Original Research

Understanding Political Polarization Based on User Activity: A Case Study in Korean Political YouTube Channels

Giang T. C. Tran[1], Luong Vuong Nguyen[1], Jason J. Jung[1] (ID), **and Jeonghun Han[2]**

Abstract

This study proposes a novel approach for measuring political polarization using a user-activity-based model. By exploiting data from comments, user activity in this study is defined based on features such as coverage, duration, and enthusiasm. To determine these features, we collect information on the activities of users from South Korean YouTube channels. Notably, the collected data of the model contains approximately 11 M comments from more than 600 K users based on 37 K videos of 77 YouTube channels. To handle the big data collection, we deploy a web-based platform called TubePlunger to collect video information (e.g., comments, replies, etc.) automatically from YouTube channels. The output of the model reveals that the users are strongly polarized because the number of neutral users is very small (approximately 8% of the total). We then applied this model to the other channels in the testing dataset to define polarization with a bias percentage and to visualize the user activity distribution. The experimental results show that there are 30 fully polarized YouTube channels (16 left-wing channels and 14 right-wing channels) with a measured bias ratio higher than 70%. Our method of analyzing social network data based on user activity provides the foundation for polarization analysis that can be applied to fields other than politics.

Keywords

user modeling, information retrieval, social networks, political polarization

Introduction

Polarization in countries with a multiparty system is not an uncommon political issue. It is widely regarded as an integral aspect of politics. It usually refers to an ideological interaction among parties. A significant issue associated with measuring polarization is the lack of consistency in defining this term and the quantifying the degree of polarization. In general, the term *polarization* is used to suggest that the political spectrum has divergent ideological views. Political polarization has serious implications in the lives of people, and its essence is still not fully understood (Morales et al., 2015). In politics, the polarization happens in tendencies, opinions, or thoughts, with many aspects and it exists in a numerous forms. Furthermore, political polarization has the ability to appear under user activity both online and offline; user activity can be collected and analyzed further. There has been considerable research on this topic from different points of view.

In multiparty systems, political bias assumes a straightforward form. For instance, political parties differentiate themselves from each other and embrace increasingly different ideological positions; thus, a compromise is often impossible.

In assessing polarization in multiparty systems, this finding has led to a focus on ideological distance (Kam et al., 2017). In addition, the ideological gap between the left-wing and right-wing was considered. However, increasing attention has been paid to less extreme forms of activism that may precede involvement in political activity (Bright, 2018).

Online social networks have seen an explosive growth over the last decade (Nguyen & Jung, 2020; Nguyen, Nguyen, & Jung, 2020). Individuals now use social networking platforms to obtain news and information instead of using conventional outlets. Therefore, traditional media, including radio, broadcast, television, and print, have been partly replaced by YouTube, Twitter, and Facebook (Amedie, 2015). YouTube is an important online video platform for discourse and the exchange of opinions and knowledge. The

[1]Chung-Ang University, Seoul, Republic of Korea
[2]Seoul National University, Republic of Korea

Corresponding Author:
Jason J. Jung, Chung-Ang University, 84 Heukseok-Ro, Dongjak-Gu, Seoul 06974, Republic of Korea.
Email: j2jung@gmail.com

website allows anyone to be a content creator and to participate instantly in a trending debate or conversation instantly. Therefore, YouTube has become a critical news source. By exploiting this platform's numerous benefits, governments and politicians have established accounts, resulting in the formation of politically homogeneous communities that are connected by the same ideology (Effing et al., 2011). Apart from the official channels of politicians or parties, YouTube has many loosely associated groups of channels and personalities. They include performers, vloggers, and essayists, and they usually publish personal discourses or speeches to share particular political opinions. YouTube also provides an online political debate space for sharing and propagating events, speeches, or campaigns. The content of each sharing video and the corresponding comments from users result in the YouTube channel obtaining a political bias. Political bias in the media occurs when the media emphasizes particular points of view, conveys selected information to further their own political view, and presents only information favoring their own political opinion (Haselmayer et al., 2017). The media also show social leaders discuss political issues with a bias toward their own political view (Strömberg, 2004). Media polarization also influences the decisions of people, as well as public opinion. Thus, an understanding of user activity and recognition of the degree of political bias in these channels is necessary for both governments and politicians, especially during elections. In this study, YouTube was used as a platform for experiments to better understand political polarization.

This study aims to improve the understanding of user activity by measuring political polarization on YouTube. In particular, we propose a model to determine user activity on YouTube channels using data collected via an API key. The first result of the model is the percentage of users of the channel that are left-wingers or right-wingers. In addition, our model also reveals the user activity distribution of each channel based on three features: coverage, duration, and enthusiasm. We apply the proposed methodology, which can measure the degree of political polarization based on user activity, to the YouTube channel in South Korea. The results from our model reveal the differences in activity between online communities in a YouTube channel. South Koreans are shifting toward a moderate political identity based on the results of a survey on national political identity conducted by the combination of Hankyoreh and the Korean Social Science Data Center (Lee, 2016). Therefore, political attitudes remain comparatively moderate (Kim et al., 2012). However, this study provides evidence of political polarization. The abundance of media options reduced the percentage of less enthusiastic and less partisan electors and increased the polarization of elections (Prior, 2013).

Nevertheless, the evidence of a causal link between more partisan messages and changing attitudes or behaviors is ambiguous at best (Prior, 2013). The determination of the bias that discriminates expression in politics and its outcome is inhibited by measurement issues. The segment of society that can be biased is high, even though only a few individuals access one-sided information. There is no clear indication that biased media cause ordinary South Koreans to become more polarized. Korean political parties are mainly divided into two groups: those that support the left-wing (liberal) and those that support the right-wing (conservative; Steinberg & Shin, 2006). For these independent Korean groups, to gain political support, lawmakers, voters, and the government need to consider the ideas and actions of users. In addition to the standard outlets that we already know belong to the left-wing or right-wing, there are still several other channels with vaguely expressed political stances. In this study, we assume that there is an intense polarization in YouTube users between the left-wing and right-wing. The political bias of a channel can be determined by viewing videos or reading associated comments. However, this method is time-consuming because of the large number of videos. Using the bias percentage obtained from the results of the proposed model, the political bias of YouTube channels can be determined without considering the content of the videos and comments.

Political polarization also has negative consequences in terms of creating an extreme political ideology, which restrict discussions and exchanges with the opposing side. Therefore, understanding this polarization is the first and essential step in the progress of democratic consolidation. Additionally, being cognizant of political cleavage will help governments and leaders of parties in decision-making. Our findings can provide information to researchers, governments, and politicians who strive to understand user activity in the context of politics. Democracy has made society more polarized with increased disagreement when there were many conflicting opinions. Understanding user activity in the context of politics, governments, and politicians has some advantages, such as having the appropriate strategies in management or dominating debates. This is also an interesting and potential topic for researchers to mine user behavior and observe their characteristics. With the exploration of social networks, we try to exploit YouTube, which has a considerable amount of user activity data, as our data source in this study. The current study examines two research questions that emerged from prior research in the area of big data analytics as follows.

1. RQ (1): How polarized are the online political communities on YouTube?
2. RQ (2): Is there any political polarization in other kinds of online communities on YouTube?

We examine whether political bias exists on YouTube, even though most people maintain that they are well balanced. In particular, we chose South Korean YouTube channels as the data source for the experiments. The study not only contributes to the determination of the degree of political bias but also the representation of user activity distribution in YouTube channels. In addition, the output results

from the proposed model were used to classify the channels into left-wing and right-wing without considering the content of the videos and comments. This study proposed a novel method of measuring the level of polarization in political YouTube channels and other kinds of YouTube channels without considering the content of videos as well as the content of comments by our hypothesis with the bias score for each channel. Additionally, we suggest a new way of analyzing user activity data from YouTube in comparison with relevant studies. There are many articles researching political polarization, but most of them focus on other social networks such as Twitter or Facebook; there are few studies focusing on YouTube. In addition, we show the user activity distribution corresponding to the left-wing and right-wing in each channel to provide new observations of a comprehensive view. The proposed method can be applied to other types of social networks using the three features of coverage, duration, and enthusiasm. In this study, we assume that there is intense polarization between the left-wing and right-wing YouTube users in South Korea. Moreover, users tend to watch, share, and discuss their ideas and opinions only on the YouTube channels that they support.

The remainder of this paper is organized as follows. The background and related works on political polarization and user modeling are briefly discussed in the next section. In the subsequent section, the goals and challenges of big data on user activity are reviewed, followed by a review of user activity modeling to understand political polarization. Experimental results are then analyzed and explained. In the final section, we present our conclusions and directions for future research.

Background and Related Work

Polarization is one of the most well-known and discussed issues in countries with multiparty systems. It is a social phenomenon associated with extreme and contradictory positions as individuals align their views, with few people maintaining neutral or moderate opinions based on a sociological perspective (Garibay et al., 2019). Therefore, the evaluation of diverse matters, such as politics and religion, results in separation over time. According to the Law of Group Polarization: *Like polarized molecules, group members become even more aligned in the direction they were already tending* (Sunstein, 1999). Several recent studies have explored the conceptualization and quantification of polarization, or the prediction of the political bias of social media users. For example, Malouf and Mullen (2007) used many current natural language processing techniques to learn how to identify users of a US political discussion site. Preoţiuc-Pietro et al. (2017) attempted to construct a fine-grained model that could classify the political ideologies of US Twitter users into seven groups using a manually labeled dataset. Dalton (2008) presented a relationship between the magnitude of a district and its polarization-variance-based calculation. Curini and Hino (2012) discussed some contradictory findings after

investigating the theoretical effects on the polarization of electoral systems and heterogeneity. Layman et al. (2006) compared the prevailing session with previous political time and studied partisan transformation theories to examine the theoretical and experiential literature on party bias and partisan movement. In South Korea, Park et al. (2011) examined methods for predicting political parties in South Korean news articles using the sentiment trends of active commentators. Moreover, Joo et al. (2016) recommended a model to recognize political bias using document embedding and deep learning algorithms in South Korean news articles.

