cysts and tumour development in multiple organs). Inline with the FDA's broader strategy of bringing together different actors for better dialogue and engagement (Bauer, 2017), the videos are curated to suit a partisan audience of clinicians, patients, policymakers, and patients. Nancy Rose, for example, describes her personal medical history and lived experience before espousing the importance of advocacy

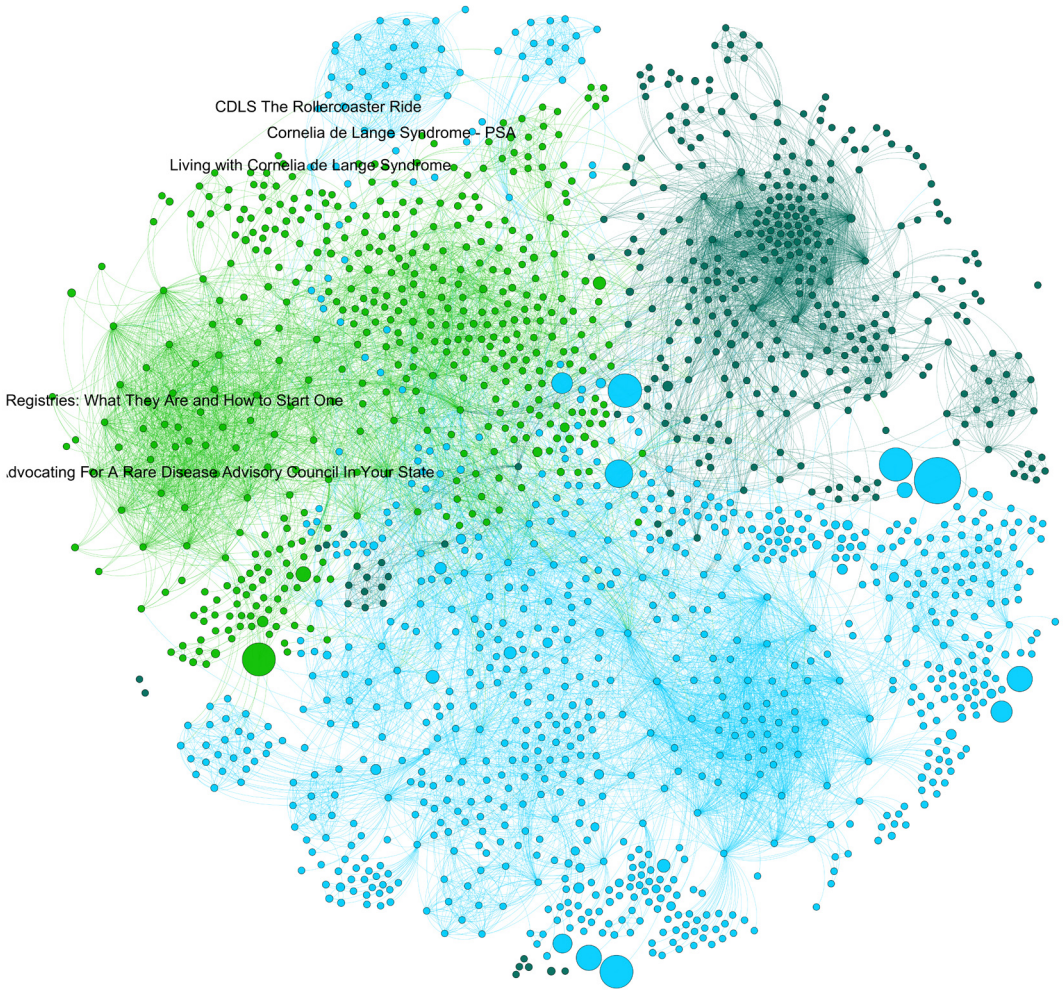

Figure 5b. Example of nodes bridging across subclusters.

for better access to care and treatment. However, despite an overarching policy rhetoric of public engagement and the FDAs formalised stance of being patient-inclusive (Bauer, 2017), they have adopted a communication strategy aimed towards a 'general' audience, composed of various publics rather than seeking to engage with 'insider audiences'. This is notable where the FDA have disabled comments on their videos, curtailing a potentially fruitful space for dialogue and exchange. Instead, YouTube is treated as a one-way means of outputting information, serving to legitimate the FDA's authority as arbiter of public knowledge about rare diseases. The same approach is taken up by other medical authorities too, as noted above in the discussion of cluster 1 videos around *What is aortic root replacement and when is it indicated?* and *Downloading Dicom studies into Desktop.* Here, a professional audience of clinicians can watch videos on YouTube but are provided no space for shared discussion; comments are again disabled.

