

Original Research

Information Interrelations in the News Ecosystem: Mapping the Dual Dimensions of Contribution and Dependence Among Media Outlets

Chankyung Pak ¹ and So-Eun Lee ²

¹ Department of Media and Communication, Kyungpook National University

² Media School, Pukyong National University

Corresponding to

So-Eun Lee

Media School, Pukyong National University, 45 Yongso-ro, Nam-gu, Busan 48513, Republic of Korea.
Email: petsilver@gmail.com

Acknowledgment

This work was supported by the Pukyong National University Research Fund in 2021 (C-D-2021-0978) and the Ministry of Education of the Republic of Korea and the National Research Foundation of Korea (NRF-2024S1A5A8026512).

Disclosure Statement

No potential conflict of interest was reported by the author.

Received

15 May 2025

Revised

1 Aug 2025

Accepted

1 Nov 2025

ABSTRACT

This study investigates the structural interrelations of news production in South Korea's online news ecosystem, introducing the dual dimensions of *contribution* and *dependence* to explore how information flows are organized—and how they may impact content diversity. Using posterior text similarity measures from over 700,000 articles collected on the Korean portal site *Daum*, this study classifies media outlets into four types—mediators, sources, dependents, and isolates. Topic-level analysis reveals that a small number of dominant players—particularly news agencies, national dailies, and economic newspapers—serve as central information sources, while smaller outlets disproportionately rely on them, limiting diversity in hard news topics such as politics, economy, and international affairs. Conversely, soft news areas like entertainment show greater independent contributions from smaller media. The study also shows that organizational characteristics, particularly reporting capacity and article volume, significantly shape contribution but have less impact on dependence. These insights highlight the structural asymmetries in the ecosystem and suggest the importance of systemic approaches to support original reporting and foster a more balanced media environment.

KEYWORDS

information interrelations, contribution, dependence, news ecosystem, media interdependence, news diversity

News diversity has long been considered essential to public discourse, yet the rise of “platform-based news ecosystem” (Dwyer & Martin, 2017) has led to increasing content uniformity. Although digital platforms initially raised hopes for greater diversity by enabling new types of content producers and delivering personalized content to segmented users, the current environment is marked by the widespread circulation of near-identical stories. As Napoli (1999) noted, content diversity cannot be captured simply by counting sources;

rather, it requires attention to the plurality of information and perspectives conveyed across the ecosystem.

In response to growing concerns over *news homogenization*, a number of large-scale studies have turned to text similarity measurements to detect duplicated reporting and assess the extent of questionable journalistic practices (Lee, 2021; Vogler et al., 2020). While these approaches help identify repetitive patterns, they often treat similarity as inherently undesirable. Such views, however, overlook the institutional and structural conditions behind news production.

News content is a distinctive form of information: it is produced within a shared temporal and societal agenda, shaped by structural routines such as beat coverage, press releases, and standardized formats. Moreover, duplication is often not only tolerated but strategically necessary. Resource-constrained outlets may rely on externally produced content to sustain operations, while larger outlets may duplicate stories to maintain visibility in algorithmically driven environments. These practices—though rational in isolation—collectively erode the depth and diversity of the information landscape.

This study addresses key methodological limitations in existing news homogenization research. While prior studies have attempted to estimate how “dependent” some outlets are (e.g., Boumans et al., 2018) or how “influential” others might be (e.g., Shin & Youn, 2025), they rarely distinguish between these two conceptually distinct roles. As a result, media outlets’ dual functions—as reproducers and providers of information—have been insufficiently captured in existing models. To address this gap, we introduce the concept of *information interrelations*, which offers a two-dimensional framework composed of *contribution* (the extent to which an outlet provides information used by others) and *dependence* (the extent to which it relies on information produced elsewhere).

We apply this framework to a large-scale dataset

of over 700,000 news articles published on the Korean portal *Daum* over the course of one month. By computing posterior similarity—a time-sensitive measure of textual similarity between articles—we estimate the flow of information between outlets and classify them according to their contribution and dependence levels. Additionally, the study explores how contribution and dependence differ by news topic and organizational characteristics.

