

Journal EXIT-Deutschland. Zeitschrift für Deradikalisierung und demokratische Kultur

Issue: 2/2014

The YouTube-Born Terrorist

Lucas Regnér¹

ABSTRACT

Personalization permeates the World Wide Web today and search engine-, social networking-, and social media websites—like Google, Facebook, and YouTube—use algorithms to tailor search results and content to web users' interest and past behavior on the Web. Author and Internet activist Eli Pariser has raised concerns about this trend and coined the term "the filter bubble" to describe the information silos he argues web users may find themselves in when browsing the Web. If Pariser's theory holds true—that personalization algorithms presents self-similar content to web users based on the users' past web behavior—what type of online world will users that consume far-right or far-left radical content find themselves in?

In this research project, I develop a methodology that studies how personalization algorithms affect YouTube users' experience of the website. I apply the methodology to the case of far-right and far-left radical web users, and I quantitatively study the videos they encounter when YouTube's personalization algorithms govern their content discovery.

Borrowing theoretical and methodological frameworks from Internet studies, political socialization, social network analysis, and communications studies, I find that the users do experience significant personalization on YouTube, while it's unclear to what extent personalization furthers radicalization processes.

Although the findings are somewhat inconclusive, the methodology provides an opportunity to systematically study a users web experience. A scaled-up version of the method could yield more decisive findings. However, considering the dynamic nature of the web, it is important that the results are considered in their temporal contexts.

¹ Georgetown University (Washington, DC, USA) | Department of Communication, Culture & Technology



Journal EXIT-Deutschland. Zeitschrift für Deradikalisierung und demokratische Kultur

Issue: 2/2014

INTRODUCTION

On July 22, 2011, Norwegian terrorist and mass murderer Anders Behring Breivik killed 77 people. The attacks were politically motivated, and the rationale behind them was clearly outlined in his manifesto 2083: A European Declaration of Independence. In this document, Breivik referenced no fewer than 525 URL links, creating an intricate web of sources that made up his totalitarian ideology.

This highlights that while Breivik might have acted alone, he is far from isolated in his farright extremist ideology—in fact, he is highly connected to a global network of far-right extremist thought. The novelty, which Breivik's case emphasizes, is that the network lives and exists online. Researchers have looked at Breivik's use of the Internet—the World Wide Web, social networking sites, as well as social media and gaming—in order to understand what impact it might have had on his radicalization. Ravndal (2013) concluded that the Internet and social media "provided Breivik with an alternative reality in which he could cultivate his radical views largely uncontested … it allowed him to isolate himself from the real world" (pp. 182).

Breivik is an extreme case—not everyone who consumes far-right and violent content online commit acts of terror, or even start a pub brawl. But what's interesting is the duality, which seems rather contradictory; how can you be both "isolated" and simultaneously "highly connected?" In the case of Breivik, it is the fact that he is highly connected to radical ideas—to one school of thought, "largely uncontested" (Ravndal, 2013)—that becomes his isolation from the real world. In plain English, Breivik lived in an online bubble.

A seemingly unrelated yet highly relevant story runs parallel to the one of Breivik, and it might partly explain the concept of online bubbles. According to author Eli Pariser, this story started on December 4, 2009, with Google's announcement of tweaks it made to its search engine algorithm. Advertised through the web service company's corporate blog, the post's title read "Personalized search for everyone" (Pariser, 2011: pp. 1). Pariser argued that personalization fundamentally changes how users experience the Web, and he coined the term the filter bubble. His book—the Filter Bubble — what the Internet is hiding from you—elaborates on the theory of how personalization of the web impacts individuals' worldviews.



ISSN: 2196-8136

Issue: 2/2014

Personalized search algorithms allegedly impact what content individual web users are presented with, creating a "content bubble." The concept of personalization partly relies on tracking cookies—pieces of code stored in your web browser, which communicate information about your online activities to websites you visit—as well as your participation on social networking sites. Based on the digital footprints left behind when browsing the Web, combined with other factors like location and time, a website might personalize its content based on the user's past web behavior: although different users visit the same website or search for the same term they might see different content. Personalization is applied to anything from Google's search results, an online newspaper's "suggested articles" section, and YouTube's "videos recommended for you" section; in fact, YouTube is the online platform I will focus on in this study. According to Pariser, "these engines create a unique universe of information for each of us [...] which fundamentally alters the way we encounter ideas and information" (Pariser, 2011: pp. 9).

While the filter bubble may seem rather harmless, Pariser raises many concerns about its effect on individuals and the fabric of society. Imagine a personalized web that is based on (and consequently reinforces) previous online behavior. It presents you only with content that these personalization engines predict you will like. What types of information would a far-right radical like Breivik be presented with? The concept of a filter bubble aligns closely with the idea of Breivik's simultaneous isolation and connectivity. He was isolated in a filter bubble, which connected him to far-right and violence promoting content from all over the world.

The filter bubble—especially in contexts in which it creates negative reinforcement—has recently become subject to empirical testing. Is it possible that a few web-clicks in the wrong direction can lead you into a tunnel of violent far-right extremist content on the Web? According to a paper on YouTube authored by researchers at University College Dublin (UCD), this might be the case:

The influence of related rankings on click through rate, coupled with the fact that the YouTube channels in this analysis originated from links posted by extreme right Twitter accounts, would suggest that it is possible for a user to be immersed in this content



ISSN: 2196-8136

Issue: 2/2014

following a short series of clicks [my emphasis] (O'Callaghan et al, 2013: pp. 9)

This is the very nut that the paper you're currently reading is trying to crack: is the filter bubble real? If so, can it potentially expose users to content about a particular ideology, in situations where a democratic society would prefer that the ideology was countered—such as with hate speech and extremist ideologies?

For this project, I developed a method to study a user's experience of social media sites—or as I will call it, content-bearing social networking sites (content-bearing SNSs). I then applied the method to YouTube—the world's largest video-sharing site—to study the presence of filter bubbles. I measured the prevalence of violent and far-right radical videos from the perspective of two types of politically extreme YouTube users. The research question I asked was "What characterizes the network structure of violent YouTube videos that promote farright radicalism, as experienced by a user?"

This paper is divided in four sections: first, it elaborates on the ideas of personalization of, and radicalization through, the Web. Second, I will introduce the research question, study design, and methodology employed in this study, followed by the results. Finally, I will discuss the findings and propose further research.



Issue: 2/2014

SECTION I: Background and Theory

In this section, I will outline the theories that informed my research and, more importantly, explain how they relate to each other in this interdisciplinary study. I will touch upon the concept of social networking sites and social media, personalization of the Web, and theories on political socialization and radicalization. I focus on connecting the overarching themes: how do the filter bubble, social networking and media sites, and the process of radicalization connect conceptually?

We may roughly regard the layering of the theories as follows: personalization of the Web and its potential to create "filter bubbles"—is a rather new concept and therefore lacks strong empirical underpinning. That is this study's purpose: to find methods to empirically explore, document, and measure personalization of the Web to test Pariser's theory. The political radicalization process—especially as it relates to hate speech—served as a case for my study; it is a meaningful vehicle I use to study the theory of the filter bubble. Lastly, social networking and media sites are platforms where a considerable amount of personalization occurs; consequently, I conducted the study on YouTube. In short, I study the case of exposure to hate speech by empirically measuring the presence of filter bubbles on social media sites.