Elsewhere medical authorities are more open. For example, one cluster 23 subcluster hosts informative videos about covering Friedreich's Ataxia (a progressive rare disease-causing nervous system damage)attracting a generally interested lay audience and/or patients with limited levels of engagement. Meanwhile another connected subcluster caters to a professional audience of clinicians with videos on rehabilitation around rare and non-rare diseases (Figure 5c), i.e. a Webinar*: Paroxysmal dyskinesias: update on clinical and genetic features by Giovanna Zorz*i and *Webinar: Hereditary Spastic Paraplegia (HSP) clinical disease course by Rebecca Schüle* - both on the European Reference Network 'ERN-RND' channel. Whilst 'ERN-RND' allows comments, its videos have received only one each and no replies. As such, they follow other educational cluster pattern videos in being treated by viewers as one-way information sources rather than sites of exchange despite the platform offering a space for potential dialogue between actors.

Discussion and conclusion

There are patterns combining the YouTube categories pre-assigned to videos at upload, that emerge within use (through viewing). Some categories are highly significant, i.e., *Music*,

Figure 5c. Example of nodes bridging between clusters.

accounts for over half of all YouTube videos about 'rare disease'. Rather than standalone categories bearing strongly on rare disease discourse, across 54 modularity clusters they are combined within eight distinct patterns (activist, current affairs, educational, entertainment, follower, infotainment, social, and specific interest) – with six relevant for rare disease discourse. Each pattern is defined by its constituent clusters' specific levels of: *focus* on rare-disease content in the videos each of their constituent clusters contain; *engagement* they garner from users in terms of likes and comments; *repetition* with which their videos are watched; and *permeability* in videos moving between patterns (as bridging nodes). Within these patterns, YouTube's 'related videos' feature is of limited importance in shaping video choice.

Examining connections between rare disease relevant clusters revealed an issue-network between across the six patterns, with three communications strategies - each steeped in a particular type of engagement: (1) a *general* one in which rare disease videos are typically watched for entertainment, infotainment, or education alongside other topics, with little engagement between actors or with content; (2) a *professional* one, with medical authorities targeting clinicians with a narrow set of repeatedly watched videos focussed on technical guidance and

advice, offering few spaces for dialogue or exchange – and thus limiting depth of engagement; and (3) an *insider* one where videos offer personal testimony, information, and education about living with a rare disease - often with underlying marketing or promotion of products/services - aimed at patients, their friends, family members, and carers. Here, viewers generate self-organised communities by sharing discussion, information, and resources in comment spaces around particular videos. Together, the three communication strategies portray an issue-network around YouTube rare disease videos in which actors are afforded few spaces to engage with one another or directly with medical authorities. It is worth noting, however, that a key limitation of the paper is that it relies on an English-language only search term and therefore different patterns may be found when applying the same method to a similar search in other languages.

Together, the three communication strategies portray an issue-network around YouTube rare disease videos in which actors are afforded few spaces to engage with one another or directly with medical authorities. Relating this to policy rhetoric reveals a disparity between medical authorities stated aims around public engagement and their actions on YouTube. Here, potential for collaboration and exhange of knowledges between clinicians, patients, and key organsiations - where patient experiences might be collated and discussed - is closed down. Instead, a top-down model is invoked as a means for medical authorities to re-legitimate their own position. Elsewhere, open spaces for dialogue and exchnage are not fully exploited, and instead sit latent amidst shallow levels of enagement with video content. As a recommendation, this paper suggests that key institutions could foster more meaningful forms of public engagement around rare disease by identifying and targeting videos/channels that bridge between insider and professional audiences, and by actively engaging with comments and discussion (and opening space for it) around those videos. Here, issue-mapping provides a useful way to identify relevant videos and channels. The paper has shown what patterns to look for and what criteria make videos relevant for rare disease discourse. As such, it contributes an understanding of how discourse around rare disease is constructed on YouTube as well as pointing to a way policymakers and key institutions might foster more inclusive public engagement with rare disease patients.

