Collaboration of Open Content News in Wikipedia: The Role and Impact of Gatekeepers

Ang Li
University of Pittsburgh
Pittsburgh, PA
ANL125@pitt.edu

Rosta Farzan
University of Pittsburgh
Pittsburgh, PA
rfarzan@pitt.edu

ABSTRACT

With the advances of social technologies, open content in social media has become an important place where people gather and communicate news. Prior studies in conventional mass media has long studied the news reporting process, and suggest that gatekeepers – editors or journalists who control the information and have a unique power to determine what gets published to the public – play an important role in the news reporting process. However, as the process of how open content are created by contributors in social media platforms is different, what we understand about content publication process in traditional mass media can not directly apply in the context of social media. Especially, it is unclear who are the gatekeepers and how do they influence the content creation and spread of information in social media. In the current proposed study, I aim to understand this new model of content generation process through the lens of gatekeepers in social media platforms such as Wikipedia. Specifically, I aim to discover ways to identify gatekeepers and assess their impact on information quality and content polarization.

CCS CONCEPTS

• **Information systems → Users and interactive retrieval**.

ACM Reference Format:
Ang Li and Rosta Farzan. 2020. Collaboration of Open Content News in Wikipedia: The Role and Impact of Gatekeepers. In *Companion Proceedings of the Web Conference 2020 (WWW '20 Companion), April 20–24, 2020, Taipei, Taiwan.* ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/3366424.3383568

1 INTRODUCTION

A series of major protests are happening in Hong-Kong. Hurricane Dorian causes 20 deaths in the Bahamas. The United Kingdom is withdrawing from the European Union. Events unfold as our days go on and attract attentions of many individuals across the world. In facing such breaking news events, people have traditionally relied on mass media to seek updated information about breaking news events by turning on the TV, listening to the radio, or checking the newspaper. Nowadays with the advantage of social technologies, there has been a significant shift in ways that people gather and communicate news, and increasingly open content in social media has become an important source of news. Based on a 2018 survey conducted by the PEW research center, 25% of the U.S. adults get their news via social media – this number is even slightly higher than the share who often do so from print newspapers (16%).[1]

Rich prior literature[19, 23] has long studied the news reporting process in conventional mass media (i.e. print, radio, TV), and provided an understanding of the journalistic news gathering and news publishing process in the mass media. As the conventional mass media has their inherent structure in the availability of column space, air time, or transmission frequencies[3], it is impossible for media outlets to present all information to audiences. Certain mechanisms or processes are established so that media outlets can control and select the information to be published according to specific criteria of newsworthiness[8]. However, the editors or journalists, known as media **gatekeepers**, may suppress or skew information in the process as news outlets usually have their own ideological preferences[5, 18, 23]. As a result, the skewed information can influence the perception of citizens about the relevant and importance of the news events[5, 30]. The Gatekeeping theory thus studies and investigates this mechanisms or processes through which the media control the construction and dissemination of information. **Gatekeeping** is defined as the process by which "*the billions of messages that are available in the world get cut down and transformed into the hundreds of messages that reach a given person on a given day*[20, 24, 32]." In this process, **Gatekeepers** – *the actors who craft and conduct what is being published to the masses* – are critical, as they decide what information the general public gets to view[24, 25, 32]. Gatekeepers not only control the **information quality** that they deem worthy to be distributed to people, but they also have a **unique power to influence audiences** by selecting and filtering certain information. As a result, they can determine what is to become the public's view of the world events with little involvement from the general public's opinions[24, 27].