This approach moves beyond moral judgements about news similarity, and instead frames it as a structural issue rooted in the survival strategies and inter-organizational relationships of media outlets within a portal-centered news distribution environment. In a system where news is consumed by article rather than by outlet, what matters is not only who publishes what, but how information circulates and accumulates across the ecosystem. By analyzing the directional flows of information, this study seeks to illuminate the underlying architecture of informational power. Rather than measuring diversity directly, the study diagnoses how structural asymmetries and relational positioning shape news production and redistribution in Korea’s distinct media environment.

THEORETICAL BACKGROUND

Reconsidering News Homogenization

Napoli (1999) proposed three dimensions of news diversity—source diversity, content diversity, and exposure diversity. While early research focused largely on source diversity, studies in the digital era increasingly examine exposure diversity, emphasizing the role of algorithmic recommendations (Reid, 2025). However, renewed concerns about rebuilding trustworthy information ecosystems have shifted scholarly attention back to content diversity. The explosion of news outlets has not translated into

greater diversity of content or perspectives. This is largely because today's online news environment is saturated with strikingly similar articles, often repeated within and across outlets and prominently displayed on news portals.

The concept of *news homogenization* captures this phenomenon, where different outlets report the same stories with little variation (Boczkowski & De Santos, 2007; Lee, 2021). Media organizations frequently observe and reference each other when selecting events and structuring coverage (Ahn, 2021; Boczkowski, 2009; Lee, 2021). As speed and quantity frequently outweigh journalistic quality in the online environment, news outlets often rely on press releases and competitor content. Concepts such as *churnalism*¹ (Davies, 2008; Jackson & Moloney, 2015; J. Lewis et al., 2008), plagiarism (N. P. Lewis, 2008, 2013), reproductive reporting (Ahn, 2021), abusing (D. Jung & Kim, 2016), news cannibalism (Phillips, 2011), and progeny churnalism (Johnston & Forde, 2017) all underscore the growing trend toward replication over original reporting. These practices, centered on reusing readily available online material rather than producing independent journalism, deepen content homogeneity. Thus, in order to assess each news outlet's true contribution to content diversity, it should be noted that content diversity is "a condition of a media system or sector and not of an individual 'channel'" (McQuail, 1992, p.146).

Empirical literature on news homogenization addresses the system-level perspective on content diversity. These studies commonly employ text similarity measures, as such a perspective entails large-scale assessment. In particular, researchers identify excessively similar articles by applying specific similarity criteria. For example, J. H. Lee (2021), using *doc2vec* embeddings and cosine distance, found that 19.86 % of articles covering the 2021 Seoul mayoral by-election exhibited

extreme textual similarity with wire services and economic newspapers playing a significant role in the production of duplicate news. Boumans et al. (2018) use Levenshtein distance to detect which article is simply copied from news agencies' products. Vogler et al. (2020) similarly employed a bag-of-words model and Jaccard similarity to detect "twin articles," revealing that corporate ownership structures often drive identical reporting across outlets.

Two characteristics of this line of research are notable. First, these studies typically adopt specific similarity criteria to determine whether an article is considered reproduced. This approach assumes that a certain level of information duplication renders a reproduced article "undesirable." However, in some domains—such as politics and international relations—it is entirely normal to follow an agenda by citing previously reported information and adding new perspectives, thereby enriching the agenda. While such outlets may score high on reliance depending on the criteria used, it is questionable to criticize them for copying.

Second, some studies conceptualize node properties in terms of *influencers* (Shin & Youn, 2025) while others focus on *followers* (Boumans et al., 2018). This difference highlights the two distinct dimensions implied by similarity relations among outlets. From a decision-making perspective, copying others serves an outlet's own interests whereas producing reproducible information benefits the market as a whole. This reflects the *positive externality* and *public good* nature of news reporting. Contributing to the market may align with self-interest when reproduction is compensated (e.g., by contract), but often it does not, depending on the nature of the information.