Social Networking Sites (SNS) and Social Media

Boyd & Ellison (2007) defined social networking sites (SNS) as a web-based service in which users can construct a profile and create connections to other users within the service. According to this definition, the first major social networking site started as early as 1996, with the launch of Six Degrees.com. As Boyd and Ellison noted, SNSs cater to different user groups and have different focuses, ranging from professional networks to networks centered on hobbies. Common for SNSs are these features: leaving private or public messages to other users, upload and create content, and to react to content by "commenting" on or "liking" it (see picture below for example from the SNS Facebook). These are all features that



Issue: 2/2014

allow the users of the website to participate, in contrast to merely consuming whatever is on the page. This trend—the proliferation of opportunities to engage with content on the Web—is part of Web 2.0, and includes a slew of features that allow users to create and publish content (e.g. write text, upload images and videos) (O'Reilly, 2007).



carried out research on the SNS YouTube, which was launched in 2005 as an independent video sharing website. The users of the website can upload videos, which they can share with their connections or the public. 17-26 year olds dominate YouTube's user-base and, perhaps as a result, Entertainment and Music are the most commons categories for content. Additionally, videos on YouTube are most likely between 3-5 minutes long (or under 200 seconds) (Cheng et al, 2008; Santos et al, 2007). Google acquired YouTube in 2006 and the website has since been integrated with other Google products. Nowadays, you authenticate



Issue: 2/2014

your identity on YouTube by using your Google account, that is to say the same account you use for Google's other services such as Gmail.

When a user uploads a video to YouTube, she provides it with a title, text description, as well as marks it with a predefined category. She also has the option to add "tags"—words or phrases that she wants the video to be associated with—so that her videos come up when other users search those terms in YouTube's search box. The "networking" part of YouTube works a little differently than traditional SNSs: a user who has uploaded one or more videos are called a "channel" in reference to traditional TV broadcast. Users can "subscribe" to channels, which is a one-directional "connection," essentially meaning that the subscriber is notified when the channel uploads new videos. YouTube's engagement functionality focuses on users reacting to videos. When watching a YouTube video, a user may "Like" or "Dislike" the video through clicking "thumb up" or "down". She may also comment on the content in a comments area under the video player, as well as subscribe or unsubscribe to the channel from which the video was uploaded.

Other web services that are similar to YouTube are Soundcloud (centered on audio), Flickr (centered on images), and Tumblr (focused on blogging/text). Traditionally, YouTube and web-services like it are often called "social media," indicative of the interactions taking place between users around the content hosted on these sites. However, I chose to be more explicit and refer to these web services as "Content-bearing Social Networking Sites." This is appropriate considering how personalization on these sites work. Networking sites that host content, like YouTube, make retrieval in the vast sea of content possible for the user in mainly two ways: categorization [1] and correlation [2].

Categories [1] or tags, as defined by the user who uploaded the content, help classify and determine what content is grouped together. Categorization examples include: a predefined YouTube category like "news & politics", key words like "Election 2012" or "Huskies", or content uploaded in the same channel. Correlation [2] refers to how a user's consumption of Content "X" may be indicative of her enjoyment Content "Y", based on the fact that other users who consumed Content "X" tend to like Content "Y". For example, one YouTube video called "Husky Cuddle" is categorized under the "Pets and Animals" category



ISSN: 2196-8136

Issue: 2/2014

and tagged with "Cute Huskies" and "Cuddle". Another YouTube video, titled "Alaskan Huskies," appears under the YouTube category "Education" and is tagged with the words "Animal Welfare" and "Alaska." The two videos would not be associated with each other since they are categorized differently. However, YouTube notices that users who watched "Husky Cuddle" also tend to watch "Alaskan Huskies"—the videos "correlate". Therefore, the "Husky Cuddle" video will appear as a "recommendation" when a user watches "Alaskan Huskies". It is through correlation that users have the power to associate seemingly unrelated content. This feature is common in content-bearing SNSs, and it is a crucial component in web personalization.2

The Web guru, Tim O'Reilly, noted that a core competence of Web 2.0 companies was the "control over unique, hard-to-recreate data sources that get richer as more people use them" (O'Reilly, 2007). It's control over these data sources—the aggregated data about the crowd of users and their behavior—that to a large extent has enabled web personalization, which I will describe in the next section.

Personalization of the Web (which can result in the filter bubbles)

So what is the purpose of web personalization? "The ultimate goal of any user-adaptive system is to provide users with what they need without them asking for it explicitly," is Bamshad Mobasher's opening line in his paper Data Mining For Web Personalization (2007: pp. 90). Since Mobasher's paper, this approach to the Web has become industry standard, according to Pariser:

"The future of the web is about personalization", [Yahoo Vice President Tapan Bhat says], "It's about weaving the web together in a new way that is smart and personalized for the user" (2011: pp. 8)

The concept Bhat refers to "implies the delivery of dynamic content [on the Web], such as textual elements, links, advertisement, product recommendations, etc., that are tailored to ... a particular user" (Mobasher, 2007: pp. 90). In reality, this can mean anything ranging from YouTube recommending you videos to the fact that the Amazon product that you

² This is a synthesis of how YouTube's algorithms work based on papers authored by Google employees; see Davidson et al (2010) and Baluja et al (2008).



Issue: 2/2014

placed in your virtual cart now starts showing up on ads across other websites you visit. Personalization is an obscure process that is by design hidden from the user. According to Mobasher, we should distinguish between hidden "automatic personalization" and "customization":

What separates these two notions is who controls the creation of user profiles [...] In customization, the users are in control of (often manually) specifying their preferences or requirements [...] Automatic personalization, on the other hand, implies that the user profiles are created, and potentially updated, automatically by the system with minimal explicit control by the user. Examples of automatic personalization in commercial systems include Amazon.com's personalized recommendations (Mobasher, 2007: pp. 90)

This is the benefit of personalization: to receive valid suggestions about ideas, information, or products that you might find useful. It's all done by gathering as much information about the user as possible and running the information through a prediction algorithm to foresee the user's future behavior or needs. It is a sort of "historical stereotyping" that may be more or less accurate depending on the particular algorithm. Therefore, the ultimate goal for an algorithm engineer is:

[D]ata mining approach to personalization ... leverages all available information about users of the Web site to deliver a personal experience (Mobasher, 2007: pp. 91).

It is this "personal experience" that worries Pariser. Web personalization creates a unique world for each individual user, eliminating the Web's function as a platform for shared meaning through universal content. "The new generation of Internet filters," Pariser explained in his book, "looks at the things you seem to like—the actual things you've done, or the things people like you like—and tries to extrapolate, [which] fundamentally alters the way we encounter ideas and information" (2011: pp. 9). Using the anecdote of two progressive female friends living in the Northeast of the U.S., Pariser recounted how Google Search returns vastly different results even though the women input the exact same search query. Presumably, this has to do with Google's assumption about what these women are interested in based on their past web behavior. In Pariser's anecdote, the search term was BP—oftentimes referred to as British Petroleum. Google presented one of Pariser's friends



Issue: 2/2014

with financial investment information regarding BP and the other with news about the 2010 BP oil spill in the Mexican gulf.

Continuing along this train of thought, Pariser discussed how personalization of the Web might perpetuate information isolation through "information bias—a tendency to believe things that reinforce our existing view, [or] to see what we want to see" (Pariser, 2011: pp. 86). He described how a "filtered environment could have [negative] consequences for curiosity" (Pariser, 2011: pp.90) as well as the social fabric of society. As long as a "large majority of online content reaches" users through search engines like Google or social networking sites like Facebook (2011: pp. 144), we are in the hands of companies that lack a code of ethics in this particular area (2011: pp. 176).

Pariser's assumptions regarding the consequences of the personalized web relies heavily on the validity and generalizability of that one anecdote regarding his female friends in the Northeast of the United States. If web personalization bluntly hides highly relevant information for certain users—and creates truly detached realities—then the consequences for all Google searches, Amazon product and YouTube video recommendations, and Facebook feeds may have the grave effect on the individual and society Pariser described. However, the issue of moving away from anecdotal data to actual empirical measures of the personalized web, and its impact on the users, remains.

However, some have researchers have attempted to measure this phenomenon. Indian researchers Majumder and Shrivastava (2013), in their work, assumed "no knowledge of the personalization algorithms and history of users maintained by them, and [worked] by comparing the personalized and vanilla content served by [the Google Search Engine]" (pp. 873). They developed a method aimed at understanding what impacts personalization, and provided an "example evidence of personalization happening on a user's account ... [and through the proposed method] the user can therefore see not just his inferred interests ... but also how it affects his results" (pp. 882). The study concluded that personalization of Google Search on the PageRank level is real. Considering that Google search users tend to follow the top suggestions, personalization may have a real impact for the user.

