



Protect the Filter Bubbles: Emphasizing User Speech Rights in Algorithmically Curated Online Forums

Priya Dames

University of Florida

Faculty Mentor: Jasmine McNealy, Department of Telecommunication

Abstract

In 2018, *Knight v. Trump* sparked discussion about the boundaries between government and citizen speech on social media. Some scholars argue that the courts erred in their decision to characterize the speech in question as government speech. Others argued that the court decided correctly and claimed that the use of forum analysis was necessary to protect both the health of our democracy and the First Amendment rights of social media users. Within the context of algorithmic curation of social media feeds, this article argues that (1) social media platforms are not designated public forums due to the algorithmic curation of online user speech, (2) due to this, the public forum doctrine should not have been applied to the *Knight v. Trump* case, (3) despite this, user speech rights should be protected online. It also reviews proposed models of thinking that could address unresolved issues of the case.

Keywords: algorithmic curation, public forum, *Knight v. Trump*

Introduction

Social and political engagement have changed a great deal in the past couple of decades as citizens, public officials, and civil society migrate online. The Internet has disrupted social and political engagement that was previously dominated by leading media institutions of the 20th century (Bruns & Highfield, 2015). This is reflected in recent presidential elections where citizens were more likely to search for information about political candidates on social media. According to a Pew Research Center study, one in five U.S. adults get their political news primarily through social media platforms such as Twitter (Mitchell et al., 2020).

While the Internet has the potential to create a new public sphere online, there are also aspects of the Internet that can “simultaneously curtail and augment that potential,” including information access inequality and fragmentation of political discourse (Papacharissi, 2002, p. 9). This article argues that algorithmic curation is an additional aspect of the Internet that can curtail its potential to be a public sphere and uses *Knight v. Trump* to demonstrate why. In *Knight v. Trump*, the Knight First Amendment Institute sued President Donald Trump for blocking seven

individuals from his Twitter account, @realDonaldTrump (*Knight First Amendment Institute v. Trump*, 2018). The court ruled that portions of Trump's Twitter account were designated public forums, making it unconstitutional for him to block citizens based on viewpoint. The case signals a significant shift in how political engagement will be conducted as social media becomes more essential to discourse between citizens and public officials.

This article argues that Trump's Twitter account is not a designated forum and that, despite this, user speech rights should be protected online using new legal tools that account for the fluid nature of online speech. It will discuss the nature of the "interactive space" defined by the court, the role of filter bubbles in shaping designated public forums, and strategies for protecting online user speech (*Knight First Amendment Institute v. Trump*, 2018, p. 549).

Knight First Amendment Institute v. Trump

In 2017, the Knight First Amendment Institute at Columbia University filed a lawsuit against President Donald Trump on behalf of seven individuals who were blocked from Trump's Twitter account, @realDonaldTrump, after expressing viewpoints critical of the President and his policies (*Knight First Amendment Institute v. Trump*, 2018). At the federal court, the case focused on whether a public official can block an individual from their Twitter account in response to the political views the person expressed. While the Knight Institute argued that unconstitutional viewpoint discrimination had taken place, Trump argued that the government was merely exercising its right to speech through its blocking of the individuals. The United States District Court for the Southern District of New York ruled that it was unconstitutional under the First Amendment for the President to block an individual from his Twitter account in response to the political views the person expressed (*Knight First Amendment Institute v. Trump*, 2018). This ruling was supported by the finding that the interactive space of a public official's Twitter account is a designated forum, making Trump's actions unconstitutional viewpoint discrimination (*Knight First Amendment Institute v. Trump*, 2018).