Data availability

Underlying data

Figshare (University of Sheffield ORDA): Orphan drugs – Dataset 2: YouTube rare disease issue-networks. https://doi.org/10.15131/shef.data.16855507 (Hanchard, 2021).

This project contains the following underlying data:

- Orphan Drugs - Dataset 2 - YouTube rare disease issue-network - Edges.csv
- Orphan Drugs - Dataset 2 - YouTube rare disease issue-network - Nodes - updated.csv

Acknowledgements

Thanks go to my colleagues Dr. Jin Ding and Dr. Eva Hilberg for constructive feedback on the narrative of an early draft, and project Principal Investigator Prof. Paul Martin for constructive feedback/comments and for allowing me time and space to write the paper. Also, to the Gephi community on Facebook for their technical advice, tips, and support.

References

Ainsworth C: **Rare diseases band together toward change in research.** *Nat Med.* 2020; **26**(10): 1496–1499.
PubMed Abstract | Publisher Full Text

Airoldi M, Beraldo D, Gandini A: **Follow the algorithm: An exploratory investigation of music on YouTube.** *Poetics.* 2016; **57**: 1–13.
Publisher Full Text

Allgaier J: **On the Shoulders of YouTube: Science in Music Videos.** *Sci Commun.* 2013; **35**(2): 266–275.
Publisher Full Text

Annemans L, Makady A: **TRUST4RD: tool for reducing uncertainties in the evidence generation for specialised treatments for rare diseases.** *Orphanet J Rare Dis.* 2020; **15**(1): 127.
PubMed Abstract | Publisher Full Text | Free Full Text

Bauer L: **Supporting Rare Disease Drug Development: CDER's Rare Diseases Program: 2017 Roadmap for Engaging with the Center for Drug Evaluation and Research.** 2017.
Reference Source

Bishop S: **Managing visibility on YouTube through algorithmic gossip.** *Soc Media Soc.* 2019; **21**(11–12): 2589–2606.
Publisher Full Text

Bossetta M: **The Digital Architectures of Social Media: Comparing Political Campaigning on Facebook, Twitter, Instagram, and Snapchat in the 2016 U.S. Election.** *Journal Mass Commun Q.* 2018; **95**(2): 471–496.
Publisher Full Text

Bruns A: **How long is a Tweet? Mapping dynamic conversation networks on *Twitter* using GAWK and Gephi.** *Inform Commun Soc.* 2012; **15**(9): 1323–1351.
Publisher Full Text

Burgess J, Green J: **YouTube: Online video and participatory culture.** Cambridge: Polity, 2009.
Reference Source

Carpentier N: **'Fuck the clowns from Grease!!' Fantasies of participation and agency in the YouTube comments on a Cypriot Problem documentary.** *Inform Commun Soc.* 2014; **17**(8): 1001–1016.
Publisher Full Text

Chapman CR, Eckman J, Bateman-House AS: **Oversight of Right-to-Try and Expanded Access Requests for Off-Trial Access to Investigational Drugs.** *Ethics Hum Res.* 2020; **42**(1): 2–13.
PubMed Abstract | Publisher Full Text

Chiang AL: **Social media and medicine.** *Nat Rev Gastroenterol Hepatol.* 2020; **17**(5): 256–257.
PubMed Abstract | Publisher Full Text

Côté A, Keating B: **What is wrong with orphan drug policies?** *Value Health.*

2012; **15**(8): 1185–1191.
PubMed Abstract | **Publisher Full Text**

Covington P, Adams J, Sargin E: **Deep Neural Networks for YouTube Recommendations.** *Proceedings of the 10th ACM Conference on Recommender Systems.* 2016; 191–198.
Publisher Full Text

