Changing News Media Landscape in South Korea

Hongjun Lim*   Choongho Chung*   Jihee Kim**
Juho Kim†   Sue Moon†   Meeyoung Cha*

*Graduate School of Culture Technology, KAIST, South Korea
**College of Business, KAIST, South Korea
†School of Computing, KAIST, South Korea
{limhj159, thegenuine, jiheekim, juhokim, sbmoon, meeyoungcha}@kaist.ac.kr

ABSTRACT

A recent disclosure of a presidential scandal and the following development of subsequent events have been a major news topic in South Korea. We conducted a data-driven study to examine public reactions to the scandal and their effect on the Korean media landscape. Our analysis is based on 59,224 news articles published by five popular newsrooms that received a total of 47,906,770 Likes on Facebook. The data reveal notable changes in media ranks throughout the scandal, where a relatively young TV news network outgrew its audience over other half-a-century old established newspapers in less than a month. Topical similarity of news headlines also remained high over an extended period of time, despite the varying political stance of each newsroom. The topical similarity further shows a gradual divergence, indicating re-positioning of media stance takes place over time. Implications of these findings and suggestions for future directions are discussed based on data analysis of this extraordinary event.

1. INTRODUCTION

With growing digital news consumers and a plethora of platforms through which news stories disseminate, the global news media landscape has been changing fast. According to studies, digital and social media platforms like Yahoo News, Facebook, and YouTube have democratized media and, as a result, digital-only platforms like Buzzfeed now have an equal chance to spread stories as any established platforms [9]. Changes in the news media landscape are sometimes driven by innovations in newsrooms themselves, for instance, the uptake of the Huffington Post via a blog-based strategy and the New York Times' shifts to visual-centric and socially share-worthy stories.

Another source of changes in media landscape is large societal events. These events incur significant changes in patterns of news cycles, in the way the public with prior knowledge of the events reacts to the follow-up news [6], and in dynamics between media outlets. Elections and political scandals affect the 24-hour news cycle and increase audience interactions. Moreover, there had been cases where a news outlet rose to a position of power by reporting news scoops for a major event and providing insights in follow-up reports, ending up with a wide public trust and favor. The role of the Washington Post on Watergate (1972), and of Al Jazeera on Arab Spring (2010) provided opportunities for the two media to be recognized by the public and gain authority, either domestically or internationally [7, 10].

This paper reports one such transformation point in the Korean media landscape that started over a political scandal in 2016 and is still on-going as of the writing of this paper in February 2017. The scandal, involving allegations for the extortion, abuse of power, and bribery of the Korean president, has led to the largest ever protest in the country on December 3rd, with the estimated number of participants over 2.3 million (or 4.5% of South Korea's population) gathering for peaceful demonstrations. The cumulative count of participants from late October to December adds up to more than 10 million. A motion of impeachment has been approved on December 9th in the National Assembly, and awaits final ruling by the Constitutional Court.

Coverage of this scandal by Korean media is noteworthy for several reasons. First, the event increased news audience. Television viewer ratings on news programs went up by 1.5 times and the total reactions on news stories on social media went up by 1.9 times. Update of news is particularly interesting given the trend that people's trust in mass media has been declining in Korea. Second, news media ranks changed. The Korean news influence until recently has been dominated by established newsrooms such as Chosun Daily, Joongang Daily, Kyunghyang, and MBC News that were founded half a century or little over a century ago. The political scandal, however, changed this ranking. A cable network-based newsroom, JTBC, that was only established several years took the lead over other half-a-century old established newsrooms. JTBC was the first to disclose a concrete evidence that sparked the presidential scandal. Its sheer focus and dedication to provide consistent high-quality coverage of the scandal might have helped gain the largest audience in terms of both television viewer ratings and social media reactions.

In this paper, we collect data from the Facebook pages of major newspaper publishers in South Korea in order to examine the changing media landscape over this notable political scandal. To quantify the differences in topical coverage and audience reactions across media outlets, we compile a set of topics seen in news headlines for each media outlet.