Many individuals worldwide use social media to express thoughts and ideas about their daily lives, as well as their general ideas and opinions. Digitalization and mediatization have changed how we connect and engage in political debates over the past few decades. Political debates may have traditionally taken place during lunch breaks or around the dinner table. These debates have entered the world of public, computer-mediated communication, owing to the rise of digital social media. The increasing role of digital media in structuring public relations is related to the current concern about the growth of divisive political discourse (Van Aelst et al., 2017). The increasing exposure to interactive political talks on social media has shifted academic attention to the potentially polarizing impact of organizations. This practice is based on a conventional emphasis on polarization as the product of political, ideological, and partisan identities (Settle, 2018). Scholars have developed new approaches to analyze the polarization of digital media. Statistical methods are generally used to extract social network structures, semantic content, emotions, and other properties from rich social media data. In this era of big data, political polarization can be detected in various ways. Digital traces of collective human activity can be used to identify and quantify various phenomena, such as polarization. Groeling (2008) examined political bias by contrasting situations in which newscasts announced presidential approval ratings, calculated based on their polls for Bill Clinton and George W. Bush. Polarization of attitudes due to divergent views in partisan media is a common consequence in the Receive-Accept-Sample model. Zaller's hypothesis, which has recently been applied to the study of media effects, has provided empirical examples of viewers attitudes polarization (Levendusky, 2013). In addition, Groseclose and Milyo (2005) proposed a novel method for rating media outlets and congress members. In news coverage and congressional records, they used the relative frequency of references to various studies to connect outlets and lawmakers. Political segregation has already been observed in political blogs and YouTube. Recent research has demonstrated that the most influential and politically engaged users communicate mostly with their allies, leaving no space for real dialogue and cross-ideological interactions. However, separation does not inherently mean division, because it is impossible to consider two different groups of individuals that hold the same viewpoint to be polarized. Therefore, to

polarize society, the views of the two classes should also be contradictory or opposed. In the later part of this paper, we describe how to adapt our methods to online data obtained from YouTube to quantify political polarization. Experiments were conducted on YouTube channels because they offer an exciting environment for polarization research. Moreover, YouTube reflects a wide range of diverse modes of communication, from individual to conventional mass media.

Relevant to user modeling, the huge amount of data obtained from social networks has motivated researchers to explore user behavior in terms of a specific domain. Using information on the number of followers, the number of retweets, and the number of mentions that were collected from Twitter, researchers in one study calculated the power of Twitter users (Cha et al., 2010). To create a user interest model, Stoyanovich et al. (2008) used the tags created and the users' social friends. Additionally, Benevenuto et al. (2009) used clickstream data for user behavior comparisons among different social media. Papagelis et al. (2011) showed the causality between individual behavior and social influence after examining user information. The experiments were conducted on YouTube to examine user behavior. To explore customer satisfaction and understand user behavior on this website, the relationship between content viewing, content production, and YouTube addiction has been investigated by Balakrishnan and Griffiths (2017). Preoțiuc-Pietro et al. (2017) found that a highly correlated relationship was shared by four popularity metrics from view count, number of comments, number of ratings, and number of favorites. YouTube users transfer more information and have longer thinking times than typical web workloads based on the length of the session, inter-transaction times, and the types of content transmitted by users (Gill et al., 2008).

In this study, we focus on exploiting data from user activity on social networks, or in YouTube, particularly. As South Korea is a democratic republic country, researching polarization in politics is also an interesting topic. The mentioned relevant studies serve as sources of reference for us to identify the direction for our research with the purpose of adding a new way of exploring user activity on social networks without spending a huge amount of time considering the contents. We developed a model for measuring the degree of political bias by exploiting user activity in a channel without considering the content of videos and comments. The model includes three features (which are coverage, duration, and enthusiasm) to explore the meaning of user activity on YouTube. We applied the model to a South Korean political YouTube channel case study to better understand user behavior and explore specific research questions.

Big Data on User Activity

Significant quantities of data are created by and about individuals, items, and interactions, thereby creating a big data era. Big data provide a contemporary society with potential experiences that have been conducted in several studies. In terms of politics, there are classifying the political ideologies of US Twitter users into seven groups by a fine-grained model (Preoțiuc-Pietro et al., 2017), recognizing political bias using document embedding and deep learning algorithms in South Korean news articles (Joo et al., 2016), or exploring customer satisfaction and understand user behavior on this website, the relationship between content viewing, content production, and YouTube addiction (Balakrishnan & Griffiths, 2017). Moreover, big data generate a variety of challenges for data scientists. In the case of small-scale data, big data have significant potential to uncover subtle community variations and unimaginable interactions. However, the massive sample scale and enormous complexity of big data present specific statistical and mathematical difficulties, including constraint optimization and storage, noise deposition, spurious correlation, incidental endogeneity, and measurement errors (Brady, 2019). In particular, there are challenges in terms of data analysis and interpretation. The first difficulty is integrating the datasets into a platform for research. This creates holes, and incorrect responses and observations may be obtained if this problem is ignored. In addition, with the exponential growth in data, online social networking (OSN) has generated an enormous demand for massive data scientists and massive data analysts (Bello-Orgaz et al., 2016). In the incorporation of methods and technologies, big data analytics is a key term for collecting and administering large structured and unstructured data in real time. Big data analytics operates on the entire dataset to sample conventional data analysis systems (Jung, 2017).

In the past decade, the number of web users worldwide has increased (Bradshaw, 2001) and is closely associated with the success of online social networks such as Facebook, Twitter, and LinkedIn. Individuals may participate directly in these networks, create their own networks of friends, and share their views, insights, knowledge, experiences, and perspectives (Nguyen, Hong, Jung, & Sohn, 2020). They can also discover and transmit information using different formats, including words, images, audio, and videos. Social interaction creates a huge amount of information, making it a user activity resource. Each user interaction on a social network is stored and considered to be personal data; this allows us to analyze and understand user behavior in a specific domain. At present, most major social networking platforms operate by storing and recalling user behavior. This results in the generation of a large amount of user activity data that are stored online and are easy to crawl in different ways. Understanding and analyzing user activity on social networks has many benefits with respect to formulating data models (Piccialli & Jung, 2017). More importantly, interactions between humans and non-human artifacts have significantly increased the productivity of data scientists. Big data analytics can build on the knowledge of a crowd, reveal patterns, and deliver best practices.

Research on social influence has become widespread in society (Saulwick & Trentelman, 2014). The high-level purpose of studies on social influence is to respond to questions such as *(1) Who can be influenced? (2) Is someone willing to control someone else? (3) Who is prone to influence? (4) Why is a specific community attractive to users? (5) Who are the most relevant social network users?* The core concept of social influence analysis is to determine how the influence of any user can be calculated, and how the most active users of social networks can be defined. To this end, we aim to understand user activity in social networking to generate an appropriate model. Specifically, we focus on exploiting user activities on YouTube, which is one of the most visited websites for online video sharing. In addition, YouTube has been the most relevant mass media in the last decade, which warrants scrutiny from academics (Bärtl, 2018). YouTube maintains accurate logs of all user experiences and openly shares some of the results, which significantly increases the quantitative analysis capacity.

Modeling User Activity for Understanding Political Polarization

Political Polarization Landscape in South Korea

Political polarization is common in democratic societies and has become a topic of academic debate. It took several centuries for democratic politics to emerge in the West. In South Korea, democracy was introduced in 1948, even though the foundations were quite weak (H. Kang & Yang, 2020). Korean democracy has emerged from a less than ideal start, as the basic rights and freedoms of individuals have been abused . There has been much debate during the late stages of the dynasty on the modern institutional system, legislative development, constitutional implementation, and republican law. The Provisional Government of the Republic of Korea chose republican democratic governance as the fundamental framework for the future government of Korea after the 1st March Movement, 1919, (Suh, 2007). This event marks one of the first public displays of Korean resistance during the Japanese occupation of Korea. The day is commemorated as a public holiday in South Korea. Of the many political parties in South Korea, some were short-lived, whereas others split from or merged with other parties. In general, the parties can be split into four main left-to-right ideological spectra named progressive, liberal, centrist, and conservative.

South Korea has attained partial democracy with direct presidential elections since 1987. In terms of politics, both the governments and the National Assembly have taken a higher level of responsibility to appearing clean to the public eye and providing social welfare. Some organizations that provide services related to social issues such as the environment, human rights, or minority issues are politically oriented by engaging in public debates. Specifically, the emergence of democracy has made society more polarized with the airing of increased disagreement when there are many conflicting opinions. More importantly, political flux and instability derived from the current "transitional politics" have aggravated the political polarization. An absolute majority of Korean voters are concerned about the growing polarization in their country.

South Korea has made considerable progress in creating a democratic political structure, considering its relatively brief history. The development of the partisan division between the left-wing (liberal) and right-wing (conservative) characterizes the current Korean political system (Steinberg & Shin, 2006). Conservatives are typically more concerned with economic development and market principles, whereas liberals are more concerned with the fair distribution of income and human rights. Conservatives were in power in the government during the repressive regimes from 1948 to 1987, and they retained a parliamentary majority. The conservative and liberal have taken turns in forming the government approximately every 10 years over the last three decades since South Korea's formal democratization in 1987, whereas centrist and progressive parties have constituted minorities in the parliament. Over the decades, many political parties in South Korea have started and then disappeared and that changed their names and their leaders, but left-wing and right-wing have remained as the most prominent ones.

It is generally acknowledged that democracy is supported by the young and liberal middle class in the southwest of South Korea, whereas conservatism is typically advocated by the old and conservative upper or working classes in the southeast. Regional competition between the southwest and southeast is deeply ingrained in South Korean politics. However, many experts maintain that regionalism, especially in the southeast, has weakened since the last general election in 2016 (H. Kang & Yang, 2020). Currently, political divisions in South Korean society seem to be more influenced by age and ideology than regionalism. Conflicts between the younger and older generations are deepening with respect to the issues of housing, jobs, childbirth, pension, and so on. In terms of human rights issues concerning refugees and homosexuality, South Korean politics is divided more by left–right ideological positions (W. T. Kang, 2008). Political polarization also has negative consequences in terms of fostering extreme political ideologies that limit debates and exchanges with opposing viewpoints. As a result, comprehending this polarization is the first and most important step toward democratic consolidation. Additionally, being aware of political divisions will aid governments and party leaders in making decisions.