Davidson J, Liebald B, Liu J, *et al.*: **The YouTube Video Recommendation System.** *Proceedings of the 4th ACM Conference on Recommender Systems.* 2010; 293–296.
Publisher Full Text

de Saille S: **Dis-inviting the Unruly Public.** *Science as Culture.* 2015a; **24**(1): 99–107.
Publisher Full Text

de Saille S: **Innovating innovation policy: the emergence of 'Responsible Research and Innovation'.** *J Responsible Innov.* 2015b; **2**(2): 152–168.
Publisher Full Text

EMA: **EMA/313148/2020: Meeting summary-PCWP/HCPWP joint meeting (Minutes).** 2020.
Reference Source

Facebook: **Control who can see what you share.** 2021.
Reference Source

Gardner J: **Patient-centred medicine and the broad clinical gaze: Measuring outcomes in paediatric deep brain stimulation.** *BioSocieties.* 2017; **12**: 239–256.
Publisher Full Text

Gardner J, Warren N, Allison C, *et al.*: **Persuasive bodies: Testimonies of deep brain stimulation and Parkinson's on YouTube.** *Soc Sci Med.* 2019; **222**: 44–51.
PubMed Abstract | **Publisher Full Text**

Giglietto F, Lee Y: **A Hashtag Worth a Thousand Words: Discursive Strategies Around #JeNeSuisPasCharlie After the 2015 *Charlie Hebdo* Shooting.** *Soc Media Soc.* 2017; **3**(1): 1–15.
Publisher Full Text

Godskesen T, Holm SF, Höglund AT, *et al.*: **YouTube as a source of information on clinical trials for paediatric cancer.** *Inf Commun Soc.* 2021.
Publisher Full Text

Google: **YouTube Data API**. 2021.
Reference Source

Grande SW, Faber MJ, Durand MA, *et al.*: **A classification model of patient engagement methods and assessment of their feasibility in real-world settings.** *Patient Educ Couns.* 2014; **95**(2): 281–287.
PubMed Abstract | **Publisher Full Text**

Guest G, Macqueen KM, Namey EE: **Planning and Preparing the Analysis.** In: Applied Thematic Analysis Learning Objectives. In Guest, G., Macqueen, K., and Namey, E. (Eds.) *Applied Thematic Analysis.* London: Sage; 2014; 21–48.

Hanchard M: **Orphan Drugs - Dataset 2 - YouTube rare disease issue-network.** The University of Sheffield. Dataset. 2021.
http://www.doi.org/10.15131/shef.data.16855507.v2

Hess DJ: **Publics as Threats? Integrating Science and Technology Studies and Social Movement Studies.** *Science as Culture.* 2015; **24**(1): 69–82.
Publisher Full Text

HM Govt: **Genome UK: the future of healthcare**. 2020.
Reference Source

Jarboe G: **How YouTube Generates and Ranks Suggested Videos.** *Search Engine Journal.* 2020.
Reference Source

Marres N: **Why Map Issues? On Controversy Analysis as a Digital Method.** *Sci Technol Human Values.* 2015; **40**(5): 655–686.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Marres N, Moats D: **Mapping Controversies with Social Media: The Case for Symmetry.** *Soc Media Soc.* 2015; **1**(2): 1–17.
Publisher Full Text

Marres N, Rogers R: **Recipe for tracing the fate of issues and their publics on the Web.** In B. Latour and P. Weibel (Eds.). *Making Things Public: Atmospheres of Democracy.* Cambridge: MIT Press, 2005; 922–935.
Reference Source

Mazanderani F, Kelly J, Ducey A: **From embodied risk to embodying hope: Therapeutic experimentation and experiential information sharing in a contested intervention for Multiple Sclerosis.** *BioSocieties.* 2018; **13**(1): 232–254.
Publisher Full Text

Mazzucato M: **Building the Entrepreneurial State: a new framework for envisioning and evaluating mission-oriented public investments.** (No. 384 - Levy Economics Institute Working Paper Collection). 2015.
Reference Source