Perhaps even more comprehensive is Hannak et al.'s study (2013), Measuring



Issue: 2/2014

personalization of web search. The research questions for the study read:

First, what user features influence Google's search personalization algorithms? ... [To] what extent does search personalization actually affect search results? ... If the delta between "normal" and "personalized" results is small, then concerns over the Filter Bubble effect may be misguided. (pp. 529)

The researchers developed a methodology that takes a comprehensive look at different aspects of personalization that Google and other search engines can potentially use to personalize search results. This includes Tracking Cookies, Geo-location, Gender and Age as documented in a given Google account, as well as Gender, Age, Income, Education and Ethnicity as inferred through browsing, search, and click history. "At a high-level, our methodology is to execute carefully controlled queries on Google Search to identify what user features trigger personalization" (pp. 529), they wrote. The results:

We applied our methodology to real Google accounts recruited from AMT and observe that 11.7% of search results show differences due to personalization. Using artificially created accounts, we observe that measurable personalization is caused by 1) being logged in to Google and 2) making requests from different geographic areas (pp. 536).

As Hannak et al's study is the first of its kind, it's difficult to say much about it. For example, how much is 11% difference due to personalization? Is it enough to cause concern? Hannak et al's findings and Majumder & Shrivastava's research informed my study's design. Now, we will look at the radicalization process and how it connects with consumption of web content. Political Radicalization on the Web

In his Manifesto, Breivik justified his violence with political motives—he had no personal disputes with the people he killed. He showcases how political, ideological, and religious identity can play a powerful role in individuals' violent behavior and reasoning behind carrying out acts of terror (Kydd & Walter, 2002: pp. 264). If political socialization is the process in which mainstream political thought is transmitted from one generation to the other, political radicalization is when that transmission fails. Somewhere in the process of attaining a political identity, an individual searched for, or was exposed to, ideologies outside of the scope of the mainstream. These can be either far-right ideologies (e.g. neo-



Issue: 2/2014

nazism and fascism, extreme nationalism, racism) or far-left ideologies (e.g. militant communism, like Trotskyists or Maoists; Anarchist; Environmentalism; or left-authoritarianism). The question is how does an individual become radicalized? There are many theories as to why people radicalize and what external factors impact an individual's radicalization process. In contrast to traditional political socialization, there is not one dominant narrative that seems to effectively span over all radicalization processes. This makes sense, since radicalization by definition is a deviant behavior. Scholars have found that in some instances radical views are in fact passed down via family; at other times, individuals from non-radical families will be attracted to radical groups' sense of association and identity (Greenberg, 1970: chapter 6).

Norwegian author and journalist Åsne Seierstad, who compiled documentation and interviews regarding Breivik, argued in her book that the massacre at Utøya was connected to Breivik's complicated childhood (Seierstad, 2013). However, Breivik himself justified his actions claiming self-defense against a phenomenon he called "the islamization of Europe". According to Breivik, his violent and non-democratic means of achieving a political goal was justified by the threat of Islam—or rather, the perceived threat of Islam.

This justification of violent behavior may in fact have support in the literature. A study published in the Journal of Conflict Resolution looks at Jewish Israelis' attitudes towards Palestinian citizens of Israel after being exposed to acts of terror. The study examined "the process that might lead individuals who are exposed to terrorism to support the denial of social and political rights [for a] minority group" that the terrorists associate with (Canetti-Nisim et al, 2009: pp. 380). The study concludes that:

... many individuals who were exposed to terrorism did not become more extreme in their exclusionist attitudes ... the findings suggest that exposure to terrorism will lead to exclusionist political attitudes principally through the multiple mediators of psychological distress and perceived threat. In other words, individuals who are exposed to terrorism may become more exclusionist particularly when they experience psychological distress, which feeds into their perception of threat posed by members of the minority group presumably associated with the source of the psychological distress (Canetti-Nisim et al, 2009: pp. 380-



Issue: 2/2014

381) (Italics are my emphasis)

The application of this concept to Breivik's case may seem non-obvious. He has no personal exposure to terrorism and should therefore not experience the psychological distress needed to develop more extreme views. However, perhaps the psychological distress may be experienced without personal exposure to terrorist attacks? The narrow worldview Breivik developed during his online activities may have resulted in "selective exposure," aligning to the idea of the filter bubble. Research suggests that "stronger social endorsements increase the probability that people select content," (Messing & Westwood, 2012) meaning that you are more likely to consume content recommended by peers in your community. Could this mean that Breivik consumed content that contributed to the perceived threat of Islam, resulting in real psychological distress and thus radicalization?

Journal EXIT-Deutschland.

und demokratische Kultur

Zeitschrift für Deradikalisierung

Some qualitative research suggests that web access and the Web's affordances may provide venues that increase the likelihood of radicalization and facilitates the process. In his research, Köhler (2012) selected and analyzed interviews with eight former members of the far-right movement in Germany and found that "the Internet provides a perceived constraint-free space and anonymity. This provokes or motivates individuals to speak or act out more radically online as they would normally do offline." (2012: pp. 7). He also found that "the Internet provides a space to share crucial information connected to the chosen lifestyle, such as banned literature, music, clothes and manuals" and that "the Internet gives individuals the perception of a critical mass within the movement" (2012: pp. 8-9).

These findings align well with those of Ravndal (2013), who analyzed Breivik's Internet and Web activities leading up to the attacks in 2011. Applying Köhler's finding to Breivik, it seems that Breivik's isolation from the real world, coupled with his connection to a global network of far-right radical thought, may have created a sense of a "critical mass." However, both authors noted the need for more research on the Web's impact on radicalization.

I use far-right extremist content as a case for this study. While it does not answer the question how the Web impacts radicalization, it might help uncover potential ways in which web users are exposed to far-right ideologies and if web personalization plays a part in that exposure.



ISSN: 2196-8136

Issue: 2/2014

In their research, O'Callaghan et al (2013a) concluded that "[a]Ithough originally composed of dedicated websites, the online extreme right milieu now spans multiple networks, including popular social media platforms such as Twitter, Facebook and YouTube" (pp. 1). In a different study, the same authors (O'Callaghan et al, 2012) found that "stable communities of related users are present within individual country networks, where these communities are usually associated with variants of extreme right ideology. Furthermore, we also [identified] the presence of international relationships between certain groups across geopolitical boundaries" (pp. 1). The international far-right community is publicly active and accessible through SNSs like Facebook and YouTube, increasing the likelihood that users of these platforms may stumble upon such content.

Finally, O'Callaghan et al (2013b) analyzed what they define as "Extreme right content on YouTube." The researchers identified far-right radical YouTube channels (practically synonymous with "users" or "accounts") "originating from links propagated by extreme right Twitter accounts" (pp. 1). Using a computational method, their study showed that the first recommended video associated with an extreme right channel is very likely to be within the same category (i.e. a far-right radical category), but that the category drops off as you go further down the recommended list. This suggests, according to the authors, a strong presence of a far-right extremist filter bubble. In my research, I take on a non-computational approach that attempts to simulate real user experiences of the filter bubble.



Issue: 2/2014

SECTION II: Research Question, Study Design, and Method

In this section, I will develop the question that drove this exploratory research project. First, we take a look at the study's purpose, which springs out of the need to empirically trace and measure Pariser's theory of the filter bubble. Then, I explain how the study's purpose results in a tangible research question with an associated hypothesis. Finally, I explain the rationale and considerations behind the design of the study.

The literature currently lacks empirical research on what personalization looks like for the individual web-user, which is crucial in determining if Web personalization is potentially harmful. Furthermore, while web-search personalization is a very important topic, it may be interesting to complement it with a replicable method that measures user-experienced personalization on content-bearing SNSs. After all, web-users consume content on these platforms. Consequently, the purpose of this study becomes:

To develop a method that can empirically trace and measure personalization on contentbearing SNSs (or social media sites) from a user-perspective.