Comparing the perceived polarization with the relationship between different news sources may not help to generalize the results of one country to another socio-political context. All media systems have their own relationships with political systems; hence, the same media platform can play different roles in providing information to citizens of different countries (Yang et al., 2016). Partisanship is the most

prominent ideological driver of polarization within research, and with the appearance of digital epoch, there are contributions to polarized political scene from ideological elements on the OSNs (Guan et al., 2021). Both partisanship and polarization are concentrated in Western countries, and South Korea is a non-Western country, but South Korea is still severely respected by Western scholars through social movements under the influence of democratic change. In addition, South Korean social movements have gained strong political influence by focusing on causal factors such as the composition of key political actors during the transition and the dynamics of relationships between social movement groups (Kwak, 2012). By choosing South Korea as a particular case and proposing the analysis of user activity in YouTube channels of this country, we aim to supplement our understanding of the deep ideological divides present not only in Western democracies but also in Eastern ones.

In this study, we assume that there is a strong polarization between the two political camps, left-wing and right-wing, in the Korean YouTube channels. This means that most of the users who interact on YouTube tend to focus on and support channels that are affiliated with their political bias. To prove this hypothesis, we propose a model to better understand user activities on YouTube channels. The channels used for the experiments were mainly focused on political issues. This model is then used to classify channels into conservative and liberal groups. Moreover, the level of user activity in each channel corresponding to each political bias was measured using this model.

Modeling User Activity

Individuals have become more biased and extreme in their opinions owing to the explosion of partisan news media platforms and strong partisanship. As a result, there are differences in user activity in OSN. After extensive viewing of certain videos and leaving comments, some users exhibited an intense fervor in their support of specific channels. In contrast, they do not stay long or are unwilling to interact with channels with opposing views. Thus, exploring and understanding user behavior in OSNs is extremely important in terms of the demands of governments and politicians. One of the main challenges is to analyze and understand an enormous dataset because users exhibit different actions on YouTube. These actions include not only viewing but also assigning like or dislike ratings, leaving comments, and subscribing. In this study, we proposed a model to investigate the extent to which users engage with YouTube channels. The three features in the model are coverage, duration, and enthusiasm, which are defined as follows.

Definition 1. Coverage. *Coverage is the number of videos to which the user comments on and replies. Because we cannot identify the watched videos of a user, it is possible to determine the list of videos that a user comments on or*

replies to. Given that user u and $v \in V$ is a video in the set of videos V in channel A, the coverage of user u is calculated as

$$\mathbb{C}(u) = \sum_{v \in V} f(v), \tag{1}$$

where

$$f(v) = \begin{cases} 1, & \text{if video } v \text{ has comment(s) from user } u \\ 0, & \text{otherwise} \end{cases}. \tag{2}$$

Definition 2. Duration. *The duration of user u is the active period determined from the history of the comments, counted each day. The duration feature has a considerable contribution in terms of measuring user activity because the longer the duration of engagement with YouTube, the more substantial the support of the user. The formulation for calculating the duration of user u is as*

$$\mathbb{D}(u) = |D_i - D_j|, \tag{3}$$

where D_i denotes the day that user u has the first activity, and D_j denotes the day that user u has the last activity.

Definition 3. Enthusiasm. *Enthusiasm is determined by the total number of comments posted by the user in the videos of a channel. The enthusiasm feature is not less important when directly showing the fervor of the user. An individual who posts frequent comments on their favorite channels may be considered an enthusiastic supporter. Given that $m \in M$ is a comment in the set of all comments M in channel A, the formula used to calculate the enthusiasm of user u can be expressed as*

$$\mathbb{E}(u) = \sum_{m \in M} f(m), \tag{4}$$

where

$$f(m) = \begin{cases} 1, & \text{if comment } m \text{ belongs to user } u \\ 0, & \text{otherwise} \end{cases}. \tag{5}$$

These features were used for the experiments in this study to show to the extent possible the activity of YouTube users after we observed the characteristic of our database. Coverage is the number of videos on which a user has left the comment(s). Duration is the period in which a user is active on YouTube in the context of this study. Enthusiasm is the total number of comments a user makes. Each user has a different way of using YouTube, as well as a different way of showing their interaction with YouTube channels. There are users with high scores for all three features, and there are users with high scores in one or two of them. Therefore, we attempt to determine the engagement of a user by using these three features.

According to these three defined features, all users of the 71 YouTube channels of the testing dataset were modeled based on their activity. The results were then used to measure

the degree of polarization of each channel and to classify these channels into left-wing and right-wing groups. As a result, the differences between the two political biases in each channel were discovered. For example, considering a new $Channel_A$, the classification and measurement of user activity are as follows.

- Step 1: Dividing users of $Channel_A$ into two groups (left-wing and right-wing) by matching with the list of users in each political polarization from the sample dataset.
- Step 2: Obtaining each feature value of all users corresponding to the left-wing and right-wing groups for the given $Channel_A$.
- Step 3: Identifying the political bias and the degree of polarization associated with a given $Channel_A$ based on the average score of the total number of users.

To separate the list of users ($List(A)$) in $Channel_A$, we require a list of left-wingers (L_0) and right-wingers (R_0) from the sample dataset. The list of left-wingers/right-wingers was obtained by identifying all the users who commented on videos in the left-wing/right-wing channels only. Hence, individuals who commented on both left-wing and right-wing channels were excluded from the scope of this research. After obtaining the standard list of users for the left-wing and right-wing groups, $List(A)$ of $Channel_A$ is divided into L_A and R_A by matching with (L_0) and (R_0) as

$$List(L_A, R_A) = List(A) \cap List(L_0, R_0), \qquad (6)$$

where L_A and R_A represent the list of users who support the left-wing and right-wing in $Channel_A$, respectively. In the next step, we measured the user activity obtained from L_A and R_A based on three features, namely coverage, duration, and enthusiasm, which are expressed as

$$\mathbb{C} = \sum_{i=1}^{n} \mathbb{C}(u_i), \mathbb{D} = \sum_{i=1}^{n} \mathbb{D}(u_i), \mathbb{E} = \sum_{i=1}^{n} \mathbb{E}(u_i), \qquad (7)$$

where n is the number of user in L_A or R_A.

The next step involves applying the proposed model to $Channel_A$. L_A and R_A obtain the value of each feature to observe the distribution of the activity in the two groups. Subsequently, we obtain the final score of the three features in the entire channel for classifying $Channel_A$ as a left-wing channel or right-wing channel. After normalizing the values of each feature in the range of [0,1], the formula for the activity score of L_A and R_A can be expressed as

$$\mathcal{A} = \mathbb{C}(u) + \mathbb{D}(u) + \mathbb{E}(u). \qquad (8)$$

The overall procedure for user activity modeling is presented in Algorithm 1.

Algorithm 1. User Activity Modeling.

Result: The political bias of the $Channel_A$
Input: $List(A)$, L_0, R_0
1 Reading the initial dataset (L_0, R_0)
2 # Separating the users in $List(A)$
 $List(L_A, R_A) = List(A) \cap List(L_0, R_0)$
3 # Getting score of users in list L_A
 $\mathbb{C}(L_A) = \mathbb{D}(L_A) = \mathbb{E}(L_A) = 0$
4 **for** u in $List(L_A)$ **do**
5 $\quad \mathbb{C}(L_A) \; += \; \mathbb{C}(u)$
6 $\quad \mathbb{D}(L_A) \; += \; \mathbb{D}(u)$
7 $\quad \mathbb{E}(L_A) \; += \; \mathbb{E}(u)$
8 **end**
9 # Getting score of users in list R_A
 $\mathbb{C}(R_A) = \mathbb{D}(R_A) = \mathbb{E}(R_A) = 0$
10 **for** u in $List(R_A)$ **do**
11 $\quad \mathbb{C}(R_A) \; += \; \mathbb{C}(u)$
12 $\quad \mathbb{D}(R_A) \; += \; \mathbb{D}(u)$
13 $\quad \mathbb{E}(R_A) \; += \; \mathbb{E}(u)$
14 **end**
15 Normalizing
 $\mathbb{C}(L_A), \mathbb{C}(R_A), \mathbb{D}(L_A), \mathbb{D}(R_A), \mathbb{E}(L_A), \mathbb{E}(R_A)$
16 # Getting user activity score for whole L_A and R_A
17 $\mathcal{A}(L_A) = \mathbb{C}(L_A) + \mathbb{D}(L_A) + \mathbb{E}(L_A))$;
18 $\mathcal{A}(R_A) = \mathbb{C}(R_A) + \mathbb{D}(R_A) + \mathbb{E}(R_A))$;
19 # Predict the political bias for $Channel_A$
20 **if** $\mathcal{A}(L_A) > \mathcal{A}(R_A)$ **then**
21 $\quad Channel_A$ in left-wing
22 **else**
23 $\quad$ **if** $\mathcal{A}(L_A) == \mathcal{A}(R_A)$ **then**
24 $\quad\quad Channel_A$ in middle
25 $\quad$ **else**
26 $\quad\quad Channel_A$ in right-wing;
27 $\quad$ **end**
28 **end**

Experiments

Data Collection

As one of the most popular media for OSNs, YouTube is used to share and disseminate information in various domains. YouTube allows users to perform numerous activities, such as uploading, viewing, commenting on, sharing, and rating videos. Although most content is created by individuals, media corporations or official organizations also share content in partnership with YouTube and develop a social impact. In this section, we describe the method used to collect and organize the data retrieved from YouTube. A crawler function was created to perform the task. We used the YouTube API to retrieve the required data, such as comments, user information, and video statistics identified by the requested parameters. However, the API has its limitations, and the number of videos and comments increases every hour owing to enthusiastic user activity. These restrictions lead to delays or failures with respect to the collection of datasets and affect the results.

To address the aforementioned issues, we built a system called TubePlunger to collect data (Tran et al., 2020). TubePlunger was developed using the model-view-controller (MVC) model to address security problems and handle multiple tasks involving large datasets using the Java programing language (Nguyen & Jung, 2020). In addition to overcoming these challenges, the system performs additional tasks related to our research purposes. Currently, the TubePlunger platform performs the following functions: (i) collecting and storing data from designated YouTube channels, (ii) presenting detailed information for each video (e.g., number of corresponding comments, publishing date, status of collection, and details of the comments), (iii) analyzing time-series statistics of the current data, and (iv) presenting the experimental results of our research. The details of the processing functions in the TubePlunger system are described as follows.