Thus, analyses of news-information relationships should move beyond a unidimensional focus

¹ Churnalism is a compound of "churn out"—implying fast and effortless mass production—and "journalism" (Harcup, 2004). It describes the reliance on recycled content rather than original reporting.

on *undesirable similarity*. Since news inherently addresses shared social agendas, a certain degree of similarity is inevitable. Moreover, media organizations may strategically draw on other outlets' reporting, depending on established production routines or financial constraints. In this sense, content diversity is shaped not merely by individual choices but by broader structural dynamics.

An Ecosystem Approach to News Content Diversity

Viewing news as an ecosystem clarifies this further. Drawing on natural ecosystem analogies (Dimmick, 2003; Dimmick & Rothenbuhler, 1984; J. Jung & Park, 2012), scholars argue that media organizations, like species, compete for limited resources—readers' attention, money, satisfaction, and advertising revenue. As resources are unevenly distributed across geography, ideology, and demographics, not all outlets compete for the same audiences, enabling diversity through differentiated niches. From an ecosystem perspective, diversity in the news ecosystem arises from competition, differentiation, and mutual dependence among organizations.

The online environment intensifies both competition for user attention and interdependence over information as a unique resource. While the copying practice is often linked to the former, the latter is more closely tied to each outlet's niche and strategic position. This duality arises from the nature of information in the news industry, which serves as both an input and an output. Even when media outlets compete for exclusive reporting in specific topical domains, the resulting information, once published, becomes available for others to use. As such, news organizations must strategically choose between producing original information and utilizing existing materials. Interdependence here is less direct competition and more symbiosis within the

ecosystem.

Although attention-economy logic suggests increasing homogenization, empirical studies indicate that inter-reliance among news outlets can also sustain diversity through complementarity, generating a symbiosis among media species and, in turn, contributing to enduring diversity (Pak et al., 2020; Weber, 2012). Based on link analysis, these studies suggest that news organizations, much like biological species, occupy diverse ecological niches, and that inter-species dependencies may support the survival and coexistence of otherwise vulnerable actors.

Further echoing our criticism of the news homogenization literature, ecosystem studies have highlighted different facets of inter-reliance among news outlets. Rather than applying fixed criteria to determine *excessive* similarity, these studies measured the *degree* of information sourcing by counting direct references or hyperlinks (Pak et al., 2020; Weber, 2012; Weber & Monge, 2011). By distinguishing between *mutualism* among legacy media and *commensalism* between legacy and emergent outlets (Weber, 2012), or between *cooperative* and *adversarial* coexistence of partisan and mainstream outlets (Pak et al., 2020), they demonstrated the multidimensional nature of inter-reliance.

The study most relevant to the present work is Weber and Monge (2011), which separately estimated authority, source, and hub scores from hyperlink structures. They reinterpreted Kleinberg's (1999) HITS algorithm within the context of news information flow: rather than simply counting links, they conceptualized authorities as trustworthy information sources and hubs as actors that collect and regenerate information. They argue that these functions are *mutually reinforcing* and together sustain the overall news ecosystem.

In sum, in an environment where information functions as a public good, media organizations engage in both competition and symbiosis to survive. Content homogeneity arises not

only from problematic practices but also from strategic responses to structural pressures. To fully understand these dynamics, it is necessary to measure each outlet's contribution to the entire ecosystem *and* reliance, which constitute separate dimensions of node property according information retrieval and news ecosystem literature.

Information Interrelations as a Dual-Dimensional Framework

To better capture the strategic choices within the online news ecosystem, this study reframes previous approaches to news similarity and homogenization through the concept of *information interrelations*. Building on ecosystem and network-based perspectives, we argue that informational relationships among media actors are not uniform but composed of distinct directional flows. It further distinguishes this concept into two dimensions: *contribution* and *dependence*. There are several reasons why content produced by different media outlets may appear similar. First, similarity may result when a media outlet becomes a source that others rely on. For example, outlets that deploy more field reporters may be more likely to have their original coverage