For the purposes of writing effective algorithms, engineers often use the science of networks. The content hosted on content-bearing SNSs can be regarded as vertices in a graph that are affiliated with each other in multiple ways. Several Google employees outlined how they think about the network that informs YouTube's recommendation and discovery algorithm users encounter when browsing the website:

By studying the viewing patterns and video discoveries of YouTube users in aggregate, we can create an effective video suggestion system that does not rely on the analysis of the underlying videos. The goal is to create a personalized page of video recommendations that not only shows the latest and most popular videos, but also provides users with recommendations tailored to their viewing habits (Baluja et al, 2008: pp. 95)

Embedded in the network approach to creating video recommendation and discovery algorithms is network structure, i.e. an inevitable structure that will emerge for each individual YouTube user interfacing with the algorithm's output. By studying the network structure of the videos YouTube recommends to users, one may observe the personalization that takes place. YouTube algorithms generate an ecosystem of videos, a network of links



Issue: 2/2014

based on predictions and associations.

What structure am I examining? This is where the specific case of far-right, violencepromoting hate speech becomes useful. Therefore, the exploratory research question for this study becomes:

What characterizes the network structure of violent YouTube videos that promote far-right radicalism, as experienced by a user?

The personal data that YouTube factors into its recommendations impacts the network structure of videos from a user-perspective. This becomes my study's independent variable, and the hypothesis reads:

Holding as many other factors as possible constant, the demographic and web-behavioral data accessed by YouTube will influence the observed network structure of far-right, violence-promoting content as experienced by distinct users.

In order to answer the research question, I designed an exploratory experiment in which I step into the shoes of a YouTube user. I analyze what content YouTube exposes to a user who follows the site's recommendations. I start at one video (called seed video) and let YouTube's recommendation engine do the rest—I simply follow the links it recommends to the user, map the different paths, and analyze the videos that the user comes across as a result of following the recommended paths. The experiment generates a graph mapping a decision tree of video recommendations YouTube puts forward to the user. The graph includes information on the type of video content the user was exposed, specifically in regards to promotion of far-right radical content and presence of violence.

In order to test the hypothesis, I created two very distinct user profiles (i.e. Google accounts with distinct web histories) and conducted the experiment twice—once with each user profile—under identical conditions (using the same seed video, time, place, and technical conditions). Any difference observed in the order and the type of content the users were exposed to is then presumably caused by YouTube's personalization algorithm. Since I aimed to study YouTube content that is of political nature—far-right radicalism—I created two profiles that have two distinct and diametrically opposing political views: one far-right (Neo-Nazi, White Supremacist) and one far-left radical (Environmental fundamentalist,



Issue: 2/2014

Figure 2: Directed graph with five nodes

and seven edges

Communist).

Mixed-Method: Social Network Analysis & Quantitative Content Analysis I mixed namely two methods, social network analysis and quantitative content analysis (QCA), to carry out this study. In this section I briefly go over the methods and how I mixed them.

Social network analysis (SNA) presents an opportunity to study relational data. SNA focuses on units' relationship to other units (Knoke & Yang, 2008: pp. 3-9). This is apparent in the network graph,3 which displays how units (nodes) are connected (through edges), and the direction of the connections (see figure 2). The graph forms the basis from which you may calculate several individual node and graph metrics.

Thus far, studies that use SNA to analyze SNSs often look at overall graph structures and metrics. They tend to focus on how network members relate to other members, or how

content relates to other content—regardless of the individual SNS' uses and affordances (Kelly et al., 2012; O'Callaghan et al, 2012 & 2010, Mislove et al, 2007). This is why I suggest

a new data collection method tailored to content-bearing social networking sites. It gathers data from a user-perspective in order create a network graph that maps realistic and plausible routes within a social networking site. In this study, I use YouTube's recommended and related videos algorithm as guide in content exploration.

An alternative SNA method needs to be



Figure 3: Example of a YouTube video-player





ISSN: 2196-8136

Issue: 2/2014

grounded in the user experience and consider how she interfaces with the SNS. In the case of content-bearing SNSs, it is particularly useful to include information about the content characteristics and how users access it. Therefore a combination of SNA and quantitative content analysis (QCA) becomes appropriate. Think about the proposed SNA method as a map of a SNS, outlining how the user not only transports herself in the network, but how she experiences its content. The goal is to map how the user moves within the content ecosystem utilizing YouTube's personalization algorithm, and what content she is exposed to as a result of it.

In my study, nodes are the content a user will consume, in this case a YouTube video. Farright ideological and violent content are of particular interests, and will inform the coding schedule.

The edges in my study represent hyperlinks to other YouTube videos that are accessible from a give YouTube-video page, as found in the "Related and Recommended Videos" section (see figure 3). I wish to simulate and recreate the user-experience within the YouTube ecosystem. If the researcher measures "exposure," she may by default select the videos that appear in the top segment of YouTube's "suggested videos" menu. I take on such "exposure"-approach in this study. Research suggests that a link's prominence in the "recommended" list increases the likelihood of a user following that link4 (Zhou et al, 2010).

This type of network exploration is referred to as "breadth-first search." Since the graph will map potential paths between content nodes in a SNS, the exploration will begin at a given "starting point" from which the search then branches out (breadth-first search) (MIT, 2013). Therefore, an appropriate

Figure 4: Breadth-first tree



⁴ "we find a strong correlation between the view count of a video and the average view count of its top referrer videos, and also discover that the position of a video on a related video list plays a critical role in the click through rate of the video." (Zhou et al, 2010: pp. 409)



Issue: 2/2014

visualization becomes the "breadth-first tree," illustrated in figure 4.

If social network analysis helps answer the part of the research question inquiring about the network structure of YouTube videos, the quantitative content analysis helps classify videos as far-right extreme and violent.

Traditionally, content analysis is defined as "a research technique for making replicable and valid inferences from data to their contexts" (Krippendorff, 1989: pp. 403). However, since the selection criterion and sampling of units in this study is based off breadth-first searches on YouTube, this analysis will only attempt to quantitatively describe core characteristics of the videos in the network analysis, and not create valid inferences. Essentially, I utilized the content analysis method to assign characteristics to the nodes in the network, in order to analyze its structure. Building off the classic work of Berelson, I adapted concepts traditionally applied to mass media to the context of my study.

Berelson wrote about the characteristics of content analysis that focuses on substance. One of the most notable uses of such analysis is comparative content analysis (1951: pp. 35) My study compared and contrasted two YouTube users' breadth-first graphs; one user had a farleft radical profile, the other a far-right.

The "case" in my content analysis is the YouTube video, since the video is a form of "selfcontained expression" (Berelson, 1952: pp. 141) and appropriate in research where the variation within the content is "small or unimportant." Considering the nature of YouTube videos, this rationale applies—videos tend to be short (3-5 minutes) and topical (Cheng, et al, 2008). However, in some cases, YouTube videos are longer than five minutes, and in this study, I encountered many videos that well exceeded the average video length on YouTube (the average video length in my study was close to 40 minutes). In cases where a YouTube video exceeds five minutes in length, only the first five minutes of the video were analyzed. The five-minute limit makes sense from a user perspective. Data show that the completion rate for videos under 30 seconds are 80%; for videos longer than five minutes, it drops to 50% (Ben, 2012). It's more likely that a user watching a 90-minute long YouTube video will be exposed to the first five minutes of the video will



Journal EXIT-Deutschland. Zeitschrift für Deradikalisierung und demokratische Kultur

Issue: 2/2014

Study Design

I want to compare how personalization may influence the structure of YouTube videos, which means I need to approach the same network from two different user perspectives while activating the personalization algorithm. This is where the breadth-first search becomes appropriate by using the same "seed video," i.e. starting point in the network search. I can explore how the users transport themselves into the network following the top recommended videos of each video the user watches. I limited the breadth-first search to five layers, and included the top four video recommendations of each page the users visited. This means that if each video never links back to a previously watched video, such approach yields 341 unique videos (1 + 4 + 16 + 64 + 256). This is the gist of the approach I deployed in this study.