- First, the data crawler function collects information related to videos that were published a day ago from the designated channels and stores them in the database. The corresponding comments on these videos were retrieved and saved a week after their publication date. This function is repeated every day according to a schedule; hence, the size of our dataset continually increases. In addition, this method of data collection requires the user to have sufficient time to debate and express their opinions about the videos. Moreover, the number of required retrieval videos should not exceed the API limit.
- Second, the system updates the status of each video and each channel to present the details of the current dataset. The comments associated with the acquired videos are displayed on the dashboard of the system with detailed information (e.g., username, comment content, comment time, number of likes, and number of replies). The pending videos that are labeled as "collecting" or "waiting" temporarily cannot reveal any information, except for the title and the published day.
- Third, after updating the dataset using information related to the videos, comments, and users according to the channel, the system obtains time-series statistics for the changes in user activity.
- Finally, the experimental results of the system are stored and presented.

We aim to collect comments posted by Koreans on YouTube channels, wherein users expose and exchange political opinions, to better understand their activity. First, we searched YouTube channels to identify those that share topics related to Korean national politics. For searching political Youtube channels, we used the word "jeongchi" (in hangul) at the searching box as well as filtered the results for only "Channel" in "Type." Therefore, more than 50 channels

were found, and most of them discussed news related either to left-wing (liberal) or right-wing (conservative). In addition, to explore whether polarization exists in other types of channels, we used the "Explore" function in YouTube. When click on "Explore," Youtube will list trending videos based on IP location (South Korea in this case). We obtained more than 40 channels from the trending videos, which mostly talked about life news or entertainment. There are rules of filtering are defined when searching YouTube channels for the experiments. When observing the results after applying the method of searching, there are many channels with a small number of subscribers or that have a few videos. Therefore, these channels have less interaction compared to channels that receive more than 5,000 subscribers or more than 50 videos. A lower interaction means less data. Therefore, we decided to remove them, and we obtained a total of 71 channels in the testing dataset. Using this method, the system collected approximately 7.5 million rows of comments from these channels. Each row of a comment contains corresponding information such as the username, comment id, comment content, number of likes, number of replies, video ID, and channel ID. The total number of usernames after removing duplicates was 430,788.

Experimental Design

Before conducting the experiments, we had an interview with three experts from Department of Politics in Seoul National University. They have suggested the list of channels as the representations of left-wing and right-wing in an accurate way and we have chosen the most common channels in both sides. Subsequently, three official channels representing the left-wing and three official channels for the right-wing were identified. These channels have a large number of subscribers and frequently upload videos. Consequently, they are effective for political communication and have become one of the official channels for conveying political news. Information regarding these official channels is provided in Table 1. We used these channels as a sample dataset to evaluate our hypothesis. The 71 other channels were used to create the testing dataset to examine the accuracy of the hypothesis, as well as the reliability of the model proposed in this study. Table 2 presents information about the 71 channels used in the testing dataset for the experiments.

In this study, we assume that there is intense polarization in YouTube users between the left-wing and right-wing. Moreover, users tend to watch, share, and discuss their ideas and opinions only on the YouTube channels that they support. To prove this assertion, after grouping the six aforementioned YouTube channels into left-wing and right-wing, we conducted username extraction from their comments for each group. Each username in our dataset was considered a user. Subsequently, we defined the number of common users who left comments in both left-wing and right-wing channels, based on the intersection of these two groups.

Table 1. Detailed Data on the YouTube Channels in the Sample Dataset.

Channel	Channel title	Political bias	Joined from	#videos	#comments	#users
#1	Allilleo	Left-wing	Jan 14, 2011	538	435,763	104,452
#2	KimyongminTV	Left-wing	Sep 27, 2017	1,239	247,385	54,033
#3	TBS Siminuibangsong	Left-wing	Feb 23, 2012	2,733	727,358	116,190
				4,510	1,410,506	274,675
#4	Sinuihansu	Right-wing	Sep 14, 2012	2,161	1,102,826	122,180
#5	PenaendeumaikeuTV	Right-wing	Feb 14, 2012	2,104	666,387	97,953
#6	TV Hongkakolla	Right-wing	Nov 29, 2018	366	356,158	61,090
				4,631	2,125,371	281,223

Table 2. About 71 YouTube Channels in the Testing Dataset.

#	Channel title	#	Channel title	#	Channel title
1	Hwangjangsuui Nyuseubeuliping	25	Ariang Culture	49	Korea Top News
2	Jeongchichodan	26	Ariang Issue	50	Korean Entertainment News
3	Jeogpyecheongsan	27	ARIRANG K-POP	51	Korean News
4	Channel A News (Korea)	28	Arirang News	52	Maichard
5	TVCHOSUN News	29	BBC News Korea	53	NBS Hangugnongeobbangsong
6	Jeongchiilhag	30	VIDEOMUG	54	NK News
7	Minjuyeogsagilogso	31	Hot news korea	55	RNX TV Korea Ent
8	Chaeneol36.5	32	KBS 1Radio	56	SBS News
9	SBS Radio	33	KBS Drama	57	Top Korean News
10	Jeongchialjidomoshaneunsalam	34	KBS Entertain	58	TV Baijin
11	JKgongsig	35	KBS Entertain: kkalkkaltibi	59	VN22Q Korea News
12	Ulijeongchiilggi	36	KBS Kpop	60	VOA Hangugeo
13	Dagchyeolajeongchi	37	KBS News	61	YTN news
14	Sisa Mich Jeongchi	38	KBS Goldeunkeipab	62	Hankuk Sisanyuseu
15	JeongchitapaTV	39	KBS Gyoyang	63	KOREA Entertainment news
16	JeongchitaunTV	40	KBS Documentary	64	New and Life Info
17	Jeongchisepo	41	KBS Seupocheu	65	KOREA News
18	Jeongchi Dabang	42	Korea Breaking News	66	Daesgeul Ilgeojuneun Kijadeul
10	Kkulseunghun	43	Korea Daily News	67	Soggosaljima KBS
20	Choitub	44	Korean Daily News	68	YeonhabnyuseuTV
21	Eonieonseu Onience	45	KOREA Dispatch	69	Jeoneollijeum Tokeusyo J
22	Yeogsa	46	HwallyeogsoTV	70	HanguggyeongjeTV
23	Podcast Sinibsawon	47	SUBUSU NEWS	71	Hangugsaenghwalcheyugnyuseu
24	Sasageongeon	48	KOREA NOW		

The number of users in the left-wing and right-wing groups after removing the duplicates were 182,984 and 193,842, respectively, whereas there were only 35,221 common users. The results show that the number of common users was small in comparison with the number of users in each group. In particular, the number of common users accounted for approximately 8. The testing dataset contained 71 channels, both personal and organizational, that discussed South Korean politics. There are more than thousand channels in the political domain. However, these 71 channels met the restrictions in the filtering step. Table 3 provides information on the dataset. There were 77 channels in total, with 37 K videos, 600 K users, and more than 11M comments, which were obtained from the Korean political YouTube

Table 3. Statistics of the Dataset Provided for the Experiments.

Dataset	#Channels	#Videos	#Users	#Comments
Sample	6	9,141	422,047	3,563,903
Testing	71	28,716	430,788	7,527,623

channel case study. The main issues in this study were the classification of a channel as a left-wing or right-wing channel, visualization of the distribution in user activity to better understand their behavior, and measurement of the degree of political bias in each channel. After the list of users for each channel was obtained and the proposed model was applied, the experimental results are presented in the next section.
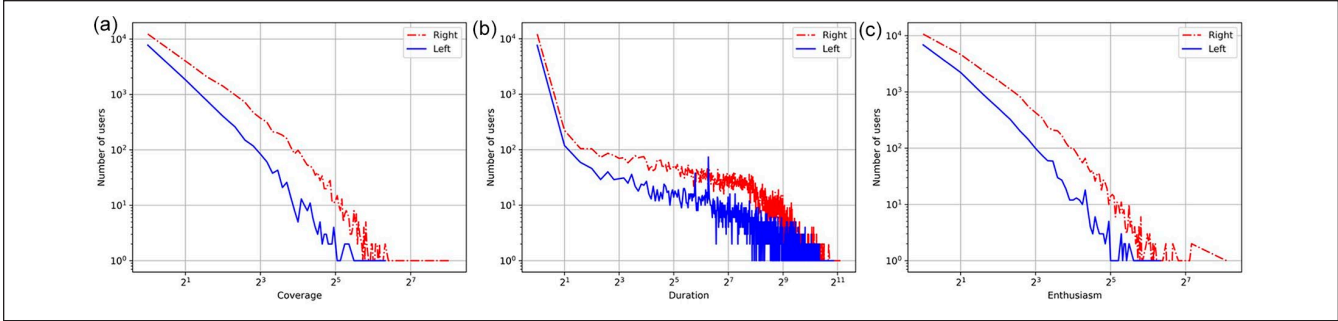
Figure 1. User activity distribution of channel *Channel A News (Korea)* in the log-scale chart based on three features: (a) coverage-based, (b) duration-based, and (c) enthusiasm-based.
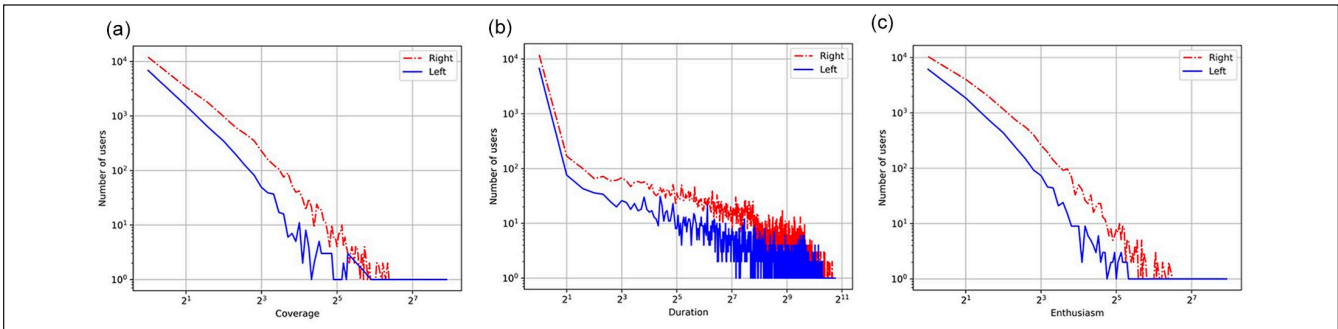
Figure 2. User activity distribution of channel *TVCHOSUN News* in the log-scale chart based on three features: (a) coverage-based, (b) duration-based, and (c) enthusiasm-based.