In contrast to this traditional process, it has been argued that the process of how open content is created by contributors in social media, and especially in Wikipedia is different. The difference can be contributed to the following three dimensions of content generation in social media. First, unlike traditional mass media, open content in Wikipedia can be **created or edited by almost anyone**. As a result, the public is not just the content consumers but also the content creators. They have the opportunity to get involved in the content generation process. Second, Wikipedia encourages **collaboration** among contributors[7], which can facilitate and support a collective determination of what content is created and disseminated. Although the editors are distributed across different

[1] https://www.pewresearch.org/fact-tank/2018/12/10/social-media-outpaces-print-newspapers-in-the-u-s-as-a-news-source/

WWW '20 Companion, April 20–24, 2020, Taipei, Taiwan

ACM ISBN 978-1-4503-7024-0/20/04.
https://doi.org/10.1145/3366424.3383568

locations, those interested in one topic may come together under Wikipedia's principle to improve the same article by simply click the "edit" button to insert additional opinions or delete the content that they think is not appropriate from the current edition that is contributed by other editors.

In addition, the **organizational structure** of Wikipedia differs from that of traditional mass media. While traditional media production has often followed a hierarchical organizational structure, Wikipedia claims to have more of a flat structure and a different type of community norm. Prior literature has found that the organizational factors can influence journalists, especially shape their decisions on what should be published to audiences through information selection and curation[13, 26, 27]. Both Wikipedia editors and professional news journalists are seldom acting on their own, but they are largely embedded and influenced by the social context of organizations[28]. As a result, the process of how Wikipedians filter and curate news information can be different from the professional journalists in conventional media organizations due to the distinct organizational environment.

We argue that due to these fundamental differences that are key to the process of content generation, what we know about news reporting process in traditional mass media can not directly apply in the context of Wikipedia. Therefore, **the overarching research goal of my study** is to understand the process of open content generation about news events in Wikipedia by asking: **how open content is *collectively* and *collaboratively* created by the general public** – the group of people used to NOT able to involve in the process in traditional media.

Moreover, the different process in the open content news generation can also lead unknown content quality as a result for Wikipedia. Especially, in contrast to the traditional media where the news articles are majorly written based on the gatekeeping journalists' editorial decisions, the content creation process in Wikipedia is expected to be done through a more *inclusive and egalitarian process* – a process that is expected to generate inclusive content that represents diverse perspectives. On the other side, such an inclusive and egalitarian content generation process can also bring side effects – for example, as it has been argued more in the recent few years, this process may lead to low-quality and polarized content proliferated and disseminated across other social media platforms[1, 4, 15].

At the same time, although this inclusive and egalitarian open content generating process suggests that everyone has the same opportunity to contribute content, the reality is different. Only a small portion of Wikipedians contribute a majority of content, and most of the Wikipedians only contribute a little[21]. Therefore, in practice, not everyone and not every perspective is represented equally, and there are still certain users who are more powerful in determining what content that audiences will read in Wikipedia. In fact, one can argue that these contributors are the **Wikipedia gatekeepers** who play the role similar as gatekeepers in traditional media – they are more powerful in influencing the process of open content creation, determining the open content that the general audience eventually read in social media, and influencing readers' perspective. However, we know very little about what is the role of **Wikipedia gatekeepers** in this process to shape the content in Wikipedia.

In this study, I propose to focus on the **gatekeepers** in Wikipedia and to investigate **what is their role and how do they impact the process of collaborative content creating?**, and particularly, **how do the role of gatekeepers and the collaborative open content creation process relate to content polarization?**

2 BACKGROUND LITERATURE ON WIKIPEDIA

It has been demonstrated [6, 14, 17, 22, 31] that, when current events happen in the society, peer production platforms like Wikipedia serve as a place for many citizens to seek and share the latest information, to collectively make sense of what is happening, and build memories of the event. Pentzold claims that Wikipedia is a place for a global memory of events and describes its essential role in this dynamic process of constructing collective memories[22]. Especially, the popularity and openness of Wikipedia together with its detailed records enable general population to present and organize different sources, information and diverse perspectives into coherent narratives. A number of case studies have been conducted to understand how editors collaboratively produce knowledge as well as manage conflicts. For example, researchers have investigated the memory-building processes for the Egyptian revolution across languages[6], the deaths of notable people[14], the Vietnam War[17], and the Black Lives Matter social movements[31].