I created two users for this study, Jack Brandon Smith and Richard Scott Williams. Jack is a far-left radical, 25 year old Google user, and Richard is Jack's far-right radical equivalent. They're characterized as far-left and far-right based on their web behavior and history; both searched on radical organizations and visited websites that are classified as extremist by either the U.S. government5, the non-governmental organization Southern Poverty Law Center6, or according to literature on domestic U.S. terrorist organizations (Smith, 1994). I logged on to each users Google accounts and spent two hours performing Google searches on these organizations names and key figures, and visiting their websites in search for YouTube videos they had uploaded. Once I identified YouTube videos associated with the organizations, I watched the videos, liked and commented on them, as well as subscribed to the channels that hosted them. All these activities left behind information that Google and YouTube use to personalize searches and video recommendations.

A look at the YouTube homepage for each user tells us that YouTube picked up on Jack and Richard's radical interests and behaviors. Jack was recommended videos about Che Guevara and revolutions, Richard about white race talks and videos about the US-Mexican border.

⁵ E.g. San Diego, FBI's most wanted terrorist, is an eco-terrorist associated with the extreme left environmentalist movement. Link: http://www.fbi.gov/wanted/wanted_terrorists

⁶ Link to SPL Center's web page outlining their work mapping hate and extremism in the U.S.: http://www.splcenter.org/what-we-do/hate-and-extremism



Journal EXIT-Deutschland. Zeitschrift für Deradikalisierung und demokratische Kultur

Issue: 2/2014

Figure 6: Richard's YouTube recommendations



Once Jack and Richard had sufficient radical "baggage" for YouTube to notice, I was ready to conduct the breadth-first search. The seed video I used for both searches was the National Socialist Movement's promotion video with the same name, uploaded by David Pope. Not only is the National Socialist Movement to date the largest neo-Nazi movement in the U.S. according to Southern Poverty Law Center, the particular video also possessed the qualities I was specifically tracing with my quantitative content analysis: it promoted a hateful ideology and contained violence. The fact that the National Socialist Movement is deemed the largest movement and that the video in question had the most views among the videos promoted on their website, made it a suitable candidate as a seed video for my study.

Google has the opportunity to consider many factors other than a user's account when personalizing search results or YouTube recommendations. Therefore, I performed the entire study on a virtual machine, using VirtualBox and Ubuntu Operating System to simulate the use of new machines; Jack and Richard had their own "brand new" computers. I carried out the study on Georgetown University's WiFi, being assigned the same IP address throughout (141.161.133.163), which is traced back to Washington D.C.

I conducted the first-breadth search manually, meaning that I visited every URL that is included in the study. As Hannak et al.'s study (2013) noted, Google and YouTube's API (application programming interface) sometimes return different results than real-world web



Journal EXIT-Deutschland. Zeitschrift für Deradikalisierung und demokratische Kultur

Issue: 2/2014

surfing.7

⁷ Since the data was retrieved "by hand," there is a slight time difference in the collection process for Jack and Richard's data: I performed Richard's breadth-first search between 10:00 and 11:30 am and Jack's search between 12:10 and 1:30 pm the same day. The influence of this time disparity on the final results on my data gathering is unclear.



Issue: 2/2014

SECTION III: Results & Analysis

I will present the graphs that are the aggregated result of the combined breadth-first search, social network analysis, and quantitative content analysis. First, however, I'll briefly go over the study's strengths and weaknesses.

Reliability Discussion

I developed a rough test-retest check, where I performed a Second [2] three-layer shallow breadth-first search approximately 20 hours after Time One [1]. I did this for both Jack and Richard's accounts. I added the nodes each three-layered deep search yielded (maximum of 21 unique videos per collection time), compared the videos that overlapped between Time One [1] and Time Two [2], and then related the overlapping nodes to all unique nodes collected through for both Jack and Richard.

This test-retest merely looks that the overlap in nominations from the seed video and consequent videos, it does not look at agreement in how the videos link to each other—a major shortcoming in the test.

	Time 1 Unique n	Time 2 Unique n	Overlap Common n	Overlap % of all unique
Richard (right)	20	21	6	6/35 = 17%
Jack (left)	20	21	9	9/32 = 28 %

Table 1: Breadth-First Search Check

It is hard to interpret these results, but I include them for comparison between the networks in my study and comparison with future studies that may take on a similar approach.

I performed an inter-coder reliability test. For the purpose of completeness, I include the level of agreement between main coder and reliability coders on all variables used in the analysis, multiple reliability statistics, as well as the decisions used in the analysis



Issue: 2/2014

computation.

Table 2: Reliability Statistics

	% Agreem ent	Scott' s Pi	Cohe n's Kappa	Krippendo rff's Alpha (nominal)	N Agr e- em ent	N Disa - gre e- me nt	N Cas es	N Deci s- ions
V1: about hateful, far-right ideology?	93.3%	0.878	0.878	0.88	42	3	45	90
V2: Includes symbols associated with hate ideology?	91.1%	0.829	0.83	0.831	41	4	45	90
V3: Supports/counters /ambiguous toward far-right ideology?	91.1%	0.848	0.848	0.85	41	4	45	90
V4: Includes instances of violence?	88.9%	0.817	0.817	0.819	40	5	45	90

Table 2 was computed using <u>http://dfreelon.org/recal/recal2.php</u>.



Issue: 2/2014

The level of agreement is relatively high between coder one and two, compared to, for example, averages between 83% and 88% in Bridges et al.'s (2010) study of violence in violent pornography. I deem the variables reliable as the agreement ranges from 88.9% to 93.3%.

Content Analysis

The content analysis allows us to describe the videos in categories of sentiment towards hateful ideologies and presence of violence. The colors in parentheses denoting each category help us study the graphs later as I use the same colors to distinguish the nodes in the visualization.

	Jack (Far-left) n=206	Richard (Far-right) n=217
Hate-supporting (Red)	6 (3%)	6 (2.8%)
Ambiguous (Blue)	15 (7%)	21 (9.7%)
Hate-Countering (Green)	24 (12%)	1 (0.5%)
Non-hate related (Orange)	114 (55%)	146 (67.3%)
Takedown/foreign8 (Black)	47 (23%)	43 (19.8%)

Table 3: Ideological Content Compared

As you may note, both Jack and Richard's networks include six videos that actively promotes a hateful ideology (including the seed video). However, I observe a significant difference in the number of videos that actively counter hateful ideologies: Jack racked up 24 videos, whereas Richard only encountered one.

Using the Chi-Squared test of Goodness of Fit for the cases that I was able to categorize (i.e.

⁸ "Takedown" means that the video became unavailable between the date of data collection and content analysis, and foreign means that the video was not in English (hence it was not analyzed further).



Issue: 2/2014

category Red, Blue, Green, and Orange) with a significance level at .05 and DF at 39, I can observe a significant difference in the distribution among the four categories.

The chi-square value generated from the test is 29.5, well past the 7.8-threshold observed in the chi-squared table (significance level.05 and DF at 3). This means that the proportion of hate-supporting, hate-countering, ambiguous, and non-hate related videos differs significantly between Jack and Richard's networks.

The large proportion of "non-hate related" videos is problematic. Several videos were documentaries about WWII and content exploring conspiracy theories targeted towards the U.S. government. Since I specifically looked at contemporary far-right ideologies, these videos did not qualify since they didn't formulate a hateful or derogatory argument toward a targeted group. However, I can see how the videos relate to far-right radical ideas, and in a future look at the data, I may want to demystify the non-hate related category.

Furthermore, looking at the presence of violence in the videos, the figures in table 4 highlights that Richard's network appears to include more violent videos than Jack's network; which makes sense since a larger proportion of Richard's non-hate related videos also displayed violence.

	Jack (Far-left) n=206	Richard (Far-right) n=217
Violent Imagery (Square)	49 (24%)	70 (32%)
No Violence (Disk)	110 (53%)	104 (48%)
Takedown/Foreign (Triangle)	47 (23%)	43 (20%)

Table 4: Violent Content Compared

⁹ In the Chi-Squared Goodness of Fit Test, I ingested Richard's frequencies as observed data, and Jack's proportions as expected data. I did this because the test is less reliable if any excepted frequency is below five.