Then using the generated topic sets, we compare topical similarity across pairs of media outlets and visualize their topic networks. We also examine how each news outlet gained or lost their influence after the scandal.

This paper makes the following observations:

1. Similarity of news headline topics can be an alternative metric to quantify the degree to which media outlets concurrently cover a large societal event.

2. A visualization helps highlight topical similarity as well as dissimilarity across media outlets, which could be used to alleviate the biasing effect of filter bubbles by exposing its audience to diverse media stance.

3. A data-driven method adopted in this study uniquely captures the unprecedented scale of topical convergence in Korean media regarding the political scandal and a resulting rise of a new winner—a relatively young news network—in terms of audience influence.

This work is limited in that Facebook news audience is not representative of all news audience in Korea. However, the trends seen on this platform help us observe prominent changes in media ranks, which before the social media era would have been hard to observe closely. Hence our study is a preliminary yet a meaningful attempt to witness the changing dynamics in the Korean media landscape. Our analysis on social media reactions on news stories provides insights on how news scoops are interpreted and conveyed to online audience differently, which are critical to uncovering how people read and react to news online.

2. DATA METHODOLOGY

2.1 Data Collection

News posts studied in this paper were accessed through the Facebook Graph API over a single day period on January 10th, 2017. We gathered a total of 166,131 posts, which are the entire set of public posts of five news media outlets in South Korea. They include two of the popular right-leaning media, Chosun Daily and Joongang Daily, and two of the popular left-leaning media, Hankyoreh and Kyunghyang, respectively. Also included in this list is a relatively newly emerged cable news network, JTBC, that has become notable for releasing a number of exclusive news and scoops on the event.

Media	Data since	Posts	Total Likes
JTBC	2011/11/28~	17,323	12,612,111
Chosun Daily	2010/04/16~	26,032	25,210,818
Joongang Daily	2010/11/09~	47,089	10,634,283
Kyunghyang	2012/05/28~	33,001	31,655,183
Hankyoreh	2010/08/05~	42,686	27,194,615
Total	2010/04/16~	166,131	107,307,010

Table 1: Summary of the dataset

Table 1 displays key information about the gathered data indicating the data period, post count, and the aggregate Likes count of news posts. The five outlets opened their Facebook pages at different points in time, attributing to different time range of data between the outlets. For each news post, we collect information about the timestamp (i.e.,

time when a post was uploaded to Facebook), main text (i.e., short text description about any linked content), news headline (i.e., text headline of the news article that is linked on the Facebook post, if any), news video (i.e., video content that is directly uploaded to the page, if any), news image (i.e., image content that is directly uploaded, if any), as well as the count of Likes and Comments. To ensure that all posts had enough time to circulate within Facebook, we analyze only those news stories posted in 2016.

Three kinds of analysis methods are utilized in this study One is the Likes count for each Facebook post. Another is topic similarity of news posts. When examining news topics of each media outlet, we do not restrain by the aforementioned circulation time and use all recent data. Last is the topic network visualization, which assists understanding of the varying keywords covered by each media outlet. We explain these methods in detail.

2.2 Topic Analysis

To analyze topics, we combined the text data from the main description and the linked news headline, if any existed, for each post. A majority of posts (78.9%) contained at least one news link with a headline; the remaining posts were featuring news videos or images and did not contain any external links to news sources, in which case only the main text description was used for topic analysis. Next, the combined text was treated with the Korean morphological analyzer Komoran (via an external Java package) to tokenize words and remove stop words (e.g., 'a', 'the')[1]. Since the gathered posts were produced by newsrooms, language usage was clean and most content could be tokenized properly. To better assist the tokenizing process, we have manually added a set of keywords related to the event for named entity recognition based on the user dictionary provided by Komoran. For instance, we added keywords indicating the name of people and organizations involved in the scandal to the user dictionary such as 'Choi Soon-sil' and 'Federation of Korean Industries'.