Experimental Results

Tables 7 to 9 present the scores calculated after normalizing for the three features of coverage, duration, and enthusiasm for the 71 channels in the testing dataset after applying the proposed model. The higher the score of a channel, the higher is the degree of user activity in that channel. After obtaining the scores of all the features, activity scores $\mathcal{A}(L_A)$ and $\mathcal{A}(R_A)$, which are used to determine the political bias of a channel, were calculated.

Each channel has activity scores corresponding to the left-wing and right-wing, based on the preceding formula. The last two columns present the bias value of the support associated with a channel for the left-wing and right-wing (rounding off to the second decimal place). The political bias values were obtained based on the final activity scores of the left-wing and right-wing in each channel. The gap between the left-wing and right-wing scores indicates the degree of political bias in the channel. Based on the table, more than 87% of the channels exhibited consistency in all three features. Therefore, if a channel is categorized as left-wing, its scores for all three features of left-wing are higher than those of right-wing. The greater the disparity between the scores, the stronger was the political bias of the channel. The nine channels that did not exhibit consistency in the scores of the three features

were mostly the channels in the middle. Consequently, users of left-wing channels exhibited better energy based on higher enthusiasm scores. Moreover, users of right-wing channels were more proactive in terms of coverage and duration.

Figures 1 and 2 reveal the distribution of user activity in the log-scale charts of channel *Channel A News (Korea)* and *TVCHOSUN News*, respectively. These two channels support the conservative. The user activity of the right-wing group, which is presented in red, is much stronger than that of the left-wing group for all coverage, duration, and enthusiasm. Figures 3 and 4 illustrate the user activity distribution in the log-scale charts of two channels in the middle, *SBS News* and *Korean News*, respectively. As these two channels do not have any polarization in terms of politics, the degree of activity between left-wing and right-wing users is the same. Therefore, which political bias has stronger activity in these channels cannot be determined. Figures 5 and 6 present the distribution of user activity in the log-scale charts of channel *Jeongchichodan* and *KBS 1Radio*, respectively. It is evident that user activity corresponding to left-wing is stronger than that of right-wing activity for all three features as *Jeongchichodan* and *KBS 1Radio* support the liberal. According to the results, 35 channels support the left-wing, 33 channels promote the right-wing, and 3 channels are in the middle.
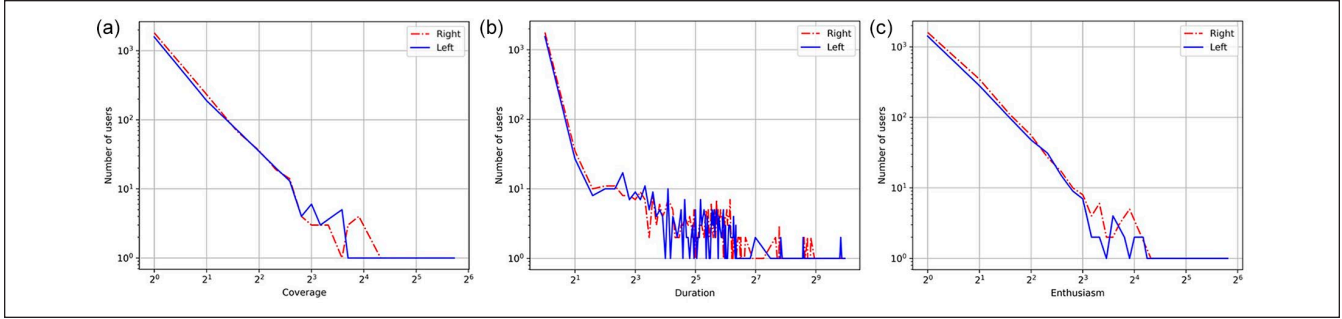
Figure 3. User activity distribution of channel *Korean News* in the log-scale chart based on three features: (a) coverage-based, (b) duration-based, and (c) enthusiasm-based.
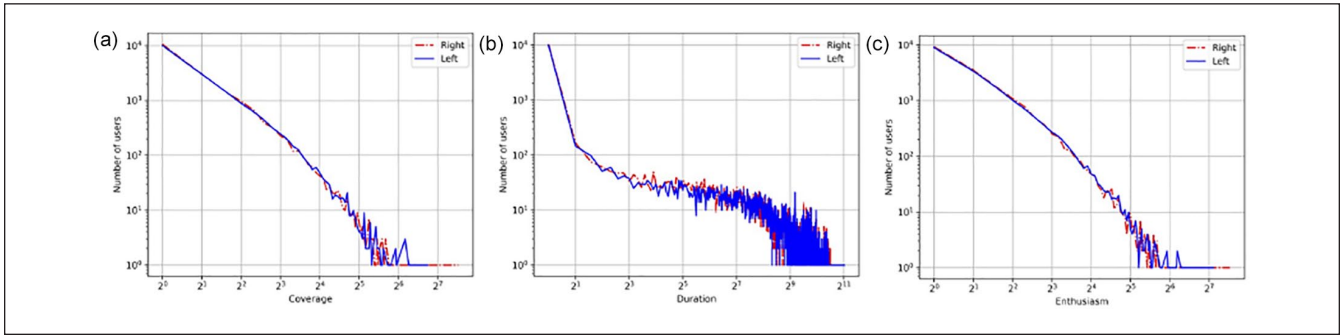
Figure 4. User activity distribution of channel *SBS News* in the log-scale chart based on three features: (a) coverage-based, (b) duration-based, and (c) enthusiasm-based.
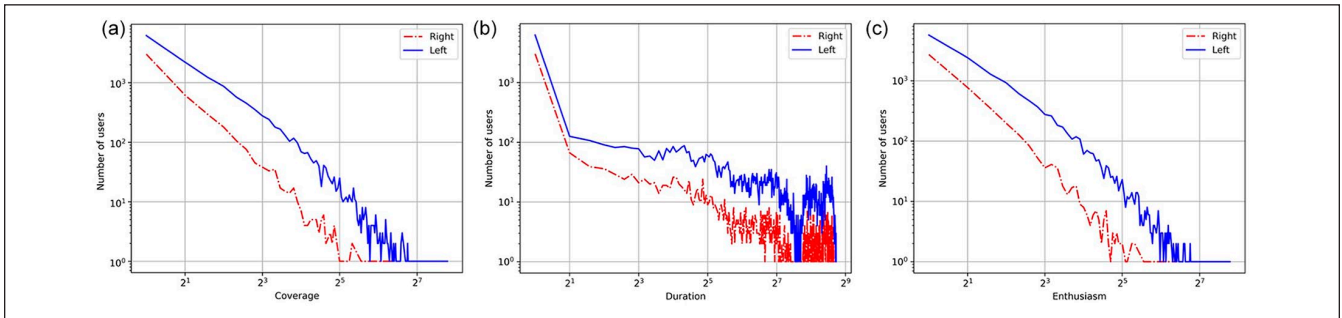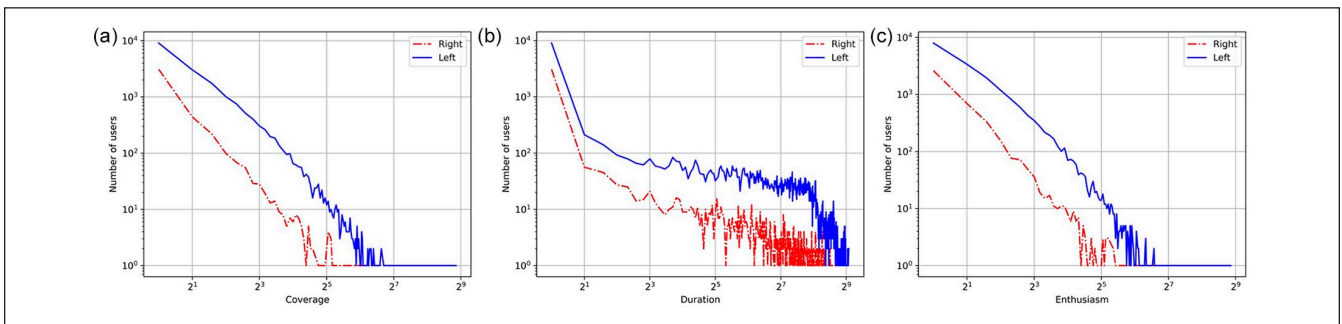

Figure 5. User activity distribution of channel *Jeongchichodan* in the log-scale chart based on three features: (a) coverage-based, (b) duration-based, and (c) enthusiasm-based.

Figure 6. User activity distribution of channel *KBS 1 Radio* in the log-scale chart based on three features: (a) from channel 1 to 24, (b) from channel 25 to 48, and (c) from channel 49 to 71.

Table 4. Bias Percentage to Liberal of Left-Wing Channels.

Channel	Title	Bias (%)
#42	Korea Breaking News	99
#69	Jeoneollijeum Tokeusyo J	94
#66	Daesgeul Ilgeojuneun Kijadeul	92
#11	JKgongsig	90
#32	KBS 1Radio	90
#21	Eonieonseu Onience	89
#14	Sisa Mich Jeongchi	88
#23	Podcast Sinibsawon	86
#07	Minjuyeogsagilogso	85
#03	Jeogpyecheongsan	83
#15	JeongchitapaTV	83
#65	KOREA News	83
#13	Dagchyeolajeongchi	82
#24	Sasageongeon	79
#06	Jeongchiilhag	77
#08	_Chaeneol36.5	76
#29	BBC News Korea	65
#02	Jeongchichodan	63
#48	KOREA NOW	61
#25	Arirang Culture	59
#49	Korea Top News	57
#18	Jeongchi Dabang	55
#27	ARIRANG K-POP	55
#28	Arirang News	55
#30	VIDEOMUG	55
#41	KBS Seupocheu	55
#37	KBS News	54
#45	KOREA Dispatch	54
#46	HwallyeogsoTV	54
#47	SUBUSU NEWS	54
#61	YTN news	53
#63	KOREA Entertainment news	53
#26	Arirang Issue	51
#40	KBS Documentary	51
#44	Korean Daily News	51
#56	SBS News	51

Table 5. Bias Percentage to Conservative of Right-Wing Channels.