Issue: 2/2014

I performed the same Chi-squared Goodness of Fit Test for the two categories classifying the presence of violence in videos (i.e., "Square" and "Disk", excluding the "Triangle") with a significance level at 0.05 and DF 1.

The Chi-square value for the test becomes 6.5, with the threshold for DF 1 and significance level 0.05 being 3.8, according to the Chi-squared table. Again, I observe a significant difference between the proportion of violent and non-violent videos in Jack and Richard's networks.

Having a basic grasp of what videos appear in the networks, the graphs will help us study how they link together and how a user could experience the content moving through the networks.

I organized the graphs to depict the breadth-first search process: the seed video appears on top of both the graphs, linking down to the first four videos in layer two, in turn linking down to videos in layer three, then four, and lastly layer five. On the next three pages I present both Jack and Richard's full graph visualizations, whereas later, I stratify the graphs in order to highlight key points and observations. Below, I include a table of keys on how to read the graph.



ISSN: 2196-8136

Issue: 2/2014

Graph Keys		
Colors (denotes info on	hateful ideologies)	
Red	Supports hateful ideology	
Blue	Is ambiguous towards hateful ideology	
Green	Counters hateful ideology	
Orange	Is about something other than hateful ideology	
	No info/foreign language	
Shape (denotes presen	ce of violence)	
Solid Square	Violence is present	
Solid Disk	Violence is not present	
Triangle	No info/foreign language	
Size (denotes in degree, i.e. number of links to video)		
Node Size	Smallest node means that no other video linked	
('small' to 7*'small')	to node, large node means that multiple videos	
	linked to node. In degree ranges from 0 to 6.	



ISSN: 2196-8136

Issue: 2/2014



Figure 7: Jack's (far-left) network graph, nodes w/o data 10% transparency



Issue: 2/2014







Journal EXIT-Deutschland. Zeitschrift für Deradikalisierung und demokratische Kultur

Issue: 2/2014

Initial graph observations

The graphs can come off as overwhelming; where should one start to look, and how should one parse the information? Having analyzed these graphs, I will present selected stratified versions where I highlight the nodes and links that are relevant to the particular observation. The hierarchy in the breadth-first graph is important to the analysis. Imagine that you start at the top red node of the graphs; you're immediately asked to make a decision about which one of the four links to follow, unaware what the next four options for travel within the network will be. If we think of the decision-making in terms of chance, every layer degrades the plausibility of travel along a particular path with approximately a division by four (however, this varies depending on backlinks). This means that the plausibility of a user reaching a node in layer five may be as little as 1%; nodes that appear close to the seed video are more important in the analysis than nodes that appear far away from the seed.

A different aspect that's important to note is nodes that cluster in silos—i.e. that the user is more likely to access certain types of content within the graph only if he makes a certain decision early on in the discovery process. Consider Jack's graph, for example: there is a large sub-network of green nodes (videos that counter hateful far-right ideologies), but it only has two access points: through a video in layer two (the green node right after the seed video) or through a blue node in layer four. The videos that counter far-right and hateful ideologies are in a silo.

Next, I will compare graphs that highlight Hate-supporting videos' proximity to seed video, The structure of videos on the topic of far-right ideologies, The structure of violent videos, and The structure of videos that are violent and on the topic of far-right ideologies. I will stratify the graphs, reducing the transparency of the nodes that aren't relevant to the particular observation to 10%. This will make it easy to focus on the links and nodes of interest against the backdrop of their position within the larger network.



Issue: 2/2014

Hate-promoting videos' proximity to seed video

The proximity of hate-supporting videos to the seed video is crucial in understanding if Richard and Jack find themselves in a far-right radical filter bubble. When examining the graphs, I observe a significant difference; Jack's five hate-supporting videos appear significantly further away from the seed video than Richard's five videos (see figure 9 and 10). Comparing the hate-supporting videos' betweenness centrality can further highlight this. Betweenness measures "the extent to which [nodes] lie on the ... shortest distance between [all other] pairs of [nodes] in the network" (Knoke & Yang, 2008: pp. 67). A zero score means that the node does not appear in any other nodes' shortest path to other nodes. Although not an ideal measure of importance in the network (the metric does not take directionality into consideration, for example), betweenness does indicate the relative position in the network; if the video receives a low score, it's likely to appear on the periphery. Furthermore, the in degree of each node will tell us if there are multiple points of entrance to the hate-supporting videos.

	Jack (Far-left)	Richard (Far-right)
In degree of each read node (excluding seed)	1, 1, 1, 1, 1 (Average = 1.0)	2, 1, 1, 1, 2 (Average = 1.4)
Betweenness Centrality for all hate-supporting videos (excluding seed)	0, 0, 0, 0, 0	1225, 1228, 1708, 34, 1002

Table 5: Red Nodes' In Degree And Betweenness Centrality

Aside from the seed video, no hate-supporting content appears in the first three layers of the breadth-first graphs. However, the betweenness centrality numbers, as well as the red nodes' positions in the graphs, clearly paint the image that Richard is more likely to encounter hate-supporting videos than Jack.



Issue: 2/2014



Figure 10: Jack's (far-left) hate supporting videos





Journal EXIT-Deutschland. Zeitschrift für Deradikalisierung und demokratische Kultur

Issue: 2/2014

The structure of videos on the topic of far-right ideologies

Considering only the videos that are about hateful far-right ideologies (the nodes that support, are ambiguous toward, and counter far-right ideologies) will help answer the question if Jack and Richard's networks further promote, or provide counter arguments to, far-right and hateful ideas. Studying figure 11 and 12, you will notice that they differ in two significant ways. First, Richard's breadth-first search presents him with only two videos about hateful far-right ideologies in layer two and three; not until three steps away from the seed do videos about far-right ideologies become prevalent. This contrasts sharply to Jack's graph, where videos about far-right ideologies dominate layer two and three. This highlights the importance of analyzing the videos in a network context and not only from a quantitative perspective. As I described earlier in this chapter, Jack and Richard have approximately an equal number of videos about hateful far-right ideologies within their network; however, those videos appear much closer to the seed video in Jack's case than in Richard's, making it more likely that Jack will encounter content about far-right ideologies (regardless of the video's sentiment towards the ideology).

I also observe that Jack is more likely to encounter videos that counter far-right ideologies than Richard. However, the green nodes in Jack's graph are well contained, meaning that the first decision he makes determines whether he'll encounter videos that counter far-right ideologies.

As for the blue nodes—i.e. videos that are about hateful far-right ideologies, but are ambiguous towards them (oftentimes documentaries)—the viewers' predisposition may determine how they will interpret such videos. For example, one video may be a documentary about the Aryan Brotherhood—a white prison gang—and Richard, who's a supporter of white nationalism, may admire the points the documentary makes about the toughness and race commitment of the gang members. On the other hand, Jack, who's a fierce opponent of white nationalism, may take away the points about the gang's lack of sophistication and intellectual reflection, and the documentary will strengthen his belief that white nationalists are dumb. If this rationale holds true, Jack was served with a lot more content relevant to his cause than Richard, as blue nodes dominate Jack's second and third



Journal EXIT-Deutschland. Zeitschrift für Deradikalisierung und demokratische Kultur

Issue: 2/2014

layers.



Issue: 2/2014

Figure 11: All Richard's (far-right) videos about hateful ideologies

Journal EXIT-Deutschland.

und demokratische Kultur

Zeitschrift für Deradikalisierung



Figure 12: All Jack's (far-left) videos about hateful ideologies





Issue: 2/2014

The structure of violent videos

Common for both figure 13 and 14 is that three out of four videos in layer two contains violence, which means that the chance that Jack and Richard will encounter a video containing violence is 75%. It's clear that the seed video is associated with violent videos, perhaps regardless of the profiles YouTube uses to personalize the results. The prevalence of violent content is similar for both Jack and Richard in the layers closest to the seed video, as summarized in the table below. Alternatively, perhaps YouTube senses Jack and Richard's past interest in violent and ideological content (although on diametrically opposed sides of the political spectrum) and therefore provides them with content that includes violence already in layer 2.