After tokenizing, we are left with a potential set of keywords in the form of nouns, verbs, and adjectives in root forms. We assume a part of the text description and news headlines published by media outlets can be a candidate for a topic, rather than to extract subtext or the hidden meaning from the text. As one such approach, we further utilized nouns as possible keywords in this research [4]. In Korean, nouns are classified into several types: common nouns (e.g., 'cookie', 'bakery'), proper nouns (e.g., 'Oreo', 'Chips Ahoy'), dependent nouns (e.g., 'what'), and rhetoric and pronouns (e.g., 'he', 'she'). Among these noun types, we further consider proper nouns and common nouns as candidate topic sets, as they are the only nouns that can have independent meanings. Each post, hence, may contain one or more nouns in its topic set. On average Facebook posts contained 10.46 nouns as topics.

We then utilized TF-IDF (Term Frequency-Inverse Document Frequency) on the pre-processed topic sets (nouns) to identify prominent topics. TF-IDF is a measure of how important a given word is in a document and is normally used to calculate weighted similarity between a document and a term in a collection. This method was applied to differentiate topical differences across media outlets. We calcu-

---

[1]Komoran 2.0 Korean morphological analyzer
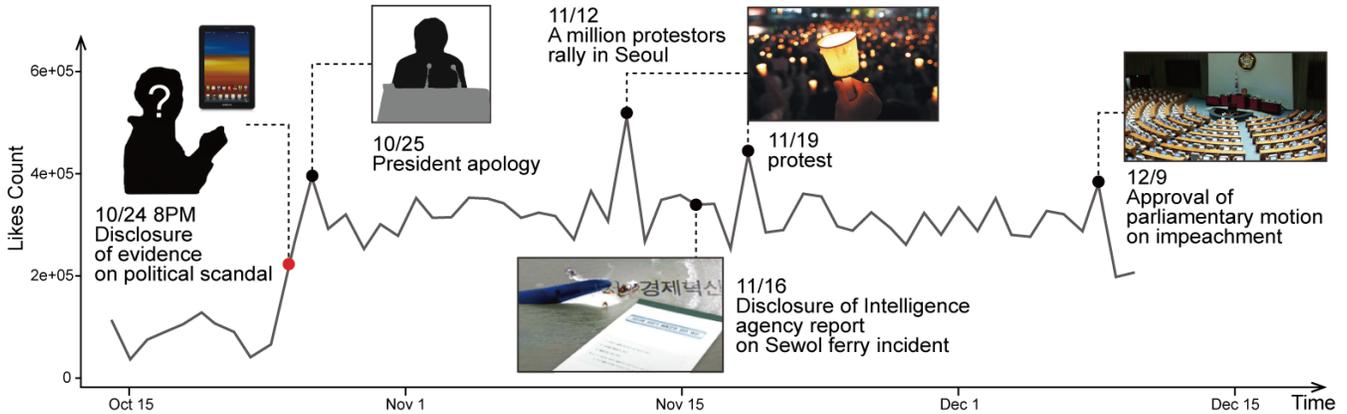http://www.shineware.co.kr/?page_id=835

Figure 1: Timeline of the 2016 South Korea's presidential scandal seen by the daily total Likes on Facebook news posts. The Likes count has been aggregated across all five news outlets.

lated similarity of the identified topic sets by constructing a term vector for each media page via a cosine function. The topical similarity between two news media outlets hence is determined by a value ranged between 0 and 1, where the value of 1 indicates a complete overlap in the topics covered and the value of 0 means otherwise.