McKee K, Richardson K: **Improving access and awareness of rare disease trials through patient and registries.** World Orphan Drug Congress USA, 2021.
Reference Source

Mikami K: **Orphans in the Market: The History of Orphan Drug Policy.** *Soc Hist Med.* 2019; **32**(3): 609–630.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Milne CP, Ni W: **The Use of Social Media in Orphan Drug Development.** *Clin Ther.* 2017; **39**(11): 2173–2180.
PubMed Abstract | **Publisher Full Text**

Moats D: **From media technologies to mediated events: a different settlement between media studies and science and technology studies.** *Inf Commun Soc.* 2019; **22**(8): 1165–1180.
Publisher Full Text

Murthy D, Sharma S: **Visualizing YouTube's comment space: online hostility as a networked phenomena.** *New Media Soc.* 2018; **21**(1): 191–213.
Publisher Full Text

NHS: **Amyloidosis.** NHS, 2020.
Reference Source

NHS England: **NHS England Patient and Public Voice Partners Policy**. 2017.
Reference Source

Oz M, Zheng P, Chen GM: **Twitter versus Facebook: Comparing incivility, impoliteness, and deliberative attributes.** *New Media Soc.* 2018; **20**(9): 3400–3419.
Publisher Full Text

Park SJ, Lim YS, Park HW: **Comparing Twitter and YouTube networks in information diffusion: The case of the "Occupy Wall Street" movement.** *Technol Forecast Soc Change.* 2015; **95**: 208–217.
Publisher Full Text

Petrakaki D, Hilberg E, Waring J: **Between empowerment and self-discipline: Governing patients' conduct through technological self-care.** *Soc Sci Med.* 2018; **213**: 146–153.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Pinto D, Martin D, Scolari C: **Chasing cures: Rewards and risks for rare disease patient organisations involved in research.** *BioSocieties.* 2018; **13**(1): 123–147.
Publisher Full Text

Pires F, Masanet MJ, Scolari CA: **What are teens doing with YouTube? Practices, uses and metaphors of the most popular audio-visual platform.** *Information, Communication, and Society.* 2021; **24**(9): 1175–1191.
Publisher Full Text

Rabeharisoa V, Callon M, Filipe AM, *et al.*: **From 'politics of numbers' to 'politics of singularisation': Patients' activism and engagement in research on rare diseases in France and Portugal.** *BioSocieties.* 2014; **9**(2): 194–217.
Publisher Full Text

Recuero R, Zago G, Soares F: **Using Social Network Analysis and Social Capital to Identify User Roles on Polarized Political Conversations on Twitter.** *Soc Media Soc.* 2019; **5**(2): 1–18.
Publisher Full Text

Reider B: **Introducing YouTube Data Tools**. 2015a.
Reference Source

Reider B: **YouTube Data Tools.** (Version 1.22). 2015b.
Reference Source

Rieder B, Matamoros-Fernández A, Coromina Ò: **From ranking algorithms to 'ranking cultures': Investigating the modulation of visibility in YouTube search results** *Convergence.* 2018; **24**(1): 50–68.
Publisher Full Text

Rogers R, Sánchez-Querubín N, Kil A: **Issue Mapping for an Ageing Europe.** Amsterdam: University of Amsterdam Press, 2015.
Reference Source

Shiffman N: **CrowdTangle for Academics and Researchers**. 2021.
Reference Source

Statista: **Average YouTube video length as of December 2018, by category**. 2019.
Reference Source

Statista: **Most used social media 2021**. 2021.
Reference Source

Struck J, Siegel F, Kramer MW, *et al.*: **Substantial utilization of Facebook, Twitter, YouTube, and Instagram in the prostate cancer community.** *World J Urol.* 2018; **36**(8): 1241–1246.
PubMed Abstract | **Publisher Full Text**

Tempini N, Del Savio L: **Digital orphans: Data closure and openness in patient-powered networks.** *BioSocieties.* 2019; **14**(2): 205–227.
Publisher Full Text