	Jack (far-left)	Richard (far-right)
Layer 2	3	3
Layer 3	6	7
Layer 4	14	18
Layer 5	25	41

Table 6: Number Of Videos Containing Violence, Per Level





Figure 13: All Richard's (far-right) violent videos

Figure 14: All Jack's (far-left) violent videos





Issue: 2/2014

The structure of all videos that are violent and far-right ideological

So what about the instances where a video is both on the topic of a hateful far-right ideology and contains elements violence? The combination would make for the most powerful and negative type of impact, where a message of hate (regardless of the video's supportive or countering position) is coupled with violence, and therefore clearly illustrates that ideology and violence are associated with each other.

Neither Jack nor Richard's graphs contain hate-supporting and violent videos other than the seed video. But the similarities stop there. Not only does Jack's graph contain more violent videos on the topic of far-right ideologies, a majority of them appear close to the seed video. This is in sharp contrast to Richard's graph, where a majority appears in layer five (see figure 15 and 16).







Figure 16: All Jack's (far-left) videos that include violence and are about far-right ideologies





Journal EXIT-Deutschland. Zeitschrift für Deradikalisierung und demokratische Kultur

Issue: 2/2014

Summary of key takeaways

To summarize the empirical observations I described in this chapter, I wish to start with discussing the benefit of incorporating network perspective in a content analysis. The most striking finding may be that, although I observe a significant difference between the types of content Jack and Richard are presented with on an quantitative level, the network analysis complicates and nuances the picture. Richard's videos includes a far larger proportion of violent videos than did Jack's (70 vs. 49 videos). However, in the network analysis, it becomes clear that Jack, in fact, is more likely to watch these violent videos since they appear much closer to the seed. The sheer number or proportion of content isn't enough to explain the impact it's likely to have on the user experience.

This study specifically searched for hateful far-right ideologies and presence of violence in YouTube videos. Yet only one video—the seed video, which I deliberately selected as a starting point for this analysis—fulfilled both of these requirements. This leads me to doubt the generalizability of the results; I wouldn't say that this methodology, applied to any YouTube video that promotes a hateful far-right ideology and includes violence, would yield similar results. Rather, I'd say that the methodology does a good job of mapping a user's experience of a content-bearing SNS, and that this study provides a compelling case for its usefulness. Furthermore, I believe my study makes an argument for both the presence of personalization on YouTube as well as its effect on the user's experience of the website.



Issue: 2/2014

SECTION IV: Conclusion

I will contextualize the findings using the theories and published research I discussed previously. Finally, I'll suggest future research that will help explore personalization of the web further.

Hypothesis: Holding as many factors as possible constant, the demographic and webbehavioral data accessed by YouTube will influence the observed network structure of violent and far-right radical content as experienced by distinct users.

I observed a statistically significant difference between the content that the two distinct YouTube users were exposed to, performing the same experiment under near identical conditions. YouTube indeed personalized the content based on the users' distinct profiles. However, due to the "blackbox" nature of Google's algorithms, I find it difficult to estimate how unknown factors and "noise" might have impacted the results. More experiments carried out across more YouTube profiles might provide a more certain result.

Research Question: What characterizes the network structure of violent YouTube videos that promote far-right radicalism, as experienced by a user?

The study did not generate a conclusive answer to this question. However, it does provide some tentative indications of structures. Based of the users' profile and the coviewership graph of the seed video, YouTube seemed to include violent content in the top four recommendations following the seed video. Furthermore, I observed that the content became more self-similar the deeper into the network the users traveled. In this experiment, subsequent content seemed to have been more strongly influenced by videos in layer two and three than by the seed video.

I can conclude that neither user ended up in a violent, hate-supporting, and radical content bubble. However, due to the small number of experiments and the lack of multiple points of comparison, I cannot decisively conclude that such a scenario is impossible. Under the "right" circumstances, it seems possible that a user could find herself in a hatesupporting and far-right content bubble on YouTube. It is also important to note what my research did not study; the numerous videos about WWII and conspiracy theories that did



Issue: 2/2014

not qualify in my pre-defined categories might still carry importance in this context. Conspiracy theories are indeed part of far-right and far-left narratives, however, because of the pre-defined coding schedule, they do not show up in the analysis.

The study's purpose: To develop a method that can empirically trace and measure personalization on content-bearing SNSs from a user-perspective.

I deem the method a successful tool in analyzing user-experiences on contentbearing SNSs. If the method is deployed in an experimental study across multiple user profiles, then it can also be used to empirically trace and measure personalization. Most importantly, other research projects that aim to study the Web's impact on users tend to utilize atomized methods of collecting and analyzing data, failing to account for human interpretation of the data. However, this method is wholeheartedly devoted to explore the Web from a user's perspective and creates a richer picture of the Web's impact on a user.

Although it is clear that YouTube's personalization algorithm did not create a pervasive farright radical filter bubble under the conditions tested in this project, I cannot debunk the theory of harmful personalization of the Web. There may still be potential for radicalization through far-right content bubbles on websites like YouTube. However, it is important to consider the entire context in which radicalization takes place. As I have stated previously, people can watch violent and hateful content without prompting them to cause so much as a pub brawl.

I observed some evidence that YouTube's recommendation algorithm did indeed generate bubbles of self-similar content, particularly in regards to content which contained violence. In regards to hate speech, however, my study could not demonstrate that there is an eminent threat from YouTube recommending hateful content to even supposed radical users. I could observe a deficit of counter-speech videos in the far-right user's video network, but there were not a significant number of hate-speech videos to counter in the first place. We need more research on this topic before we can put forward serious recommendations on the political and legal treatment of personalization algorithms.

My study assumes that users will follow YouTube's top recommendations regardless of the recommended videos' metadata (such as title, thumbnail, length, uploader, and number of



ISSN: 2196-8136

Issue: 2/2014

views). However, my sense is that those metadata, in combination with the video's ranking in the recommended list, will impact whether a users is likely to follow the link or not, and that different user motives will influence a user's decision-making process. For example: a user that actively researches a topic on YouTube will probably be more likely to follow links to longer videos with fewer views, if the title and thumbnail of the link seem relevant to what the user is researching. However, a user who's taking a micro pause in her office might not follow a link to a video that's 54 minutes long, even though that video is YouTube's first recommendation. In order to further refine the methodology I developed, detailed research on how users search and discover videos on YouTube would help fine-tune the rationale that goes in between what links are included in future iterations and applications of my method. As this study explored how YouTube's personalization algorithms may or may not drive a user's exposure to radical and violent content, it does not provide a clear connection

between exposure to radical content and radicalization as a process itself. More research on the Web's explicit role in an individual's radicalization process would inform the relevance of research like this in the study of political radicalization.



Issue: 2/2014

APPENDICES

Appendix 1: Coding Schedule

Variable	Variable Values	Definitions
Meta Data on the video included:	 A. Entry ID B. Full URL to YouTube Video C. Video Title D. Uploaded By (User/Channel) E. Date of upload F. Video Length (HH:MM:SS) G. Number of Views H. Primary Language I. YouTube Video Category 	The information as is at time of coding. If the video was made unavailable between breadth-first search and time of coding, the reason for removal was noted in section B and then further analysis stopped. If primary language was other than English, only meta data was collected but further analysis stopped.10
Variable 1: Is the video about the topic of ideology, politics, race, nationality, and/or religion from a far-right radical perspective?	 Yes (as defined in coding scheme) 90. Yes, but in other hateful ideas expressed towards group of people not defined in coding scheme (note answer qualitatively) 99. NO 	A minimalist criterion applied, meaning that the mere occurrence of the topic records as 1-90.
Variable 2: Do symbols or logos associated with hateful ideologies appear	1. Yes 2. No	A minimalist criterion applied, meaning that the mere occurrence of symbol records as

¹⁰ Since I assume the user perspective, non-English videos were not included in the content analysis (they're noted as "missing cases" and blank nodes in the network analysis).