The Interactive Space

The court defined the interactive space of @realDonaldTrump to be the space "where Twitter users may directly engage with the content of the President's tweets" (*Knight First Amendment Institute v. Trump*, 2018, p. 549). This engagement includes replying to, retweeting, or liking a tweet sent by Trump (*Knight First Amendment Institute v. Trump*, 2018). Although it has been called by a variety of terms, the concept of interactive spaces as the Court has defined it, has

been studied by many researchers. Rathnayake and Suthers (2018) characterize the interactive spaces of the Internet as a “networked public,” a type of public space that is modified by networked technologies (boyd, 2010, p. 1; Rathnayake & Suthers, 2018). As boyd (2010) emphasizes, “networked publics” are both the space itself and “the imagined collective that emerges as a result of the intersection of people, technology, and practice” (boyd, 2010, p. 1). Tully and Ekdale (2014) demonstrate this exact quality of online interactive spaces in their documentation of the way hashtags have been used to create interactive spaces within the Kenyan Twittersphere (Tully & Ekdale, 2014). They go on to state that these spaces are used as a place for leisure as much as they are used for civic engagement.

The courts used the public forum doctrine in order to analyze the interactive space of @realDonaldTrump. In forum analysis, a forum is identified by pinpointing the access sought by the speaker who, in this case, are the plaintiffs blocked by Trump (*Cornelius v. Naacp Legal Defense & Ed. Fund, Inc.*, 1985). The plaintiffs were seeking access to the interactive space of Trump’s tweets, where they could directly interact with Trump’s tweets. They were not seeking access to the entire account of @realDonaldTrump in order to send tweets as the President or receive his notifications. Nor were they seeking access to the comment thread of Trump’s tweet which consisted of the initial tweet, direct replies, and second-order replies. The comment thread of Trump’s tweets were still accessible to the blocked plaintiffs who could still view replies to Trump’s tweets and post replies to those replies. They were seeking access to the interactive space of Trump’s tweets, in which they can directly reply to Trump and retweet his tweets.

Knight v. Trump advanced to the United States Court of Appeals for the Second Circuit where it was ruled, once again, that Trump’s actions were unconstitutional under the First Amendment.

Public Forum Doctrine and the Government Speech Doctrine

Much of the debate regarding *Knight v. Trump* centered on the court’s correct or incorrect use of the public forum doctrine as opposed to the government speech doctrine. The public forum doctrine is a tool used in First Amendment law to determine if speech restrictions administered by the government on public property are constitutional or not (*Forums*, 2021). It is applied to spaces that qualify as a forum under the First Amendment. If the forum a speaker is seeking access to is public property or private property dedicated to public use, it can qualify as a forum for First Amendment purposes (*Cornelius v. Naacp Legal Defense & Ed. Fund, Inc.*,

1985). If a space is a forum under the First Amendment, the public forum doctrine states that the government cannot exercise viewpoint discrimination unless it serves “compelling state interest and [is] narrowly tailored to meet the needs of that interest” (*Forums*, 2021). The public forum doctrine also categorizes government property into four groups: traditional, designated, limited, and non- public forums (*Perry Educ. Ass’n v. Perry Educators’ Ass’n*, 1983). The government speech doctrine states that the government is able to determine the contents of its speech without being restricted by the Free Speech Clause of the First Amendment (*The Government Speech Doctrine*, 2021).

Traditional public forums are areas that have been traditionally used for political speech and debate (*Forums*, 2021). While the government cannot engage in content-based discrimination in these forums, it can impose content-neutral restrictions on the time, place, and manner of speech. Designated public forums are public property that have been opened by the government for public expression (*Forums*, 2021). The speech within the forum is given the same First Amendment protections as traditional public forums.

Some scholars argue that the courts were correct to use the public forum doctrine to classify the interactive space of Trump’s tweets as a designated public forum because it preserves democratic values (Morales, 2020). Citing the “pluralist theory of democracy” and *Packingham v. North Carolina*, which refers to social media platforms as “the modern public square,” Benson (2019) argues that the courts were correct to use public forum analysis because it “enables [a] wide variety of interest groups to access and directly responds to their representatives’ policies” (Benson, 2019, pp. 87-110; *Packingham v. North Carolina*, 2017, p. 1732). Other scholars argue that the courts were incorrect to use the public forum doctrine and should’ve used the government speech doctrine instead. According to Beausoleil (2019), Trump’s choice to continue using his personal Twitter account after becoming President further proves his intentions to use the platform as a regular user rather than a government entity (Beausoleil, 2019). As a result, Trump’s control of his Twitter account is not enough to establish state action and, consequently, is not enough to apply the public forum doctrine.