In addition, Wikipedia has long been considered as an open system that facilitates the *mass collaboration* among contributors. In addition, Wikipedia has strong *community principles* that all contributors should follow[2]. Although the Wikipedia editors are distributed across different locations, those who interested in one topic come together under Wikipedia's principle to improve the same article collaboratively.

3 RESEARCH METHOD

In the current study, I plan to focus on discovering ways to identify gatekeepers in Wikipedia and to quantitatively measure their power of gatekeeping. As discussed before, gatekeepers in the traditional mass media are the newsrooms or journalists who have the absolute power to determine the news content that is presented to the readers (i.e. general public). The traditional news audience is the "gated", bounded by the scope of events and perspectives presented in coverage – they rely mainly on the gatekeepers to get information and thus their perspectives are also largely influenced by them. However, in the context of online social media where newsrooms or journalists do not hold absolute power of gatekeeping any more, the general news audience (i.e. those being "gated" in traditional mass media) can also be involved in this process and create measurable impact to influence other's perspectives – they can produce and broadcast their own opinions, to choose various content, and to even interact with traditional gatekeeping newsrooms or journalists and influence them.

The theory of Networked gatekeeping emphasizes this shifted power dynamics in the digital age and describes the relationships between news actors who hold diverse levels of power and positions. Particularly, the theory suggests that the power of "gated" (i.e.

[2]https://en.wikipedia.org/wiki/Wikipedia:Expectations_and_norms_of_the_Wikipedia_community

news audiences) in the networked gatekeeping process can be described by four attributes – their opinion leadership, information production ability, relationship with traditional gatekeepers, and their autonomy to choose diverse information[2, 33]. Therefore, in the current study, *I plan to operationalize the four attributes proposed by Network gatekeeping theory and quantify the power of social media gatekeepers for the open content contributors.*

To do so, a mixed-method study including machine learning techniques and qualitative content analysis will be conducted, aiming to not only discover ways in identifying gatekeepers in different social media platforms but also to evaluate the proposed method. Specifically, I will first collect a data corpus that represents news events related *open content* and their associated *open content contributors* in Wikipedia. I will then operationalize the four attributes discussed previously (i.e. the opinion leadership, information production ability, relationship with traditional gatekeepers, and their autonomy to choose diverse information) as a set of *variables* to infer each contributors' gatekeeping power in Wikipedia.

Before modeling contributors' power of gatekeeping, one challenge remains: we don't have a ground truth for each contributor's gatekeeping power in social media. Following the definition of social media gatekeepers – they are the users who are more powerful in determining the open content that the general audience eventually read in social media, and influencing readers' perspective. As a result, the content created by powerful gatekeepers in social media should reach and influence more readers, and should stay in the open content for a longer time. Therefore, the *influence of the content* created by each contributors – how many readers the content can be eventually reach and like, how long time the content still stays and being transmitted – can be a proxy for the outcome of the gatekeeping. I plan to use the *content influence* as outcome variable of gatekeeping, and check their relationship with the *measures of gatekeeping* proposed before as independent variable. The *hypothesis* is that content contributed by users with more gatekeeping powers should reach more readers, and be able to stay in open content for good amount of time. I will test this hypothesis to validate the proposed measures.

To further evaluate the validity of the proposed variables, I plan to also adopt qualitative methods. Especially, ground truth of who are gatekeepers can be obtained through human annotation of a small random sample of open content contributors. In addition, qualitative content analyses can be conducted to further examining and characterizing the identified gatekeepers and their gatekeeping strategies.

4 EXPECTED RESULTS AND DISCUSSION

The current study is expected to identify a group of gatekeepers in Wikipedia who have more influential power to determine the news open content that most audiences read. Future studies can be done to investigate the role of gatekeepers on the process that contributors collaboratively create open content in Wikipedia, especially, to study how does the power of gatekeepers in relation with the open content quality and polarization in document news events.