Journal EXIT-Deutschland. Zeitschrift für Deradikalisierung und demokratische Kultur

Issue: 2/2014

in the video?		1.
Variable 3: does the video support, counter, or not take any stance toward the hateful ideology?	 Supports Counters It's Ambiguous Does not apply due to V1==99. 	From the perspective of the producer of the video. Coder may use video's description for guidance.
Variable 4: Does the video's imagery display violence?	1. Yes 2. No	A minimalist criterion applied, meaning that the mere occurrence of violence records as 1.



Issue: 2/2014

Appendix 2: Dates Of Web Research Activities

Date	Activity	
February 15, 2014	Creation of Jack and Richard's Google accounts on individual virtual machines running on Ubuntu OS. (IP: 141.161.133.152)	
March 1, 2014	2 hours of search on far-right and far-left web content and YouTube videos each for Richard and Jack (IP: 141.161.133.163)	
March 4, 2014	Breadth-first search using same seed video. (IP: 141.161.133.163)	
March 5, 2014	Replication of breadth-first search using same seed video. (IP: 141.161.133.163)	
March 13-17, 2014	Performing quantitative content analysis and inter-coder reliability test.	



Journal EXIT-Deutschland. Zeitschrift für Deradikalisierung und demokratische Kultur

Issue: 2/2014

BIBLIOGRAPHY

Baluja, S., Seth, R., Sivakumar, D., Jing, Y., Yagnik, J., Kumar, S., ... Aly, M. (2008). Video suggestion and discovery for youtube: taking random walks through the view graph. In Proceedings of the 17th international conference on World Wide Web (pp. 95–904). Mountain View. Retrieved from http://dl.acm.org/citation.cfm?id=1367618

Berelson, B. (1952). Content analysis in communication research (Vol. 44, p. 220). New York, NY: Hafner Publishing Company. doi:10.1086/617924

Boyd, D. M., & Ellison, N. B. (2008). Social Network Sites: Definition, History, and Scholarship. Journal of Computer-Mediated Communication, 13(1), 210–230. doi:10.1111/j.1083-6101.2007.00393.x

Bridges, A. J., Wosnitzer, R., Scharrer, E., Sun, C., & Liberman, R. (2010). Aggression and sexual behavior in best-selling pornography videos: a content analysis update. Violence against Women, 16(10), 1065–85. doi:10.1177/1077801210382866

Canetti-Nisim, D., Halperin, E., Sharvit, K., & Hobfoll, S. E. (2009). A New Stress-Based Model of Political Extremism: Personal Exposure to Terrorism, Psychological Distress, and Exclusionist Political Attitudes. The Journal of Conflict Resolution, 53(2), 363–389. doi:10.1177/0022002709333296

Cheng, X., Dale, C., & Liu, J. (2008). Statistics and Social Network of YouTube Videos. 2008 16th Interntional Workshop on Quality of Service, 229–238. doi:10.1109/IWQOS.2008.32

Davidson, J., Livingston, B., Sampath, D., Liebald, B., Liu, J., Nandy, P., Lambert, M. (2010). The YouTube video recommendation system. In Proceedings of the fourth ACM conference on Recommender systems - RecSys '10 (p. 293). New York, New York, USA: ACM Press. doi:10.1145/1864708.1864770



Issue: 2/2014

Greenberg, E. S. (1970). Political Socialization. Palo Alto: Atherton Press. Hannak, A., Sapiezynski, P., Kakhki, A. M., Krishnamurthy, B., Lazer, D., & Mislove, A. (2013). Measuring personalization of web search. In 22nd international conference on World Wide Web (pp. 527–538). Retrieved from http://dl.acm.org/citation.cfm?id=2488435

Kelly, J., Barash, V., Alexanyan, K., Etling, B., Faris, R., Gasser, U., & Palfrey, J. (2012). Mapping Russian Twitter. Berkman Center Research Publication (Vol. 7641). Retrieved from <u>http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2028158</u>

Knoke, D., & Yang, S. (2008). Social Network Analysis. Social Networks (2nd ed., Vol. 8, p. 133). Los Angeles, CA: Sage Publications. Retrieved from http://books.google.com/books?id=buiJi6HtGusC&pgis=1

Köhler, D. (2012). Internet and Radicalization: Connecting the Dots - the Role of the Internet in the Individual Radicalization Processes of Right-Wing Extremists (No. 01/12). ISRM -Institute for the Study of Radical Movements. Berlin.

Krippendorff, K. (1989). Content analysis: An introduction to its methodology (p. 7). Retrieved from <u>http://repository.upenn.edu/asc papers/226</u>

Kydd, A., & Walter, B. F. (2002). Sabotaging the Peace: The Politics of Extremist Violence. International Organization, 56(2), 263–296. doi:10.1162/002081802320005487

Majumder, A., & Shrivastava, N. (2013). Know your personalization: Learning topic level personalization in online services. In 22nd international conference on World Wide Web (pp. 873–884). Retrieved from <u>http://dl.acm.org/citation.cfm?id=2488464</u>



Issue: 2/2014

Messing, S., & Westwood, S. J. (2012). Selective Exposure in the Age of Social Media: Endorsements Trump Partisan Source Affiliation When Selecting News Online. Communication Research, 1–23. doi:10.1177/0093650212466406

Mislove, A., Marcon, M., Gummadi, K. P., Druschel, P., & Bhattacharjee, B. (2007). Measurement and Analysis of Online Social Networks. In Proceedings of the 7th ACM SIGCOMM conference on Internet measurement (pp. 29–42). Retrieved from http://dl.acm.org/citation.cfm?id=1298311

Mobasher, B. (2007). Data mining for web personalization. In The Adaptive Web (pp. 90–135).SpringerBerlinHeidelberg.Retrievedfromhttp://link.springer.com/chapter/10.1007/978-3-540-72079-93

O'Callaghan, D., Greene, D., Conway, M., Carthy, J., & Cunningham, P. (2012). An Analysis of Interactions Within and Between Extreme Right Communities in Social Media. arXiv:1206.7050v1, 1–15. Retrieved from <u>http://arxiv.org/abs/1206.7050</u>

O'Callaghan, D., Greene, D., Conway, M., Carthy, J., & Cunningham, P. (2013a). The Extreme Right Filter Bubble. arXiv:1308.6149v1. Retrieved from <u>http://arxiv.org/abs/1308.6149</u>

O'Callaghan, D., Greene, D., Conway, M., Carthy, J., & Cunningham, P. (2013b). Uncovering the Wider Structure of Extreme Right Communities Spanning Popular Online Networks. arXiv:1302.1726v2. Retrieved from http://arxiv.org/abs/1302.1726

Pariser, E. (2011). The Filter Bubble - What the Internet is hiding from you (1st ed.). New York: The Penguin Press.

Ravndal, J. A. (2013). Anders Behring Breivik's use of the Internet and social media. JEX Journal for Deradicalization and Democratic Culture, 2, 172–185. Retrieved from



Issue: 2/2014

http://www.doaj.org/doaj?func=fulltext&ald=1930764

Santos, R. L. T., Rocha, B. P. S., Rezende, C. G., & Loureiro, A. A. F. (2007). Characterizing the YouTube video-sharing community. Federal University of Minas Gerais (UFMG), Belo Horizonte, Brazil, Tech. Rep. Retrieved from <u>http://security1.win.tue.nl/~bpontes/pdf/yt.pdf</u>

Seierstad, Å. (2013). En Av Oss (p. 541). Stockholm: Albert Bonniers Förlag. Smith, B. L. (1994). Terrorism in America: Pipe bombs and pipe dreams. SUNY Press.

Zhou, R., Khemmarat, S., & Gao, L. (2010). The impact of YouTube recommendation system on video views. In Proceedings of the 10th ACM SIGCOMM conference on Internet measurement (pp. 404–410). New York, New York, USA: ACM Press. doi:10.1145/1879141.1879193