Before further examining the applicability of either of these doctrines, it is important to first consider this debate within the context of algorithmic curation of social media feeds.

Algorithmic Curation

It is difficult to pinpoint the beginning of the use of algorithms for curative purposes online. In 2009, Google was one of the first companies to use personalization features on their platform to handle the recent massive increase in information available on the Internet (Bozdag, 2013). The algorithms assist the Google search engine in gathering information about a user's search history and using this data to recommend a search that is most suited to their preferences. This goal of predicting what a user wants and recommending something accordingly is at the core of curation algorithms and has since then been used on Facebook, Twitter, Instagram, and YouTube to curate the experience of users. Twitter, specifically, uses curation algorithms to manage, rank, and filter tweets on a user's timeline to the user's preference (Burrell et al., 2019). Twitter discloses their use of personalization and recommendation systems which is called "suggested content" by the company (*About Your Twitter Timeline*, 2021). Regarding the "home timeline" of a user, Twitter states that it uses "a variety of signals, including how popular [a tweet] is and how people in [a user's] network are interacting with it" (*About Your Twitter Timeline*, 2021). While this statement gives users some insight into the factors used to suggest content to them, it is not an exhaustive list. The popularity of a tweet as well as its relevancy to a user are crucial metrics to curation algorithms but there are a number of undisclosed metrics the company uses as well. This can make algorithmic curation a difficult topic to study directly since the metrics used by a recommendation algorithm can be kept private by companies for proprietary purposes (Buolamwini, 2017).

Twitter has also disclosed its use of curation algorithms for the comment sections of tweets, which it calls "conversation ranking" (*About Conversations on Twitter*, 2021). This feature groups replies to tweets into "sub-conversations" and ranks them in order of "relevance" to the user rather than displaying them in chronological order. The documentation does not provide an exhaustive list of the factors used to group reply tweets but does note that relevancy, credibility, and safety are all factors considered when grouping and ranking tweets. For example, a reply may rank higher in the comment section if the author of the reply is someone a user follows (*About Conversations on Twitter*, 2021). It is important to note, then, that users play a large role in shaping these algorithms through their usage of the platform. Through "liking," "retweeting," "following," clicking links, and interacting with advertisements, users are continuously

supplying curative algorithms with data about their preferences which the algorithm uses to predict what to show them next (Burrell et al., 2019, p. 2).

Scholars have alerted the public to the risks of bias in ranking systems. Twitter has recently come under fire for possible racial bias in its automated image cropping neural network, resulting in its switch to fully displaying photos on user's timelines (Canales, 2020). Bias in natural language processing algorithms used by the company to flag hate speech can also have inherent racial and ethnic biases (Sap et al., 2019). There is also the concern of bias in Twitter search results for politics-related queries (Kulshrestha et al., 2017). However, bias is not the only concern here. In the context of public forum analysis, algorithmic curation is also concerning. According to a Pew Research Center study, one in five U.S. adults get their political news primarily through social media (Mitchell et al., 2020). This gives curation algorithms an incredible power to shift public political discussion, to "grant visibility," and to "certify meanings" in the public sphere (Burrell et al., 2019 p. 2; Teevan et al., 2011). Curation algorithms can determine what information a user is exposed to by pushing certain perspectives to the top of a user's feed and pushing others to the bottom. Most importantly, they can control how information is disseminated in a conversation taking place between users online – what conversations a user is exposed to, what replies they can view, and who they can see discussing a particular topic.

Filter Bubbles

The use of algorithmic curation can lead to the creation of filter bubbles around users. A filter bubble as defined by Eli Pariser, who coined the term in 2010, is "a unique universe of information for each of us [...] which fundamentally alters the way we encounter ideas and information" (Pariser, 2011, p. 10). As a personalization algorithm continuously refines its perceptions of a user's preferences, the recommendation algorithm filters out certain information that does not align with this perception and selectively guesses information that does. This results in the user being shielded from diverse information and becoming isolated in their own cultural or ideological bubble.