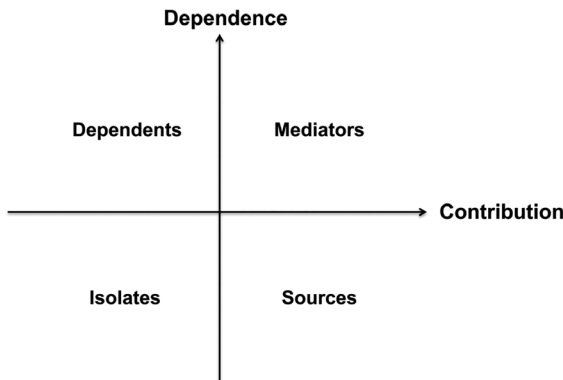
picked up by others. This study defines the extent to which a media outlet provides content that others use as *contribution*. Second, similarity may arise when one outlet depends on another for reporting. For instance, an outlet may rely on wire services or republish stories originally produced by another organization. This study defines the extent to which a media outlet uses information produced by others as *dependence*.

These two dimensions decompose news similarity by considering inter-organizational information flows. Yet, they are not necessarily two sides of the same coin. Some outlets produce significant original content while also referencing others; others may neither contribute much nor rely heavily. As such, media outlets may function both as providers and reproducers of information within the news ecosystem.

Treating contribution and dependence as distinct dimensions allows for the construction of a conceptual quadrant, capturing the diverse positioning of outlets within the information ecosystem. Plotting outlets along contribution (x-axis) and dependence (y-axis) yields four types: *Mediators* (first quadrant), *Sources* (fourth quadrant), *Dependents* (second quadrant), and *Isolates* (third quadrant).

If both contribution and dependence are high,

Figure 1. Four Types of Media Outlets by Contribution and Dependence



the outlet is not only heavily reliant on others but also actively producing new information that other outlets rely on. This reflects strong integration within the broader information ecosystem. If contribution is high while dependence is low, the outlet primarily functions as a source of original reporting, but contributes less to the reproduction of discourse generated within the ecosystem. If dependence is high and contribution is low, the outlet follows prior reporting from other organizations without contributing much original content. Finally, if both contribution and dependence are low, the outlet is relatively disconnected from the overall ecosystem and produces content that is more isolated.

Meanwhile, not all news similarity stems from inter-organizational dependence; some arises naturally from covering the same events—topic-driven similarity. Some may specialize, relying on external content in some areas while producing original coverage in others. To address this issue, this study calculates contribution and dependence by news topic and introduces a decomposition method that separately estimates and controls for topic-specific common information similarity. This constitutes an improvement over the HITS algorithm, which did not account for topic-wise text similarity.

Through this decomposition, the study seeks to uncover differentiated symbiotic relationships within the news ecosystem and their implications for content diversity. The research questions guiding the study are:

RQ1-1. How are media organizations distributed based on their information interrelations, as decomposed into contribution and dependence?

RQ1-2. How does the distribution of media organizations differ across different news topics?

Additionally, the study introduces *net contribution*,

calculated by subtracting dependence from contribution. While it is less suited for capturing complex roles like mediation or segmentation, net contribution highlights outlets that provide original information without heavily relying on others—a useful indicator for identifying core contributors to the ecosystem.

Determinants of Information Interrelations

If media organizations differ in their levels of information interrelations, what factors account for these differences? Under what conditions do certain outlets contribute to the news ecosystem, while others rely on their peers?

Previous studies suggest that financial resources, ownership structures, and production practices influence information dependence. Outlets lacking independent business models and relying on platform-based revenue often lose the capacity for original reporting (Bell et al., 2017; Nielsen & Fletcher, 2022). Although some financially stable media companies have adopted subscription models and prioritized quality journalism (Klopčič et al., 2020; Picard, 2011, 2014), most outlets depend heavily on advertising-driven models, reinforcing churnalism.

Ownership consolidation also exacerbates dependence. Assmann (2023) documents how many German regional outlets absorbed by national dailies now operate as “zombie papers,” maintaining brand names but lacking independent reporting functions. These trends highlight that source diversity may be illusory when structural consolidation occurs.