2.3 Topic Visualization

Visualization was used to illustrate differences in topical trends across media outlets. Here, rather than individually treating each noun separately as a topic, we tried to identify any phrases or co-occurring patterns among topical nouns. This allows us to employ a network-based visualization, where a node represents a topic and an edge represents a link between two topics if they are *associated*. A number of association rules can be applied here, where the main idea is to highlight closely related nouns in the topic sets. This study relied on one such heuristic based on the Apriori algorithm [1], whose association rule tries to infer causal relationship. Keyword co-occurrences were examined at the post level. The algorithm allows a single keyword or a noun in a post to appear across multiple associations. For instance, a keyword 'president' can appear as a singleton and at the same time be associated with another keyword to form a word pair ('president','impeachment') or ('president','respond'). Such subtle difference in co-appearance of keywords in news headlines reveal the careful choice of words and political stance across media outlets.

In applying the association rule, we only considered top 150 nouns from each media outlets, so that only prominent associations are learned. We applied the arules package in R programming language [5]. The Apriori algorithm was run for each news post to form a topic network. Then, the resulting topic associations were visualized as a directed network by the Gephi tool, an open-source network analysis and visualization software [2].

3. RESULTS

Effect of the political scandal on the Korean news media landscape is highlighted from three perspectives: popularity of news stories, shifts in media ranks, and topical similarity of news headlines over time. We also examine an increase

in the production of 'exclusive' stories.

• **The public reception of the scandal measured by content popularity.** We first examine the temporal evolution of the 2016 presidential scandal from the public reactions and the media coverage, in turn. First, the public reaction can be represented by the aggregate Likes count received on news posts. This trend is shown in Figure 1, where the news story ('Disclosure of evidence on political scandal' in the figure) published on October 24th evening by the JTBC newsroom marks the beginning of the huge public interest. That news story was the very first instance for any news media outlet to disclose a tangible piece of evidence supporting the prior allegations on the president. Compared to the week before the scandal, news articles on Facebook received on average 1.9 times more Likes per post. While not shown in the figure, the total number of news posts per day, however, is similar to early October in that none of the five media outlets produced significantly more news posts on their Facebook page upon the event.

The public reaction to news posts shows peaks on certain days. The largest peak appears on November 12th, when the first million-people protest was organized at the Gwanghwamun Square in Seoul, which demanded the ouster of the president. Protests were organized regularly and voluntarily every Saturday and the graph shows subsequent peaks corresponding to such events, albeit smaller than the peak on the first million-people protest. It is also clear that some key news events do not correlate with peaks in Figure 1. However, it is still reasonable to assume that high Likes count levels and peaks generally represent public attention to important events.

We then extend our analysis to the news media's perspective on the scandal. In occurrence of a large event, we expect various newsrooms to report more on the event, regardless of their political stance. This will lead to a temporary increase in similarity in topics across news media outlets. Figure 2 shows similarity of news topics between two news outlets of opposite leaning, plotted from the beginning of 2016 to early 2017. Topical similarity at each data point (appearing every 3 days) is aggregated over an overlapping 2-week moving window. It is easy to see that similarity in news headlines
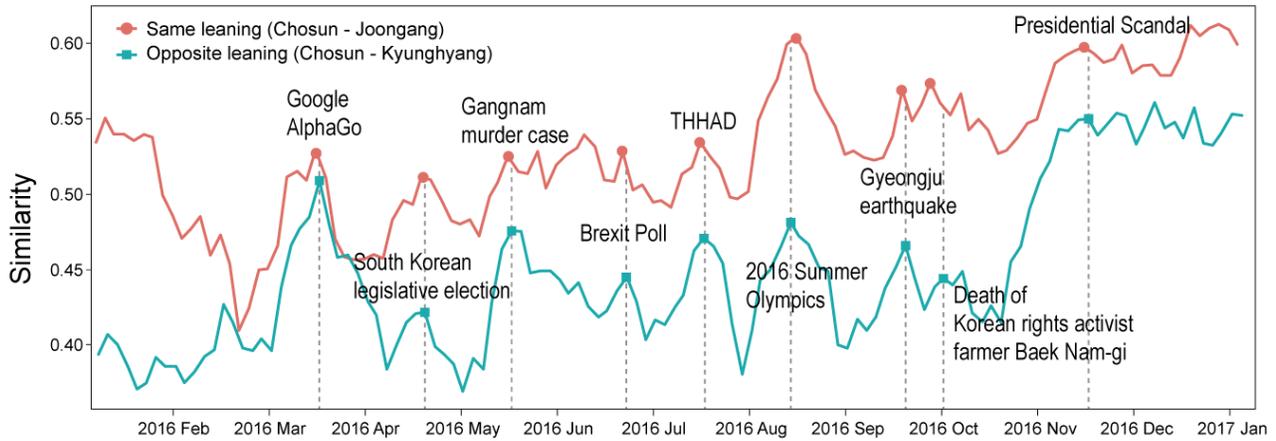
Figure 2: A one year-long trend on topical similarity of news headlines between two prominent news outlets: Chosun Daily (right-leaning media) and Kyunghyang Shinmun (left-leaning media)