News events are unique as they are highly popular and usually involve controversial topics. As identified by many prior studies, reporting them can be polarized in traditional mass media as different media outlets hold different preferences or political stands[10, 11, 29], and this polarization has been a source of many debates in recent years. Meanwhile, news events can also attract the diverse interests from the general public across the world to collectively build open content around them in social media platforms. Such diversity can play a critical role in achieving the goal of "Neutral point of view" – one of the Wikipedia's major principles. The results of current study can help future studies to provide a set of empirical evidence on whether or how the collective model of news related open content creation process can serve as a step towards a solution to this critical issue.

The world has become more divided than ever before, especially as we face controversial events that are so common in our daily lives. A biased perspective can easily be framed by selecting and curating the information that supports that perspective, and the general public is constantly being influenced by the biased information they are exposed to. Recently, many researchers have also noticed this critical issue, and most of their work focus on discovering ways to identify or detect content bias or polarized opinion (i.e. [4, 9, 12, 16]). However, instead of focusing on detecting content bias, my proposed work focuses on understanding the content generation process and the underlying mechanism, which can be more constructive in both understanding the polarization and a solution to it. Without a better understanding of how the content is selected, curated and spread, the society is in danger of getting more polarized to such an extent that reconciliation between opposing perspectives will become even more difficult.

Therefore, I pursue my study as a step towards solving this critical issue by providing guidelines for practitioners and designers of social media systems that aim to present their users a diverse perspective with unbiased information. Overall, the study will have implications for information scientists, technology developers and community practitioners seeking to build social media platforms that can generate inclusive information that is both high quality and represents diverse perspectives.

REFERENCES

[1] 2009. Users say they regularly encounter false and misleading content on social media. *Pew Research Center* (2009). https://www.pewresearch.org/internet/2019/05/13/users-say-they-regularly-encounter-false-and-misleading-content-on-social-media-but-also-new-ideas/
[2] Karine Barzilai-Nahon. 2008. Toward a theory of network gatekeeping: A framework for exploring information control. *Journal of the American society for information science and technology* 59, 9 (2008), 1493–1512.
[3] Axel Bruns. 2003. Gatewatching, not gatekeeping: Collaborative online news. *Media International Australia Incorporating Culture and Policy* 107, 1 (2003), 31–44.
[4] Michael D Conover, Jacob Ratkiewicz, Matthew Francisco, Bruno Gonçalves, Filippo Menczer, and Alessandro Flammini. 2011. Political polarization on twitter. In *Fifth international AAAI conference on weblogs and social media.*
[5] Jessica T Feezell. 2018. Agenda setting through social media: The importance of incidental news exposure and social filtering in the digital era. *Political Research Quarterly* 71, 2 (2018), 482–494.
[6] Michela Ferron and Paolo Massa. 2011. Collective memory building in Wikipedia: the case of North African uprisings. In *Proceedings of the 7th International Symposium on Wikis and Open Collaboration.* ACM, 114–123.
[7] Andrea Forte and Cliff Lampe. 2013. Defining, understanding, and supporting open collaboration: Lessons from the literature. *American Behavioral Scientist* 57, 5 (2013), 535–547.
[8] Johan Galtung and Mari Holmboe Ruge. 1965. The structure of foreign news: The presentation of the Congo, Cuba and Cyprus crises in four Norwegian newspapers. *Journal of peace research* 2, 1 (1965), 64–90.
[9] Kiran Garimella, Gianmarco De Francisci Morales, Aristides Gionis, and Michael Mathioudakis. 2018. Quantifying controversy on social media. *ACM Transactions*