Channel	Title	Bias (%)
#12	Ulijeongchiilggi	95
#01	Hwangjangsuui Nyuseubeuliping	94
#58	TV Baijin	94
#17	Jeongchisepo	92
#20	Choitub	92
#62	Hankuk Sisanyuseu	91
#16	JeongchitaunTV	90
#19	Kkulseunghun	88
#71	Hangugsaenghwalcheyugnyuseu	88
#60	VOA Hangugeo	87
#22	Yeogsa	79
#04	Channel A News (Korea)	76
#05	TVCHOSUN News	73
#10	Jeongchialjidomoshaneunsalam	69
#31	Hot news korea	68
#43	Korea Daily News	63
#52	Maichard	63
#55	RNX TV Korea Ent	61
#50	Korean Entertainment News	57
#53	NBS Hangugnongeobbangsong	57
#64	New and Life Info	56
#38	KBS Goldeunkeipab	55
#39	KBS Gyoyang	55
#59	VN22Q Korea News	55
#70	HanguggyeongjeTV	54
#51	Korean News	53
#68	YeonhabnyuseuTV	53
#35	KBS Entertain: kkalkkaltibi	52
#09	SBS Radio	51
#33	KBS Drama	51
#36	KBS Kpop	51
#57	Top Korean News	51
#67	Soggosaljima_KBS	51

Table 6. Middle Channels.

Channel	Title	Bias (%)
#34	KBS Entertain	50
#54	NK News	50

Table 4 illustrates the descending polarization percentages of the channels belonging to the left-wing. Approximately 45% of the channels exhibited strong polarization in the left-wing group when the percentage of bias was higher than 70%. *Korea Breaking News, Jeoneollijeum Tokeusyo J, Daesgeul Ilgeojuneun Kijadeul, JKgongsig*, and *KBS 1Radio* are the channels that exhibit the highest bias in politics of more than 90%. Table 5 illustrates the descending polarization percentage of the channels that belong to the right-wing. Approximately 40% of the channels exhibited extreme polarization in the right-wing group, because the percentage of bias was higher than 70%. *Ulijeongchiilggi* is the most polarized channel with 95% bias for conservatives. *Hwangjangsuui Nyuseubeuliping, TV Baijin, Jeongchisepo, Choitub, Hankuk Sisanyuseu*, and *JeongchitaunTV* are extreme right-wing channels given that

the percentage of bias is higher than 90%. Table 6 presents the polarization percentages of the three channels in the middle with no polarization. However, channels with a gap of less than 10% between the left-wing and right-wing can also be considered as middle channels. Therefore, approximately 21 channels can be considered middle channels because they do not exhibit much bias, based on the results of our model. Most of these channels discuss general daily news, documentaries, or entertainment. As they are mostly created by national broadcasting systems, they should maintain a balance in the
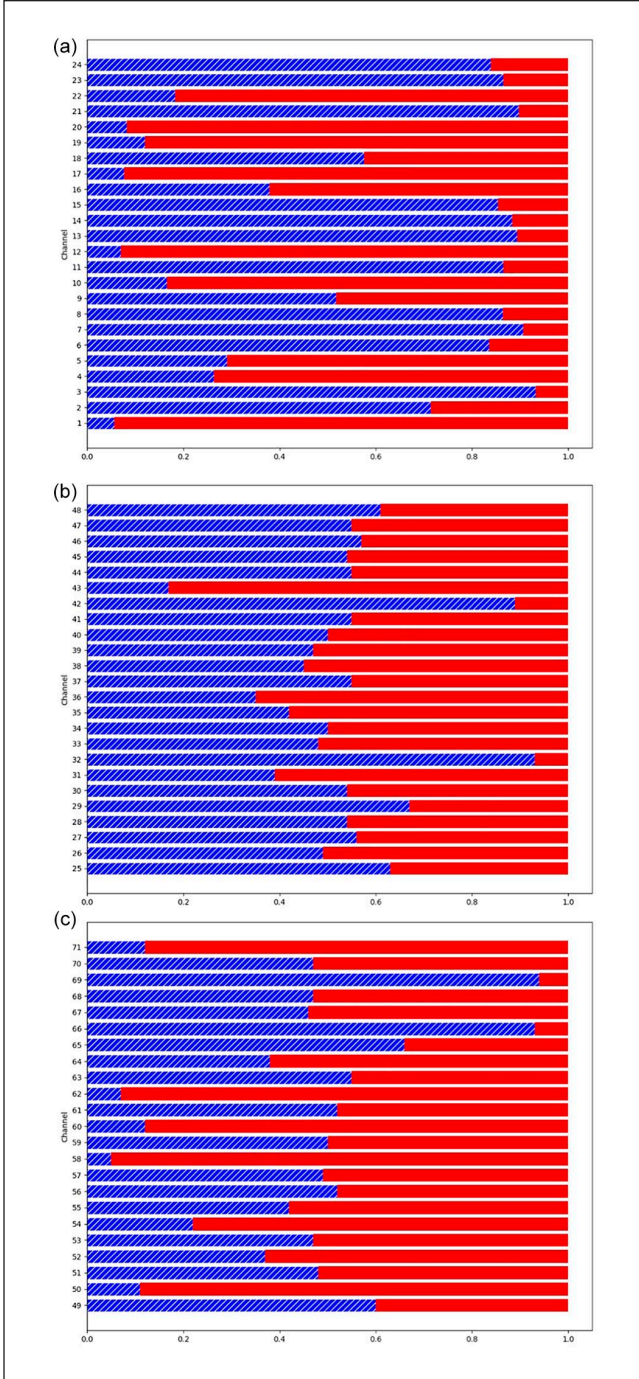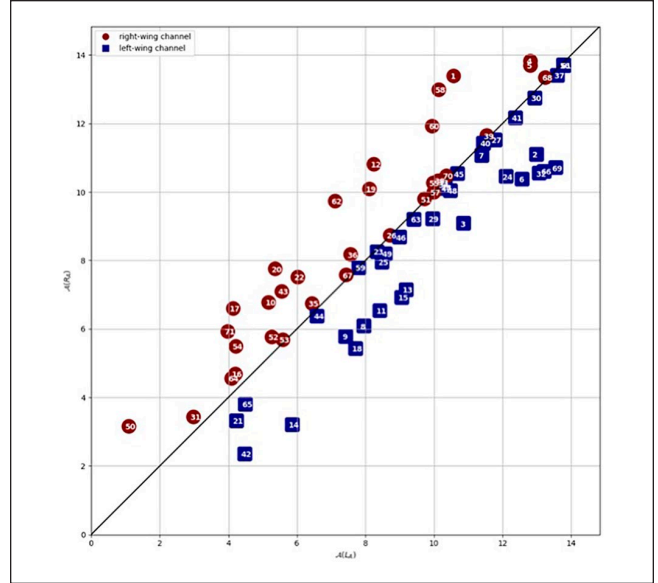
Figure 7. Percentage of the political bias labels predicted for the 71 channels in the testing dataset. The blue and red colors present the left-wing and the right-wing, respectively.

Figure 8. Classification results for the testing dataset (71 YouTube channels) in the log-scale chart.

sharing of information instead of expressing strong political bias like individual channels.

Figure 7 presents the experimental results for the measurement of political bias. The blue represents the left-wing, and the red is indicative of the right-wing. To respond to skewness toward large values, we decided to use the log-scale chart to

present the classification results for the 71 YouTube channels in the testing dataset, which are shown in Figure 8. The blue nodes (squares) denote left-wing channels, whereas the red nodes (circles) denote right-wing channels. Tables 7 to 9 present the detailed experimental results for the 71 YouTube channels of the testing dataset. Columns 8 and 9 present the average scores for users $\mathcal{A}(L_A)$ and $\mathcal{A}(R_A)$, which were measured using the coverage, duration, and enthusiasm features. $\mathcal{A}(L_A)$ and $\mathcal{A}(R_A)$ are the decisive elements for classifying the channels in the testing dataset. The results reveal that 36 channels support the left-wing, 33 channels support the right-wing, and 2 are in the middle. In Figure 8, the distance from the nodes to the boundary (the red line) denotes the reliability (strong polarization) of the left-wing or right-wing of each channel. The channels in the middle, *KBS Entertain* and *NK News*, which are created by organizations, express balanced opinions.

Conclusion and Future Work

We proposed a user-activity-based model using data from YouTube, one of the most popular video-sharing platforms, to explore political polarization. By applying the proposed model, we exploited the implicit meaning of user activity to better understand bias in online users. Using coverage, duration, and enthusiasm as features, the model yielded outputs to examine the research questions. We conducted experiments based on a case study of Korean political YouTube channels. To collect daily data, the TubePlunger system was built using the MVC model to address issues related to the API's key limitations and security problems. The data were split into a sample dataset with six official channels, which

Table 7. Experimental Results for the Political Polarization of YouTube Channels (From #1 to #24).