Twitter: **Counting characters**. 2021.
Reference Source

Vicari S: **Twitter and Non-Elites: Interpreting Power Dynamics in the Life Story of the (#)BRCA Twitter Stream.** *Soc Media Soc.* 2017; **3**(3): 2056305117733224.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Vicari S, Cappai F: **Health activism and the logic of connective action. A case study of rare disease patient organisations.** *Inf Commun Soc.* 2016; **19**(11): 1653–1671.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Waller L, Gugganig M: **Re-visioning public engagement with emerging technology: A digital methods experiment on 'vertical farming'.** *Public Underst Sci.* 2021; **30**(5): 588–604.
PubMed Abstract | **Publisher Full Text**

Wu J, Wang C, Toh S, *et al.*: **Use of real-world evidence in regulatory decisions for rare diseases in the United States-Current status and future directions.**

Pharmacoepidemiol Drug Saf. 2020; **29**(10): 1213–1218.
PubMed Abstract | **Publisher Full Text**

Yáñez-Muñoz RJ: **10 Years of Rare Disease Day.** *Gene Ther.* 2017; **24**(2): 67.
PubMed Abstract | **Publisher Full Text**

Yang G, Gong NZ, Cai Y: **Fake Co-visitation Injection Attacks to Recommender Systems.** In D. Choffnes (Ed.), *Network and Distributed System Security (NDSS) Symposium Proceedings.* 2017; 1–15.
Publisher Full Text

Young LE, Fujimoto K: **The Co-evolution of online social networks and syphilis incidence among young black men who have sex with men.** *Soc Sci Med.* 2021; **272**: 113764.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Yu H, Schroeder S: **Distribution and Popularity Patterns of Chinese Music on YouTube: A Case Study of Local Music's Representation on a Global Internet Platform.** *J New Music Res.* 2018; **47**(1): 68–77.
Publisher Full Text

Open Peer Review

Current Peer Review Status: ?

Version 1

Reviewer Report 24 February 2022

https://doi.org/10.21956/wellcomeopenres.19154.r47779

?

Narelle Warren
School of Social Sciences, Faculty of Arts, Monash University, Melbourne, Vic, Australia

Thank you for a very interesting paper, that shifts from a focus on users to focus on the videos themselves as agentic and as the focus of analysis. This shift in lens potentially offers particular insights in terms of discourse construction that are useful for people working directly with patients and also for other social media researchers. In reading the paper, the discourse construction is only a small part of the analysis, with greater attention given to the algorithms and what they reveal. It may perhaps benefit the paper to make these two different goals - and the different types of data generated in achieving them - explicit at the start.

The 'framework' section (p.4) is quite brief, especially given the importance of the approach to the analysis itself, and could be strengthened through outlining more specifically how each of controversy analysis and social network analysis is conceptualised for this paper, and how they fit together for the operationalization of the analysis. I suspect that quite a lot of information relevant to this comment is already provided but is obscured by the logic of the paragraph/section.

Explain what is a crawl depth (p5).

A lot of information is provided on DTFY but little on the qualitative analysis. Please provide more detail on the content and interaction analysis. Were videos analysed for their content? How many videos were subject to this analysis? Can any further information be provided about these? Having this level of detail is important given the focus on discourse construction and the concern with disparities in the information for different users.

Under 'ethical considerations', were any attempts made to seek consent from any video producers? Samuel and Buchannan's 2020 Guest editorial on Ethical Issues in Social Media Research (doi: 10.1177/1556264619901215)[1] may be useful to expand this section, and more fully explore the ethical implications of YouTube-based research.

The network analysis was interesting and includes a lot of detailed data. It may be useful for a

reader - and aid in keeping track of all of the different clusters - to include a table of the different clusters identified and how each of those contributes to the different patterns. The relevance and insights generated by analysing surrounding videos, for example, seem unclear - this may be emphasised in the revisions.

The three engagement categories offer great insights to policy makers and also people working in the broad field of engaging publics. These are elucidated in the discussion, which indicates the value of this research.