In South Korea, longstanding news production practices deepen information dependence. Journalists frequently share information, cross-reference each other’s coverage, and engage in pack journalism (Park et al., 2020; Song & Lee, 2007). Access to sectors like the Prosecutors’ Office and Ministry of National Defense is limited to a small number of outlets, creating heavy dependence on the news they produce. Within

these beats, information is often shared among informal reporter networks, often called as *kumi* (くみ組). These practices show that information dependence stems not only from the digital environment but also from entrenched industry norms.

Focusing on these dynamics, this study examines four factors: legacy status, media type, resource levels, and news topics.

First, differences are expected between legacy media and newer outlets. Although market forces strongly influence the online news environment, organizational norms and ideals rooted in traditional journalism continue to influence editorial decisions. As Lowrey (2012) argues, the evolution of the news ecosystem is best understood as a long-term process of adaptation to disruptive technologies, during which traditional journalistic ideals continue to serve as a source of institutional legitimacy. Over time, emerging outlets—including partisan websites, political blogs, fact-checkers, and news curators—undergo a process of normalization, gradually aligning with legacy standards to expand their legitimacy and audience reach. Thus, if organizational ideals of independent reporting prevail, legacy outlets should contribute more original content.

Second, differences are expected across types of media organizations, such as between national and regional newspapers or between terrestrial and regional broadcasters. Media organizations do not adapt uniformly to the news ecosystem; adaptation is a long-term process shaped by technological, economic, ideological, and audience-related factors. As a result, various types of media organizations coexist, reflecting distinct institutional structures and resource conditions. Information interrelations is closely tied to production costs and staffing levels. Although imperfect, media type serves as a practical proxy for differences in financial and human resources.

Third, differences are expected based on the

resource scale of media organizations, indicated by metrics such as the number of journalists and the volume of articles produced. Outlets with limited resources may adopt strategies that prioritize content reproduction or focus on specific topical domains, while better-resourced organizations are more likely to invest in original reporting across a wider range of areas. Consequently, resource scale is anticipated to significantly influence the degree to which outlets contribute new information across different news domains.

Fourth, variations in information interrelations are also expected by news topic. Although prior studies have explored topic selection patterns (Maier, 2010; Stempel, 1985), little attention has been paid to how inter-organizational dependence differs across topics. However, previous studies suggest that the strategies media organizations adopt for news distribution on social media reflect an intersection of popularity, economic incentives, and institutional goals (Pak, 2019). For instance, financially constrained outlets tend to concentrate on soft news, whereas legacy organizations—with stronger editorial infrastructures and specialized reporters—are more likely to produce original reporting on hard news topics such as politics, international affairs, and finance.

Based on the above discussion, the following research questions are proposed:

- RQ2. What factors influence information interrelations?
- RQ2-1. Do information interrelations vary according to the legacy status of media organizations?
- RQ2-2. Do information interrelations vary according to the type of media organizations?
- RQ2-3. Do information interrelations vary according to the resource scale of media organizations?
- RQ2-4. Do information interrelations vary according to the news topic?

METHODOLOGY

Data Collection and Preprocessing

To measure media interrelations on an online platform environment, this study utilizes large scale news data scraped from the Web. The dataset consists of 701,893 articles published by 143 media outlets on Daum, one of South Korea's major online news portals, during December 2022.

According to *Media Users in Korea 2023*, a report by the Korea Press Foundation (2023), the vast majority of South Korean news users access news through internet portals. As of 2023, the usage rate of online news aggregators stood at 69.6%, a significantly high figure compared to many other countries. Among portal news users, 92.1% use Naver and 23.1% use Daum. Despite Daum's lower usage rate, this study focuses on Daum for both practical and structural reasons.