Pariser's theory emphasizes an unprecedented characteristic of this type of media consumption: we each are alone in our own filter bubbles. Pariser demonstrates this by comparing filter bubbles to specialized cable channels. While a viewer of a home improvement or golf channel shares a frame of reference with viewers of the same channel, a user in a filter

bubble cannot. A personalization algorithm ensures that each user's feed is narrowly tailored to their interest, making it difficult for users to share the same "universes" of information. This is significant when discussing public forum analysis on social media. If we are each alone in our own filter bubbles, it is important to question whether public forums are still possible on social media.

Are Interactive Spaces Designated Public Forums?

Interactive spaces are curated in such a way that conversations are experienced differently from user to user. Replies, retweets, and likes are the tools inherent to the interactive space of a particular tweet. The landscape on which these tools are used, the comment thread, is curated to each user's preferences. When deciding whether interactive spaces should be classified as designated public forums, municipal meeting rooms can be used as an example. Municipal meeting rooms are generally considered designated public forums because they are public property opened by the government for public expression (*Forums*, 2021). However, if a particular municipal meeting room was full of speakers who each had their own perceptions of how a particular meeting was transpiring – of who is currently speaking, of what conversations are "relevant" to the meeting, of what conversations are heard at all – is it correct to classify the meeting room as a designated public forum? The speakers each experienced the meeting in very different ways and left with different perceptions of how the meeting took place. Additionally, although the government has the capacity to restrict speech in a designated public forum, it is allowed to do so only if the restriction serves a "compelling state interest and [is] narrowly tailored to meet the needs of that interest" (*Forums*, 2021). In the case of this municipal meeting room and the interactive space of a tweet, there is neither a compelling interest nor a narrowly tailored restriction of speech. There is only the organization of the political, personal, and social information flows of every user by Twitter, a private technology company, for the purpose of user personalization at the cost of diversity of perspectives (Tufekci, 2014). It is incorrect to classify the aforementioned municipal meeting room as a designated public forum in the same way that it is incorrect to classify the interactive space of a tweet as a designated public forum. Within their filter bubbles, each user is experiencing an overrepresentation of certain views and an underrepresentation of others in their interactive spaces making it difficult to express ideas equally in the forum (as speakers would in a public forum) or to control the contents of speech taking place (as the government would in accordance with the government speech doctrine). It is

clear, then, that the debate between the use of the public forum doctrine versus the government speech doctrine is not relevant to the key issues of *Knight v. Trump* if it is not considering the confounding factor of algorithmic curation.

Protecting the Filter Bubbles

Regardless of the doctrine applied, online user speech rights should still be prioritized. The utilization of algorithmic curation does not erase the need for an online public sphere, “a domain of social life where public opinion can be formed” and where matters of public concern can be argued unencumbered by state and corporate interests (Corner, 1995; Habermas, 1991; Mwingenmeir, 2014).

Scholars have proposed new models of thinking that could address this need. Some scholars have suggested a “mixed speech” approach to the public forum doctrine versus government speech debate (Corbin, 2008; Wiener, 2020). This approach allows online speech to be dual in nature by acknowledging that it is neither purely private nor purely governmental while also ensuring that the interests of all parties are protected (Corbin, 2008). Corbin argues that classifying mixed speech as government speech infringes on speaker’s free speech interests while classifying mixed speech as private speech infringes on the government’s compelling state interests. Corbin concludes that a better approach to handling mixed speech as it relates to First Amendment jurisprudence is to subject viewpoint restrictions to intermediate scrutiny (*Intermediate Scrutiny*, 2021). Using this approach, a statute governing how public officials block citizens online would be constitutional only if “(1) it has a closely tailored, substantial interest that is clearly and publicly articulated; (2) it has no alternate means of accomplishing the same goal; and (3) private speakers have alternate means of communicating to the same audience” (Corbin, 2008, p. 675).

Multi-party negotiation is another tool that can be used to come to a consensus about how speech should be categorized online. Mongiello proposes the use of positional bargaining between the President, a Twitter representative, and potential plaintiffs in order to negotiate rules for Twitter users (Mongiello, 2019). This is a more participatory approach to addressing the need for a public sphere and would give the courts a roadmap for future decision making regarding online speech issues.