References

1. Samuel G, Buchanan E: Guest Editorial: Ethical Issues in Social Media Research. *Journal of Empirical Research on Human Research Ethics*. 2020; **15** (1-2): 3-11 Publisher Full Text

Is the work clearly and accurately presented and does it cite the current literature?
Yes

Is the study design appropriate and is the work technically sound?
Yes

Are sufficient details of methods and analysis provided to allow replication by others?
Yes

If applicable, is the statistical analysis and its interpretation appropriate?
Not applicable

Are all the source data underlying the results available to ensure full reproducibility?
Yes

Are the conclusions drawn adequately supported by the results?
Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: qualitative research, patient communities, patient communication, health sociology, medical anthropology

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Author Response 30 Mar 2022

Matthew Hanchard, University of Sheffield, Sheffield, UK

Thank you for your comments. I have addressed each below, point-by-point, and have detailed any changes made to the text, which is now far stronger for your input. I have submitted revised version including the changes below.

Reviewer 1 - Point 1
Thank you for a very interesting paper, that shifts from a focus on users to focus on the videos themselves as agentic and as the focus of analysis. This shift in lens potentially offers particular insights in terms of discourse construction that are useful for people working directly with patients and also for other social media researchers. In reading the paper, the discourse construction is only a small part of the analysis, with greater attention given to the algorithms and what they reveal. It may perhaps benefit the paper to make these two different goals - and the different types of data generated in achieving them - explicit at the start.

Response to: Reviewer 1 - Point 1
Likewise, thank you for the kind comments. I have now added a sentence to the end of Introduction section to clarify these as two separate goals. The text reads: “As such, the paper examines both the construction of discourse and the role of an algorithm in shaping it, with more weight given to examining the latter

---

Reviewer 1 - Point 2
The 'framework' section (p.4) is quite brief, especially given the importance of the approach to the analysis itself, and could be strengthened through outlining more specifically how each of controversy analysis and social network analysis is conceptualised for this paper, and how they fit together for the operationalization of the analysis. I suspect that quite a lot of information relevant to this comment is already provided but is obscured by the logic of the paragraph/section.

Response to: Reviewer 1 - Point 2
I have redrafted the Framework section, adding more detail in the second and third paragraphs.

Reviewer 1 - Point 3
Explain what is a crawl depth (p5).

Response to: Reviewer 1 - Point 3
I have now added a sentence to clarify what crawl depth means to t methods section. T reads: “Crawl depth depicts the depth to which search is conducted, and thus provides an incremental set of results. For this query, a crawl depth of 0 returned 50 videos, while a crawl depth of 1 returned 250”

Reviewer 1 - Point 4
A lot of information is provided on DTFY but little on the qualitative analysis. Please provide more detail on the content and interaction analysis. Were videos analysed for their content? How many videos were subject to this analysis? Can any further information be provided about these? Having this level of detail is important given the focus on discourse construction and the concern with disparities in the information for different users.

Response to: Reviewer 1 - Point 4

I have now added further detail on the qualitative research into the Methods section (in the last paragraph).

Reviewer 1 - Point 5

Under 'ethical considerations', were any attempts made to seek consent from any video producers? Samuel and Buchannan's 2020 Guest editorial on Ethical Issues in Social Media Research (doi: 10.1177/1556264619901215)[1] may be useful to expand this section, and more fully explore the ethical implications of YouTube-based research.

Response to: Reviewer 1 - Point 5

I have now included producers within the ethical considerations section. They should have been included in the first version. The suggested paper will also be useful for an upcoming one I ma writing based on Twitter data around rare diseases. My thanks to the reviewer.

Reviewer 1 - Point 6

The network analysis was interesting and includes a lot of detailed data. It may be useful for a reader - and aid in keeping track of all of the different clusters - to include a table of the different clusters identified and how each of those contributes to the different patterns. The relevance and insights generated by analysing surrounding videos, for example, seem unclear - this may be emphasised in the revisions.

Response to: Reviewer 1 - Point 6

I have created a table (Table 2) and submitted it as part of the revised paper.

Competing Interests: No competing interests were disclosed.