First, although Naver and Daum differ slightly in their evaluation processes for media partners whose articles appear on the main page, there is little difference in the number of search-affiliated media outlets. As of 2023, Daum had approximately 1,350 affiliated media outlets (Ji, 2023), while Naver had 981 (AD-PR & Media, 2023). Given their structural similarity in content supply mechanisms, Daum can serve as a representative model of Korea's portal-based news ecosystem. Second, data scraped from Daum may be closer to the full set of published news stories than data from Naver. Interviews with media professionals conducted during this study indicate that while articles are typically submitted to both portals, deletion and management efforts focus more heavily on Naver due to its greater visibility. Consequently, duplicated or corporate-sponsored articles circulated within informal journalist networks (*kumi*) often remain on Daum but

are removed from Naver. This makes Daum particularly suitable for examining the underlying interdependencies and information flows that shape news homogenization. Third, Daum offers more transparent and consistent access to article-level metadata. Because Naver allows greater editorial autonomy in article formatting, essential metadata such as publication time and reporter names are not consistently structured, complicating automated collection. While Naver provides a news search API, it significantly limits the number of retrievable articles. By contrast, Daum's standardized HTML structure facilitates large-scale scraping and allows for more comprehensive and replicable analysis. An & Kwak (2017), for instance, used Daum's news data to analyze patterns of news consumption, illustrating its practical utility and reliability as a data source.

The news crawler developed for this study utilized the Python requests library to visit the "All News" page on Daum. The scraper accessed each collected URL to extract the article text, the news topic categories assigned by Daum, bylines, and publication date and time. The code used for this scraping process is provided in online Appendix 1.²

For the sake of analytical convenience, we collapsed the automatically assigned news topics into six categories. We retained *Economy*, *International Relations*, *Politics*, and *Social* categories as-is. But the following subcategories were merged into a single *Sports* category: 'golf', 'basketball', 'volleyball', 'baseball', 'international football', 'football', 'sports', 'e-sports', 'international baseball.' Similarly, we grouped 'culture', 'broadcasting', 'entertainment', 'movie', 'general entertainment', 'drama', 'pop music', 'international entertainment', 'entertainment hot issues' into *Culture/Entertainment* category.

In addition, we gathered information about news organizations from *The Korea Press Yearbook 2022* (Korea Press Foundation Industry

² The online appendices are available at <https://osf.io/hqg8z/>

Analysis Team, 2022). Specifically, we used the age of each organization as a proxy for their institutional legacy, the number of international correspondents and the number of press award-winning as indicators of its commitment to traditional news reporting. Taken together, these measures capture each organization's legacy status. Further, we calculated the number of unique reporter names from the byline data as a proxy for news organization's resource scale.

Finally, we utilized the categories offered by the Yearbook to reflect typology of news outlets. Yet again, we collapsed some of the categories. These categories for news outlets are all defined according to Act on the Promotion of Newspapers,³ Broadcasting Act,⁴ and Act on the Promotion of News Communications.⁵ Specifically, we took *National Newspapers*, *Regional Papers*, *Internet News*, *Magazines and News Agencies* as they are, but collapsed Financial News, Sports News, Newspapers in Foreign Language, Other Specialized Newspapers into *Specialized Newspapers* category. One large deviation we took is categorizing national TV networks as one group, and regional broadcasters and specialized cable channel providers as another. This is because the law still categorizes TV channel providers based on the physical medium they use—that is, terrestrial, cable, and so on—while audiences no longer feel these differences. The size and internal organizations and norms of news outlets are rather determined by the breadth of audience (Dwyer, 2010). Thus, we define National Broadcasters as TV networks that target a national audience with comprehensive programming, whereas Regional/Program Providers (PP) refers to channels that serve specific regions or interest-based subgroups of viewers.

Before the analysis, the standard text pre-

processing procedures were applied. In particular, nouns and verbs were detected and extracted using the OKT (Open Korean Text) morphological analyzer, and stopwords were removed. A unigram document-term matrix (DTM) was then created, including only nouns and verbs that appeared in more than 1% but less than 90% of all articles. Finally, non-Korean characters were removed from the dataset.