peaks when notable societal events occur. Decreasing trend in similarity after the peak implies that newsrooms simply shifted their focus to other new topics or are providing follow up reports with different wordings due to their own interpretation of the major event.

Nonetheless, the political scandal in 2016 stands out among previous peaking events in two ways. First, the scandal is associated with the highest degree of similarity. Retention of high similarity levels is also unprecedented and forms a plateau rather than peaks as in other social events. Such observations indicate that effects of the scandal on media attention is longer lasting. This showcases a need for a deeper analysis on how news outlets adapt to the scandal over extended periods of time, which we aim in this study.

• **Shifts in media ranking and effects of exclusive news post.** While the previously introduced similarity graph calls for further analysis, the significant increase in Likes count in Figure 1 also raises question on which media outlet attributes to such increase in public attention. Figure 3(a) shows the average Likes count per post for each media outlet along with error bars measured in standard errors before and after the event. Like counts were calculated over a non-overlapping month period before and after October 24th. The political leaning of each news outlet is denoted under its name as R (right-leaning) or L (left-leaning). Most platforms show an increment in audience engagement. However, the most popular right-wing outlet that has the largest offline newspaper circulation in South Korea, Chosun Daily, shows a slight decrease in the Likes count. JTBC, the youngest newsroom, in contrast, more than triples the Likes count over the same time period.

This finding demonstrates that circumstances surrounding the political scandal did lead to a change in the South Korean media landscape in terms of which media the public favors. A further analysis on nature of news posts was performed to identify the underlying causes of this change. Among many factors that might have affected to this dynamics, here we focus on two such causes. One is that we saw a frequent mention of 'exclusive' in news headlines of JTBC posts. Hence we assume that exclusive news posts

on the scandal gathers more Likes than non-exclusive news, and that any media outlet's superiority on producing such news results in gathering more Likes than the other news providers.

Figure 3(b) shows the percentage of news that we deem as 'exclusive' stories for each news media outlet. For this analysis, Facebook posts between the time of Oct 24th to Nov 24th were classified as either i) exclusive news or ii) general news. This was done by looking for explicit expressions in the news headlines; all posts containing either 'exclusive news' or 'special news' were classified as an the exclusive news type, while all other posts were dubbed as the general news type. JTBC was the leading media when it comes to reporting more news scoops, at the selected time frame. In fact, the rate of news deemed as exclusive type across media outlets were proportional to the rate of likes news outlets gathered—as seen by the reverse ordering of media outlets in Figures 3(a) and (b). The results show that JTBC's rise in popularity can be partially explained by its higher percentage of exclusive news posts to those of other news media outlets.