Conclusion

Knight v. Trump was a landmark case in defining the bounds of forum analysis on social media. However, the case did not fully consider how algorithmic curation affects the classification of social media as a designated public forum. Mischaracterization of interactive spaces as designated public forums could lead to infringement of citizen and government speech rights. While the public forum and government speech doctrines may not be the most suitable for handling conflicts between public officials and citizen's speech rights online, citizen and government free speech rights should aim to be preserved. New legal tools will be needed to continue to deal with unprecedented cases related to online speech issues and to protect the free speech rights of users.

Acknowledgements

I would like to thank Dr. Jasmine McNealy for her continued guidance and support as my faculty mentor on this project. I would also like to recognize the assistance that I received from the UF University Scholars Program.

References

- About conversations on Twitter. (2021). <https://help.twitter.com/en/using-twitter/twitter-conversations>
- About your Twitter timeline. (2021). <https://help.twitter.com/en/using-twitter/twitter-timeline>
- Beausoleil, L. (2019). Is trolling Trump a right or a privilege: The erroneous finding in *Knight First Amendment Institute at Columbia University v. Trump*. *Boston College Law Review*, 60, II–31. https://lawdigitalcommons.bc.edu/bclr/vol60/iss9/3?utm_source=lawdigitalcommons.bc.edu%2Fbclr%2Fvol60%2Fiss9%2F3&utm_medium=PDF&utm_campaign=PDFCoverPages
- Benson, S. J. (2019). @PublicForum: The argument for a public forum analysis of government officials' social media accounts. *Washington University Jurisprudence Review*, 12, 85. https://openscholarship.wustl.edu/law_jurisprudence/vol12/iss1/7/
- boyd, danah. (2010). Social network sites as networked publics: Affordances, dynamics, and implications. *Networked Self: Identity, Community, and Culture on Social Network Sites*, 39-58. <https://www.taylorfrancis.com/chapters/edit/10.4324/9780203876527-8/social-network-sites-networked-publics-affordances-dynamics-implications-danah-boyd>
- Bozdag, E. (2013). Bias in algorithmic filtering and personalization. *Ethics Inf Technol*, 15, 209–227. <https://doi.org/10.1007/s10676-013-9321-6>
- Bruns, A., & Highfield, T. (2015). Is Habermas on Twitter? Social media and the public sphere. <https://www.taylorfrancis.com/chapters/edit/10.4324/9781315716299-5/habermas-twitter-axel-bruns-tim-highfield>

- Buolamwini, J. A. (2017). Gender shades: Intersectional phenotypic and demographic evaluation of face datasets and gender classifiers. [Thesis, Massachusetts Institute of Technology]. <https://dspace.mit.edu/handle/1721.1/114068>
- Burrell, J., Kahn, Z., Jonas, A., & Griffin, D. (2019). When users control the algorithms: Values expressed in practices on Twitter. *Proceedings of the ACM on Human-Computer Interaction*, 3(CSCW), 138:1-138:20. <https://doi.org/10.1145/3359240>
- Canales, K. (2020, October 2). Twitter is changing how it crops photos after reports of racial bias. Business Insider. <https://www.businessinsider.com/twitter-change-photo-cropping-machine-learning-racial-bias-2020-10>
- Corbin, C. M. (2008). Mixed speech: When speech is both private and governmental. *New York University Law Review*, 85. <https://papers.ssrn.com/abstract=1114642>
- Cornelius v. Naacp Legal Defense & Ed. Fund, Inc., 473 US 788 (Supreme Court 1985).
- Corner, J. (1995). *Television form and public address*. Edward Arnold. <https://doi.org/10.1080/08821127.1996.10731823>
- Forums. (2021). LII / Legal Information Institute. <https://www.law.cornell.edu/wex/forums>
- Habermas, J. (1991). *The structural transformation of the public sphere: An inquiry into a category of bourgeois society*. MIT Press. <https://mitpress.mit.edu/books/structural-transformation-public-sphere>
- Intermediate Scrutiny. (2021). LII / Legal Information Institute. https://www.law.cornell.edu/wex/intermediate_scrutiny
- Knight First Amendment Institute v. Trump, 302 F. Supp. 3d 541 (Dist. Court 2018).
- Kulshrestha, J., Eslami, M., Messias, J., Zafar, M. B., Ghosh, S., Gummadi, K. P., & Karahalios, K. (2017). Quantifying search bias: Investigating sources of bias for political searches in social media. *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*, 417–432. <https://doi.org/10.1145/2998181.2998321>
- Mitchell, A., Jurkowitz, M., Oliphant, J. B., & Shearer, E. (2020, July 30). Americans who mainly get their news on social media are less engaged, less knowledgeable. *Pew Research Center's Journalism Project*. <https://www.journalism.org/2020/07/30/americans-who-mainly-get-their-news-on-social-media-are-less-engaged-less-knowledgeable/>
- Mongiello, A. B. (2019). Is President Trump violating the First Amendment when blocking citizens on Twitter?: Exploring Multi-Party Negotiation as a Way to Protect Citizens' Rights in the Wake of the New Digital Age Notes. *Cardozo Journal of Conflict Resolution*, 21(1), 217–246. <https://heinonline.org/HOL/P?h=hein.journals/cardcore21&i=227>
- Morales, A. (2020). A little bird told everyone but me: Establishing public forums in the twittersphere. <https://etda.libraries.psu.edu/catalog/17779aeg284>