on Social Computing 1, 1 (2018), 3.
[10] Matthew Gentzkow and Jesse M Shapiro. 2010. What drives media slant? Evidence from US daily newspapers. *Econometrica* 78, 1 (2010), 35–71.
[11] Tim Groseclose and Jeffrey Milyo. 2005. A measure of media bias. *The Quarterly Journal of Economics* 120, 4 (2005), 1191–1237.
[12] Pedro Calais Guerra, Wagner Meira Jr, Claire Cardie, and Robert Kleinberg. 2013. A measure of polarization on social media networks based on community boundaries. In *Seventh International AAAI Conference on Weblogs and Social Media*.
[13] Thomas Hanitzsch and Claudia Mellado. 2011. What shapes the news around the world? How journalists in eighteen countries perceive influences on their work. *The International Journal of Press/Politics* 16, 3 (2011), 404–426.
[14] Brian C Keegan and Jed R Brubaker. 2015. 'Is' to'Was': Coordination and Commemoration in Posthumous Activity on Wikipedia Biographies. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*. ACM, 533–546.
[15] Jae Kook Lee, Jihyang Choi, Cheonsoo Kim, and Yonghwan Kim. 2014. Social media, network heterogeneity, and opinion polarization. *Journal of communication* 64, 4 (2014), 702–722.
[16] Haokai Lu, James Caverlee, and Wei Niu. 2015. Biaswatch: A lightweight system for discovering and tracking topic-sensitive opinion bias in social media. In *Proceedings of the 24th ACM International on Conference on Information and Knowledge Management*. ACM, 213–222.
[17] Brendan Luyt. 2016. Wikipedia, collective memory, and the Vietnam war. *Journal of the Association for Information Science and Technology* 67, 8 (2016), 1956–1961.
[18] Maxwell E McCombs and Donald L Shaw. 1972. The agenda-setting function of mass media. *Public opinion quarterly* 36, 2 (1972), 176–187.
[19] Denis McQuail. 1987. *Mass communication theory: An introduction.* Sage Publications, Inc.
[20] Denis McQuail. 1994. *Mass communication theory: An introduction.* Sage Publications, Inc.
[21] Katherine Panciera, Aaron Halfaker, and Loren Terveen. 2009. Wikipedians are born, not made: a study of power editors on Wikipedia. In *Proceedings of the ACM 2009 international conference on Supporting group work*. ACM, 51–60.
[22] Christian Pentzold. 2009. Fixing the floating gap: The online encyclopaedia Wikipedia as a global memory place. *Memory Studies* 2, 2 (2009), 255–272.
[23] Eugene F Shaw. 1979. Agenda-setting and mass communication theory. *Gazette (Leiden, Netherlands)* 25, 2 (1979), 96–105.
[24] Pamela J Shoemaker. 1991. Gatekeeping. (1991).
[25] Pamela J Shoemaker, Martin Eichholz, Eunyi Kim, and Brenda Wrigley. 2001. Individual and routine forces in gatekeeping. *Journalism & mass communication quarterly* 78, 2 (2001), 233–246.
[26] Pamela J Shoemaker and Stephen D Reese. 1996. *Mediating the message.* White Plains, NY: Longman.
[27] Pamela J Shoemaker and Timothy Vos. 2009. *Gatekeeping theory.* Routledge.
[28] Jane B Singer. 2004. More than ink-stained wretches: The resocialization of print journalists in converged newsrooms. *Journalism & Mass Communication Quarterly* 81, 4 (2004), 838–856.
[29] Daniel Sutter. 2000. Can the Media be so Liberal-The Economics of Media Bias. *Cato J.* 20 (2000), 431.
[30] Yue Tan and David H Weaver. 2013. Agenda diversity and agenda setting from 1956 to 2004: What are the trends over time? *Journalism Studies* 14, 6 (2013), 773–789.
[31] Marlon Twyman, Brian C. Keegan, and Aaron Shaw. 2017. Black Lives Matter in Wikipedia: Collective Memory and Collaboration Around Online Social Movements. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing (CSCW '17)*. ACM, New York, NY, USA, 1400–1412. https://doi.org/10.1145/2998181.2998232
[32] David Manning White. 1950. The "gate keeper": A case study in the selection of news. *Journalism Bulletin* 27, 4 (1950), 383–390.
[33] Weiai Wayne Xu and Miao Feng. 2014. Talking to the broadcasters on Twitter: Networked gatekeeping in Twitter conversations with journalists. *Journal of Broadcasting & Electronic Media* 58, 3 (2014), 420–437.