In fact, news posts that were of the exclusive news type on average gained 2.17 times more Likes than the general news type in our data, indicating that the news audience appreciated the media's efforts to disclose new facts about the event. We also mention that, while JTBC undoubtedly contributed largely to reporting the scandal, other news outlets also covered non-negligible fraction of exclusive news in that this event was covered by multitudes of media outlets in a *collaborative* fashion. This could be confirmed by the varying topics of exclusive news covered by each media outlet; the newly published news stories were on slightly different topics from one media outlet to another, rather than repeating of the same narrative.

• **Convergence of news topic similarity.** After seeing how JTBC gained popularity, we are left with another question of how the other news media outlets, with varying political stances, reacted to the scandal. As a first step in answering the question, we explored the change of topical similarities for each news provider after the scandal. Fig-
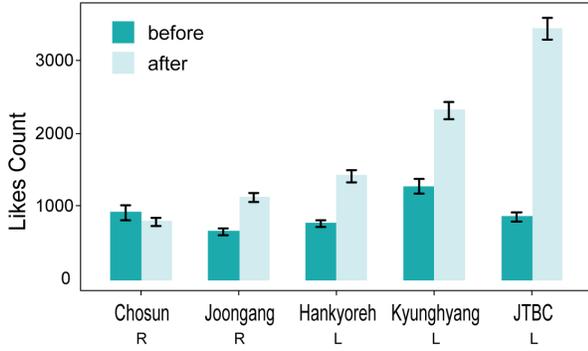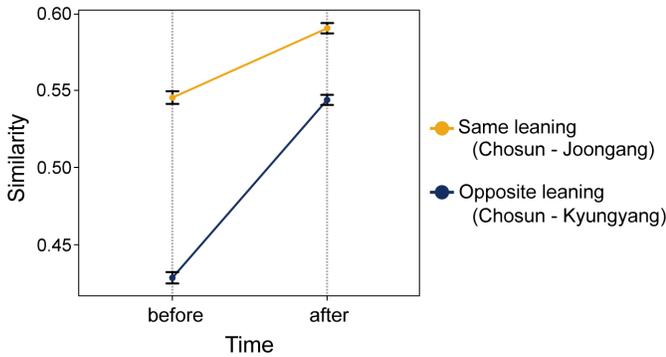
(a) Average Likes count per media
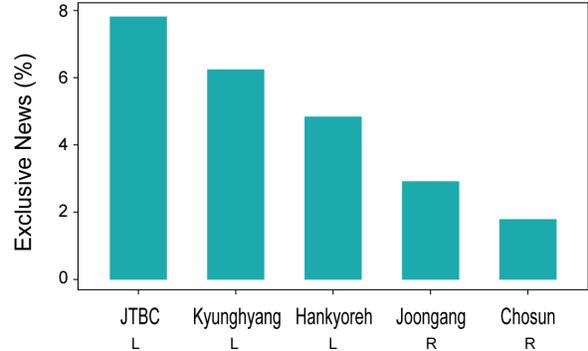
(b) Percent of exclusive news per media

Figure 3: Likes count change and exclusive news percentage for each media outlet

Figure 4: Topical similarity changes for different media pairs

ure 4 shows two such comparisons for the two right-wing media (Chosun Daily and Joongang Daily) and two right-left media pairs (Chosun Daily and Kyunghyang). Topical similarities, measured by the cosine similarity based on TF-IDF, news headlines between newsrooms of similar political leaning already exhibit high similarity before the event (i.e., nearly 0.5 or higher), whereas the cross-pairs cover different topics prior to the event. Upon the scandal, all media pairs converge in their topical similarity due to the significance of the event.

While omitted due to space limitation, other media pairs also showed changes; headline topics between all media pairs showed increased similarity to a different degree that reflects political leanings. This finding implies that each news provider, originally choosing topics that fit more amicably with its set political stance, opted to report on similar news topics after the political scandal. Thereby, the magnitude and gravity of the political scandal reached a certain threshold where a convergence of topics of the news media with different political viewpoints took place.