- Mwengenmeir. (2014). Habermas' public sphere. *Media Studies 101*. BCcampus.
<https://opentextbc.ca/mediastudies101/chapter/habermas-public-sphere/>
- Packingham v. North Carolina, 137 S. Ct. 1730 (Supreme Court 2017).
- Papacharissi, Z. (2002). The virtual sphere: The internet as a public sphere. *New Media & Society*, 4(1), 9–27. <https://doi.org/10.1177/14614440222226244>
- Pariser, E. (2011). *The filter bubble: How the new personalized web is changing what we read and how we think*. Penguin.
- Perry Educ. Ass'n v. Perry Educators' Ass'n, 460 U.S. 37 (1983). (n.d.). Justia Law. Retrieved March 27, 2021, from <https://supreme.justia.com/cases/federal/us/460/37/>
- Rathnayake, C., & Suthers, D. D. (2018). Twitter issue response hashtags as affordances for momentary connectedness. *Social Media + Society*, 4(3), 2056305118784780.
<https://doi.org/10.1177/2056305118784780>
- Sap, M., Card, D., Gabriel, S., Choi, Y., & Smith, N. A. (2019). The risk of racial bias in hate speech detection. *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, 1668–1678. <https://doi.org/10.18653/v1/P19-1163>
- Teevan, J., Ramage, D., & Morris, M. R. (2011). #TwitterSearch: A comparison of microblog search and web search. *Proceedings of the Fourth ACM International Conference on Web Search and Data Mining*, 35–44. <https://doi.org/10.1145/1935826.1935842>
- The government speech doctrine. (2021). LII / Legal Information Institute.
<https://www.law.cornell.edu/constitution-conan/amendment-1/the-government-speech-doctrine>
- Tufekci, Z. (2014). Engineering the public: Big data, surveillance and computational politics. *First Monday*, 19(7). <https://doi.org/10.5210/fm.v19i7.4901>
- Tully, M., & Ekdale, B. (2014). Sites of playful engagement: Twitter hashtags as spaces of leisure and development in Kenya. *Information Technologies & International Development*, 10(3), 67–82.
https://www.researchgate.net/publication/265509607_Sites_of_Playful_Engagement_Sites_of_Playful_Engagement_Twitter_Hashtags_as_Spaces_of_Leisure_and_Development_in_Kenya
- Wiener, J. (2020). Social media and the message: Facebook forums, and First Amendment follies notes. *Wake Forest Law Review*, 55(1), [i]-244.
https://heinonline.org/HOL/Page?handle=hein.journals/wflr55&div=9&g_sent=1&casa_token=&collection=journals