	Coverage		Duration		Enthusiasm		Activity score		Political bias	
#	L	R	L	R	L	R	$\mathcal{A}(L_A)$	$\mathcal{A}(R_A)$	L	R
1	0.0602	1.0000	0.0605	0.9645	0.0372	0.6216	0.1579	2.5861	0.0576	0.9424
2	0.0005	0.0003	0.0005	0.0003	0.0000	0.0000	0.0010	0.0006	0.6303	0.3697
3	0.0011	0.0003	0.0012	0.0003	0.0003	0.0000	0.0026	0.0005	0.8288	0.1712
4	0.2767	0.9864	0.2963	1.0000	0.3623	1.0000	0.9352	2.9864	0.2385	0.7615
5	0.2228	0.6598	0.2356	0.6807	0.3675	0.8853	0.8258	2.2258	0.2706	0.7294
6	0.0057	0.0017	0.0058	0.0019	0.0014	0.0002	0.0129	0.0038	0.7700	0.2300
7	0.0080	0.0013	0.0083	0.0017	0.0018	0.0001	0.0181	0.0031	0.8525	0.1475
8	0.0040	0.0014	0.0039	0.0015	0.0027	0.0004	0.0106	0.0033	0.7637	0.2363
9	0.0099	0.0106	0.0133	0.0145	0.0036	0.0032	0.0269	0.0283	0.4866	0.5134
10	0.0005	0.0008	0.0005	0.0008	0.0002	0.0009	0.0011	0.0025	0.3129	0.6871
11	0.0175	0.0019	0.0176	0.0019	0.0036	0.0006	0.0388	0.0044	0.8972	0.1028
12	0.0095	0.1845	0.0094	0.1789	0.0034	0.0408	0.0223	0.4042	0.0522	0.9478
13	0.0192	0.0044	0.0192	0.0051	0.0078	0.0008	0.0463	0.0103	0.8179	0.1821
14	0.0660	0.0093	0.0651	0.0094	0.0061	0.0008	0.1373	0.0194	0.8760	0.1240
15	0.1181	0.0241	0.1133	0.0236	0.0473	0.0078	0.2788	0.0555	0.8340	0.1660
16	0.0001	0.0008	0.0001	0.0012	0.0001	0.0001	0.0002	0.0020	0.1040	0.8960
17	0.0004	0.0044	0.0004	0.0048	0.0000	0.0005	0.0009	0.0097	0.0832	0.9168
18	0.0775	0.0647	0.0801	0.0697	0.0876	0.0639	0.2452	0.1983	0.5529	0.4471
19	0.0160	0.1199	0.0166	0.1195	0.0025	0.0178	0.0351	0.2573	0.1200	0.8800
20	0.0008	0.0092	0.0008	0.0095	0.0002	0.0019	0.0017	0.0206	0.0781	0.9219
21	0.5687	0.0707	0.5491	0.0715	0.2634	0.0296	1.3812	0.1718	0.8894	0.1106
22	0.0020	0.0070	0.0020	0.0070	0.0003	0.0015	0.0042	0.0155	0.2143	0.7857
23	0.8138	0.1259	0.7825	0.1273	0.4008	0.0618	1.9972	0.3151	0.8637	0.1363
24	0.2587	0.0733	0.2719	0.0850	0.1713	0.0315	0.7019	0.1897	0.7872	0.2128

Table 8. Experimental Results for the Political Polarization of YouTube Channels (From #25 to #48).

	Coverage		Duration		Enthusiasm		Activity score		Political bias	
#	L	R	L	R	L	R	$\mathcal{A}(L_A)$	$\mathcal{A}(R_A)$	L	R
25	0.0056	0.0042	0.0059	0.0044	0.0047	0.0027	0.0162	0.0113	0.5891	0.4109
26	0.0029	0.0026	0.0037	0.0031	0.0062	0.0064	0.0128	0.0121	0.5145	0.4855
27	0.0466	0.0383	0.0479	0.0407	0.1331	0.1041	0.2276	0.1831	0.5542	0.4458
28	0.0263	0.0212	0.0299	0.0252	0.0300	0.0252	0.0862	0.0716	0.5462	0.4538
29	0.0431	0.0234	0.0561	0.0314	0.0188	0.0090	0.1180	0.0638	0.6489	0.3511
30	0.2366	0.1956	0.2562	0.2104	0.4147	0.3500	0.9075	0.7560	0.5455	0.4545
31	0.0001	0.0002	0.0001	0.0002	0.0000	0.0000	0.0002	0.0005	0.3215	0.6785
32	0.8362	0.1019	0.8455	0.1131	0.4429	0.0335	2.1246	0.2485	0.8953	0.1047
33	0.0264	0.0275	0.0286	0.0294	0.0272	0.0293	0.0822	0.0862	0.4881	0.5119
34	0.0299	0.0299	0.0313	0.0307	0.0266	0.0272	0.0879	0.0878	0.5003	0.4997
35	0.0028	0.0029	0.0030	0.0032	0.0005	0.0007	0.0063	0.0068	0.4806	0.5194
36	0.0055	0.0051	0.0062	0.0056	0.0016	0.0034	0.0134	0.0141	0.4872	0.5128
37	0.4596	0.3898	0.4835	0.4219	0.8136	0.6733	1.7566	1.4850	0.5419	0.4581
38	0.0580	0.0701	0.0624	0.0756	0.0221	0.0276	0.1425	0.1732	0.4512	0.5488
39	0.1078	0.1326	0.1168	0.1457	0.1005	0.1114	0.3251	0.3897	0.4548	0.5452
40	0.1676	0.1621	0.1808	0.1736	0.0863	0.0864	0.4346	0.4221	0.5073	0.4927
41	0.1709	0.1373	0.1766	0.1408	0.2354	0.1935	0.5829	0.4716	0.5528	0.4472
42	0.0004	0.0000	0.0005	0.0000	0.0001	0.0000	0.0009	0.0000	0.9903	0.0097
43	0.0021	0.0031	0.0022	0.0035	0.0001	0.0011	0.0045	0.0077	0.3683	0.6317
44	0.0010	0.0012	0.0016	0.0014	0.0007	0.0005	0.0032	0.0031	0.5061	0.4939
45	0.0359	0.0298	0.0385	0.0312	0.0432	0.0376	0.1176	0.0985	0.5440	0.4560
46	0.0235	0.0206	0.0262	0.0222	0.0068	0.0049	0.0565	0.0477	0.5423	0.4577
47	0.0314	0.0264	0.0353	0.0296	0.0292	0.0242	0.0959	0.0803	0.5443	0.4557
48	0.0499	0.0319	0.0527	0.0339	0.0339	0.0220	0.1365	0.0879	0.6084	0.3916

Table 9. Experimental Results for the Political Polarization of YouTube Channels (From #49 to #71).

	Coverage		Duration		Enthusiasm		Activity score		Political bias	
#	L	R	L	R	L	R	$\mathcal{A}(L_A)$	$\mathcal{A}(R_A)$	L	R
49	0.0037	0.0031	0.0044	0.0034	0.0053	0.0036	0.0135	0.0100	0.5734	0.4266
50	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.4342	0.5658
51	0.0320	0.0356	0.0345	0.0384	0.0152	0.0166	0.0817	0.0905	0.4744	0.5256
52	0.0005	0.0008	0.0005	0.0008	0.0002	0.0003	0.0011	0.0018	0.3702	0.6298
53	0.0008	0.0011	0.0008	0.0011	0.0002	0.0002	0.0019	0.0025	0.4315	0.5685
54	0.0005	0.0005	0.0008	0.0006	0.0000	0.0002	0.0014	0.0013	0.5033	0.4967
55	0.0436	0.0719	0.0439	0.0714	0.0195	0.0258	0.1070	0.1692	0.3874	0.6126
56	0.6023	0.5980	0.6141	0.6122	0.9635	0.9034	2.1799	2.1136	0.5077	0.4923
57	0.0098	0.0105	0.0110	0.0109	0.0220	0.0225	0.0427	0.0438	0.4940	0.5060
58	0.0353	0.4948	0.0366	0.4920	0.0241	0.4279	0.0960	1.4147	0.0635	0.9365
59	0.0045	0.0058	0.0049	0.0064	0.0023	0.0022	0.0117	0.0144	0.4483	0.5517
60	0.0441	0.2966	0.0491	0.2943	0.0188	0.1410	0.1121	0.7319	0.1328	0.8672
61	0.5828	0.5005	0.6006	0.5219	0.9837	0.9059	2.1672	1.9283	0.5292	0.4708
62	0.0032	0.0327	0.0036	0.0345	0.0011	0.0158	0.0079	0.0831	0.0869	0.9131
63	0.0176	0.0159	0.0188	0.0168	0.0116	0.0094	0.0480	0.0421	0.5328	0.4672
64	0.0006	0.0008	0.0007	0.0009	0.0000	0.0000	0.0014	0.0018	0.4354	0.5646
65	0.0005	0.0001	0.0005	0.0001	0.0001	0.0000	0.0010	0.0002	0.8270	0.1730
66	0.7769	0.0636	0.7585	0.0677	0.5212	0.0382	2.0566	0.1695	0.9238	0.0762
67	0.0077	0.0084	0.0098	0.0096	0.0012	0.0015	0.0187	0.0194	0.4904	0.5096
68	0.2623	0.2923	0.2800	0.3152	0.5693	0.6370	1.1115	1.2445	0.4718	0.5282
69	0.9170	0.0625	0.8974	0.0660	0.7345	0.0432	2.5489	0.1717	0.9369	0.0631
70	0.0417	0.0483	0.0455	0.0529	0.0301	0.0343	0.1173	0.1355	0.4639	0.5361
71	0.0003	0.0025	0.0004	0.0028	0.0000	0.0002	0.0007	0.0055	0.1192	0.8808

were the channels attributed to known political bias, and an experimental dataset with 71 channels of various types. First, we extracted the usernames from the sample dataset, which are considered political supporters, corresponding to the left-winger and right-winger lists. These two lists were then considered representative lists of left-wingers and right-wingers. Second, the lists of usernames obtained from the 71 channels in the testing dataset were matched to the representative lists to filter the usernames needed for the model input. Finally, we applied the model to measure political polarization and user activity distribution in each channel. The experiments were processed using a dataset that included 77 channels, 37 K videos, 600 K users, and more than 11 M comments, obtained from the study of the Korean political YouTube channels.

In this study, we attempted to answer two research questions. We extracted usernames from six official channels in response to RQ (1). After dividing them into left-wing and right-wing groups, there was a small portion of common users who left comments in both left-wing and right-wing channels, based on the intersection of these two groups. Notably, the number of common users only accounted for 18% of the number of left-wingers, 19% of the number of right-wingers, and only 8% of the number of users in the entire sample dataset. Therefore, based on this case study on Korean YouTube channels, it was determined that there is

intense polarization in the political communities on YouTube. In response to RQ (2), data from 71 Korean YouTube channels were used for the proposed model in the experiments. As a result, in the other categories of online video channels, there is still a political bias, depending on the channel. The channels that usually posted videos on politics or topics related to politics had a stronger degree of political bias than entertainment channels. Some channels also hosted videos on news about politics, but they exhibited less polarization because they were national channels. In addition, the experimental results do not classify the channels according to their beliefs; however, they reveal the devotion of users that follow a channel, corresponding to the left-wing and right-wing groups. The final results illustrate the consistency in user activity distribution in all three proposed features of most channels without an understanding of the content of the comments. Based on the results of this study, the proposed model can be used in other online sharing platforms that have many political communities such as Facebook or Twitter with the same original hypothesis and customized features depending on its characteristics.