• **Divergence of topics over time.** Having recognized a sudden uptake of JTBC and the lasting impact of the scandal on the news media, we compare further how this newly emerged platform compares to other newsrooms during the course of time after the convergence of topics among all selected news media. Figure 5(a) shows a change of topical similarity between JTBC and the most established media, Chosun Daily. The week 0 in the x-axis represents October 24th and topical similarity at each data point is calculated in the same way as Figure 2. Topical similarity between the two news providers that start as low as 0.35 rises over the event to a peak of 0.52, at the disclosure of the political scandal. Basically, topical similarity increases if there exists more number of same topic pairs between the two news outlets and such increase is moderated by TF-IDF weights, which have differing values for different topics. After reaching its peak on the second week, headline topics start to gradually diverge over time.

In order to look deeper into such topic divergence, we visualized topical networks of the two media, where Figures 5(b) and (c) are network plots of headline topics covered by JTBC and Chosun Daily between December 22th, 2016 and January 5th, the following year. Topics that are directly related to the presidential scandal are marked in a bubble, whereas topics that are indirectly related or not related appear outside the bubbles.

Figure 5(b) shows that most of the news stories published by JTBC are about the scandal, indicated by the large coverage of the bubble. Prominent topics here are 'president', 'impeachment', 'recording', and 'judgment' in the graph. Some topics appear as not directly related to the event such as 'Article', 'Social department'. JTBC's headlines were mostly dedicated to the scandal in the time period we examined. The visualization shown in Figure 5(c) for Chosun Daily, in contrast, shows a smaller coverage of the bubble. A number of topics appear as not directly related to the scandal such as 'establishment of a new party', 'North Korea', and 'the THAAD US army missile system'. The difference in topical similarity is known to be caused by differing preferences of each newsroom to focus on the current biggest issue in more detail, or to focus on other topics unrelated to the political scandal, respectively. [3]

4. DISCUSSION

News content has been traditionally disseminated via platforms like television, newspapers, local radio, and maga-

(a) News headline similarity   (b) Topic network of JTBC   (c) Topic network of Chosun Daily
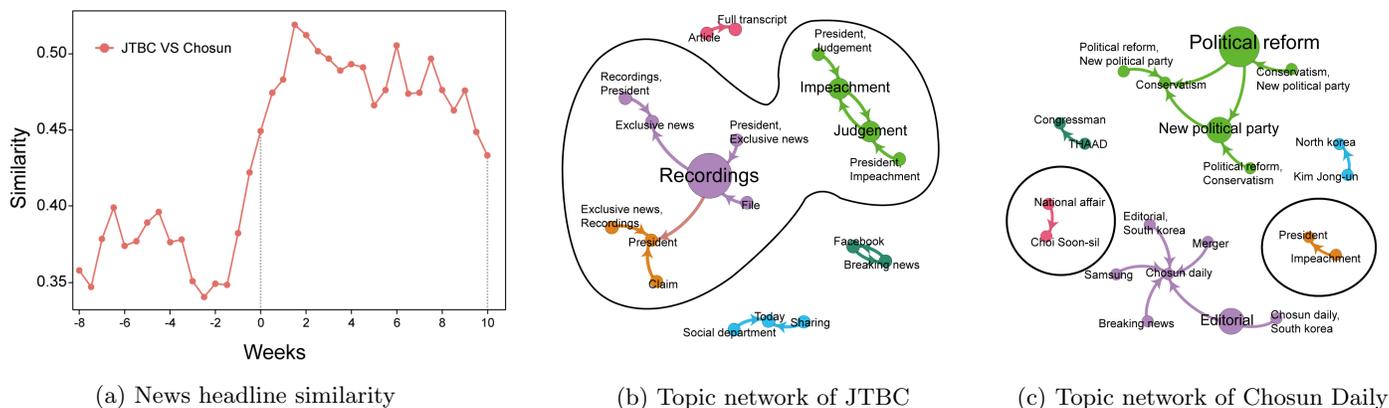
Figure 5: Diverging topical similarity in news headlines between two news outlets: JTBC and Chosun Daily

zines. Now more people are accessing news through the Internet and the social web. This paper utilized data that can be gathered from one such popular platform and analyzed it to report on news media coverage and public reactions of a large political scandal in South Korea in 2016. In the following we discuss some implications of our study and suggestions for future research.

First, our findings have implications to media outlets. Similarity and dissimilarity in headline topics can be utilized in positioning the newsroom's stance better and further be used to predicting how competing media outlets will cover a given story. Newsrooms can then re-evaluate what kinds of headlines are reacted popularly by the Facebook audience based on the topic network analysis (e.g., what kinds of keywords are used in such headlines).

Second, our methodology and findings can be used for building tools that assist news readers. An automated visualization algorithm that plots a timeline summary of key events based on detection of peaks [8] and plateaus could help news readers quickly catch up with trending news. Furthermore, the map of headline topics can provide useful information for interested news readers to learn the political stance of different media outlets. A comprehensive visualization tool that integrates both functions can easily introduce diverse perspectives to news readers, which can alleviate media bias caused by filter bubbles.

Third, this research contributes to journalism research in three major ways. One is uniquely capturing the converging and diverging headline topics through a large scandal across the right-wing and left-wing media. The degree of convergence and its lasting period are valid indicators in classifying exceptional public events, while the gradual process of topical divergence upon the onset of a political scandal opens up a new research topic on discovering how different media employ unique strategies to cope with their own set of audience over an extended period. Another contribution is in recognizing an uptake of a young cable news outlet (JTBC) in the Korean news media landscape; a swift coverage of exclusive news content likely contributed to this changing dynamics.

Finally, our analysis opens up multiple directions for future work: For instance, (1) Do other online (e.g., Twitter, news websites) and offline (e.g., newspaper readership) channels exhibit similar patterns observed on Facebook? (2) How do media and the public interact with each other in such big events? How does the public discourse affect or get affected by media? (3) Within topical convergence, what strategies do media outlets employ to further their political stance? (4) What do the patterns of bursts and peaks tell us about the nature of events themselves? For example, do smoother and longer peaks reflect prolonged public interest and high significance of an event? We plan to investigate these questions to better understand and support the intricate interplay between a societal event, a media landscape, and the public discourse.

5. REFERENCES

[1] R. Agrawal, R. Srikant, et al. Fast algorithms for mining association rules. In *proc. of the VLDB*, 1994.

[2] M. Bastian, S. Heymann, M. Jacomy, et al. Gephi: an open source software for exploring and manipulating networks. In *proc. of the ICWSM*, 2009.

[3] R. M. Entman. Framing: Toward clarification of a fractured paradigm. *Journal of Communication*, 43(4):51–58, 1993.

[4] V. Gupta and G. S. Lehal. A survey of text summarization extractive techniques. *Journal of Emerging Technologies in Web Intelligence*, 2(3):258–268, 2010.

[5] M. Hahsler, C. Buchta, B. Gruen, and K. Hornik. *arules: Mining Association Rules and Frequent Itemsets*, 2016. R package version 1.5-0.

[6] S. R. Kairam, M. R. Morris, J. Teevan, et al. Towards supporting search over trending events with social media. 2013.

[7] H. H. Khondker. Role of the new media in the arab spring. *Globalizations*, 8(5):675–679, 2011.

[8] A. Marcus, M. S. Bernstein, O. Badar, et al. Twitinfo: aggregating and visualizing microblogs for event exploration. In *proc. of the ACM CHI*, 2011.

[9] N. Newman, R. Fletcher, D. Levy, and R. Nielsen. *Reuters Institute Digital News Report 2016*. University of Oxford, 2017.

[10] M. Schudson. Notes on scandal and the watergate legacy. *American Behavioral Scientist*, 47(9):1231–1238, 2004.