This study had several limitations. First, there are some cases of different users having the same username. Our current experiment only considered usernames without identifying the exact users. This flaw leads to inaccurate measurement of user activity. Second, not all users who commented on

videos in a particular channel were talking about political issues. The proposed model does not consider the content of comments or the content of the videos. Hence, in the next step, we need to remove videos that are not related to political topics. Third, the current version of the proposed model is not fully able to identify accurate supporters because we do not consider the sentiment of comments.

Future work should consider determining the weight of each feature to improve the performance of the model. Furthermore, the model should be upgraded by adding new meaningful features to guarantee greater output. Analysis of the sentiment and content of videos, as well as comments, is an accurate way to obtain a comprehensive perspective on user behavior in OSNs.

ORCID iD

Jason J. Jung 🆔 https://orcid.org/0000-0003-0050-7445

References

Amedie, J. (2015). The impact of social media on society. *Advanced Writing: Pop Culture Intersections September*, *1*(1), 1–20.

Balakrishnan, J., & Griffiths, M. D. (2017). Social media addiction: What is the role of content in YouTube? *Journal of Behavioral Addictions*, *6*(3), 364–377.

Bärtl, M. (2018). Youtube channels, uploads and views: A statistical analysis of the past 10 years. *Convergence*, *24*(1), 16–32.

Bello-Orgaz, G., Jung, J. J., & Camacho, D. (2016). Social big data: Recent achievements and new challenges. *Information Fusion*, *28*, 45–59.

Benevenuto, F., Rodrigues, T., Cha, M., & Almeida, V. (2009). *Characterizing user behavior in online social networks* [Conference session]. Proceedings of the 9th ACM SIGCOMM Conference on Internet Measurement, Chicago, USA (pp. 49–62).

Bradshaw, A. C. (2001). Internet users worldwide. *Educational Technology Research and Development*, *49*(4), 112–117.

Brady, H. E. (2019). The challenge of big data and data science. *Annual Review of Political Science*, *22*, 297–323.

Bright, J. (2018). Explaining the emergence of political fragmentation on social media: The role of ideology and extremism. *Journal of Computer-Mediated Communication*, *23*(1), 17–33.

Cha, M., Haddadi, H., Benevenuto, F., & Gummadi, K. (2010). Measuring user influence in twitter: The million follower fallacy. *Proceedings of the international AAAI Conference on Web and Social Media*, *4*(1), 10–17.

Curini, L., & Hino, A. (2012). Missing links in party-system polarization: How institutions and voters matter. *The Journal of Politics*, *74*(2), 460–473.

Dalton, R. J. (2008). The quantity and the quality of party systems: Party system polarization, its measurement, and its consequences. *Comparative Political Studies*, *41*(7), 899–920.

Effing, R., Van Hillegersberg, J., & Huibers, T. (2011). Social media and political participation: Are facebook, twitter and youtube democratizing our political systems? In E. Tambouris, A. Macintosh, & H. de Bruijn (Eds.), *International conference on electronic participation* (pp. 25–35). Springer,.

Garibay, I., Mantzaris, A. V., Rajabi, A., & Taylor, C. E. (2019). Polarization in social media assists influencers to become more influential: Analysis and two inoculation strategies. *Scientific Reports*, *9*(1), 1–9.

Gill, P., Arlitt, M., Li, Z., & Mahanti, A. (2008). Characterizing user sessions on youtube. In: *Multimedia computing and networking 2008* (Vol. 6818, p. 681806). International Society for Optics and Photonics.

Groeling, T. (2008). Who's the fairest of them all? An empirical test for partisan bias on abc, cbs, nbc, and fox news. *Presidential Studies Quarterly*, *38*(4), 631–657.

Groseclose, T., & Milyo, J. (2005). A measure of media bias. *The Quarterly Journal of Economics*, *120*(4), 1191–1237.

Guan, T., Liu, T., & Yang, Y. (2021). Moving away from partisanship: Diversifying the motives behind political polarization. *Political Studies Review*, *19*(4), 656–667.

Haselmayer, M., Wagner, M., & Meyer, T. M. (2017). Partisan bias in message selection: Media gatekeeping of party press releases. *Political Communication*, *34*(3), 367–384.

Joo, W. T., Jeong, Y. S., & Oh, K. (2016). *Political orientation detection on korean newspapers via sentence embedding and deep learning* [Conference session]. 2016 International Conference on Big Data and Smart Computing (BigComp) (pp. 502–504), Hong Kong, China, IEEE.

Jung, J. J. (2017). Computational collective intelligence with big data: Challenges and opportunities. *Future Generation Computer Systems*, *66*, 87–88.

Kam, C., Indridason, I., & Bianco, W. (2017). Polarization in multiparty systems.

Kang, H., & Yang, J. (2020). Quantifying perceived political bias of newspapers through a document classification technique. *Journal of Quantitative Linguistics*, *29*(2), 127–150.

Kang, W. T. (2008). How ideology divides generations: The 2002 and 2004 South Korean elections. *Canadian Journal of Political Science 41*(2), 461–480.

Kim, K. R., Kang, J. S., & Yun, S. (2012). Moral intuitions and political orientation: Similarities and differences between South Korea and the United States. *Psychological Reports*, *111*(1), 173–185.

Kwak, K. S. (2012). *Media and democratic transition in South Korea*. Routledge.

Layman, G. C., Carsey, T. M., & Horowitz, J. M. (2006). Party polarization in american politics: Characteristics, causes, and consequences. *Annual Review of Political Science*, *9*, 83–110.

Lee, J. (2016). *A study on the party system in South Korea after democratization* [Doctoral Dissertation], The University of Michigan.

Levendusky, M. S. (2013). Why do partisan media polarize viewers? *American Journal of Political Science*, *57*(3), 611–623.

Malouf, R., & Mullen, T. (2007) *Graph-based user classification for informal online political discourse* [Conference session]. Proceedings of the 1st Workshop on Information Credibility on the Web, Miyazaki, Japan.

Morales, A. J., Borondo, J., Losada, J. C., & Benito, R. M. (2015). Measuring political polarization: Twitter shows the two sides of venezuela. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, *25*(3), 033114.

Nguyen, L. V., Hong, M. S., Jung, J. J., & Sohn, B. S. (2020). Cognitive similarity-based collaborative filtering recommendation system. *Applied Sciences*, *10*(12), 4183.

Nguyen, L. V., & Jung, J. J. (2020). Crowdsourcing platform for collecting cognitive feedbacks from users: A case study on movie recommender system. In H. Pham (Ed.), *Reliability and statistical computing* (pp. 139–150). Springer.

Nguyen, L. V., Nguyen, T. H., & Jung, J. J. (2020). *Content-based collaborative filtering using word embedding: A case study on movie recommendation* [Conference session]. Proceedings of the International Conference on Research in Adaptive and Convergent Systems (pp. 96–100), Gwangju, Seoul, South Korea.

Papagelis, M., Murdock, V., & van Zwol, R. (2011). Individual behavior and social influence in online social systems [Conference session]. Proceedings of the 22nd ACM conference on Hypertext and hypermedia(pp. 241–250), Eindhoven, The Netherlands, June 6–9, 2011.

Park, S., Ko, M., Kim, J., Liu, Y., & Song, J. (2011). *The politics of comments: Predicting political orientation of news stories with commenters' sentiment patterns* [Conference session]. Proceedings of the ACM 2011 conference on computer supported cooperative work (pp. 113–122), Hangzhou, China, March 19–23, 2011.

Piccialli, F., & Jung, J. E. (2017). Understanding customer experience diffusion on social networking services by big data analytics. *Mobile Networks and Applications*, *22*(4), 605–612.

Preoţiuc-Pietro, D., Liu, Y., Hopkins, D., & Ungar, L. (2017). Beyond binary labels: Political ideology prediction of twitter users [Conference session]. Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers., pp. 729–740), Vancouver, Canada.

Prior, M. (2013). Media and political polarization. *Annual Review of Political Science*, *16*, 101–127.

Saulwick, A., & Trentelman, K. (2014). Towards a formal semantics of social influence. *Knowledge-Based Systems*, *71*, 52–60.

Settle, J. E. (2018). *Frenemies: How social media polarizes America*. Cambridge University Press.

Steinberg, D. I., & Shin, M. (2006). Tensions in south korean political parties in transition: From entourage to ideology? *Asian Survey*, *46*(4), 517–537.

Stoyanovich, J., Amer-Yahia, S., Marlow, C., & Yu, C. (2008). Leveraging tagging to model user interests in del. icio. Us. In AAAI Spring Symposium: Social Information Processing [Symposium] (pp. 104–109), Stanford, CA, USA.

Strömberg, D. (2004). Mass media competition, political competition, and public policy. *The Review of Economic Studies*, *71*(1), 265–284.

Suh, H. K. (2007). The historical origins of the national founding constitution of the Republic of Korea [Conference session]. KSAA Conference (p. 218).

Sunstein, C. R. (1999). *The law of group polarization* (John M. Olin Law & Economics Working Paper, 91). University of Chicago Law School.

Tran, G. T., Nguyen, L. V., Jung, J. J., & Han, J. (2020). *Modeling user loyalty for Korean political youtube channels* [Conference session]. Proceedings of the International Conference on Research in Adaptive and Convergent Systems (pp. 101–106), Gwangju, Seoul, South Korea, October 13–16, 2020.

Van Aelst, P., Strömbäck, J., Aalberg, T., Esser, F., De Vreese, C., Matthes, J., Hopmann, D., Salgado, S., Hubé, N., Stepińska, A., Papathanassopoulos, S., Berganza, R., Legnante, G., Reinemann, C., Sheafer, T., & Stanyer, A. (2017). Political communication in a high-choice media environment: A challenge for democracy? *Annals of the International Communication Association*, *41*(1), 3–27.

Yang, J., Rojas, H., Wojcieszak, M., Aalberg, T., Coen, S., Curran, J., Hayashi, K., Iyengar, S., Jones, P. K., Mazzoleni, G., Papathanassopoulos, S., Rhee, J. W., Rowe, D., Soroka, S., & Tiffen, R. (2016). Why are "others" so polarized? Perceived political polarization and media use in 10 countries. *Journal of Computer-Mediated Communication*, *21*(5), 349–367.