Construction of Indicators

Measuring News Similarity: Posterior Similarity

To measure information interrelations, this study employs *posterior similarity*, calculated as the average text similarity between each outlet's articles and articles *previously published by other outlets*, taking into account the timing of publication. Posterior similarity represents how similar an outlet's articles are, on average, to those already published by other media organizations. The posterior similarity calculation using the news text database collected from Daum was conducted in following steps.

First, a matrix was constructed where each row represents a news article and each column represents a feature vector corresponding to linguistic attributes of the article. While various methods could be used to create such a matrix, including topic distributions or embedding matrices derived from neural network-based algorithms like *doc2vec*, this study employed the simplest approach, constructing a Document-Term Matrix (DTM) based on the frequency of terms appearing in each article. The matrix was sorted chronologically by the time of article publication to ensure that similarity is measured only against articles already published at the time of comparison.

Second, using the DTM constructed in the

³ https://elaw.klri.re.kr/eng_service/lawView.do?hseq=42723&lang=ENG

⁴ https://elaw.klri.re.kr/eng_service/lawView.do?hseq=49315&lang=ENG

⁵ https://elaw.klri.re.kr/eng_service/lawView.do?hseq=46197&lang=ENG

previous stage, pairwise distances between articles were computed. Among the various available distance metrics, this study adopted cosine similarity to measure the proximity between articles. Cosine similarity is a natural choice when distance is measured based on frequency-based DTMs, as it reflects the semantic information contained in the document rather than its length.⁶ Third, from the distance matrix, articles published prior to the focal article were selected. Since the matrix was chronologically sorted by publication time, articles located above the focal article represented earlier publications. From all previously published articles, *the most similar one from each outlet* was identified and recorded in a new table. This procedure was repeated for every article in the database. The similarity scores calculated for each article were then aggregated at the outlet level by taking the average similarity for each media organization. This average was used as the outlet's measure of *posterior similarity*. The Python code used for text preprocessing and calculating inter-organizational posterior similarity is provided in online Appendix 3.

An intuitive interpretation of posterior similarity is the tendency of an outlet to publish articles that are semantically similar to those previously published by another specific outlet. Even though posterior similarity is designed to account only for the previously published news stories, it cannot be interpreted as isolating a specific outlet's tendency to rely on others as such. This is because posterior similarity can also be high when the counterpart serves as a common information source to the entire news ecosystem—a reflection of the counterpart's tendency. This can also occur when news topic inherently involves common agents, institutions, or organizations across different

stories, which naturally results in higher textual similarity.

Decomposing Posterior Similarity Using a Bayesian Gravity Model

Now to disentangle the multiple sources of posterior similarity and estimate the underlying ('true') dependency, we analyze the network structure of posterior similarity among news outlets. In particular, we implement the following Bayesian gravity model, which decomposes posterior similarity into topical similarity, dependence of a media outlet and contribution of the counterpart. The statistical model about topic t is as follows:

$$y'_{ij} = \alpha' + d'_i + s'_j + \epsilon'_{ij}$$

where y'_{ij} , d'_i , and s'_j respectively denote posterior similarity from outlet i to j , outlet i 's dependence on j and outlet j 's contribution. Essentially, the gravity model above decomposes the dyadic and directional relationships into the subcomponents. The model has two more components. α' is the intercept that is common within the same topic, which is designed to capture the topical similarity, and finally ϵ'_{ij} is the error term for this statistical model.

Unlike the centrality scores that are commonly used in network analysis, this method separately generates two distinct measures which reflect distinct tendency of outlets in forming interdependent relationships across outlets. While conceptually similar to the HITS (Kleinberg, 1999) that simultaneously calculates hub and authority scores from the hyperlink structure, our model can also capture and thereby control the topic-wise inherent text similarity.

Since each outlet retains two latent variables,

⁶ When documents are quantified using different approaches, different distance measures become more appropriate. For instance, if DTMs are constructed in a binary fashion, the Jaccard distance is a natural choice, as shown by Vogler et al. (2020). Yet, binary DTMs discard much of the available information about word frequencies. If the analysis concerns only string-level copying—as in Boumans et al. (2018)—the Levenshtein distance is suitable. However, when the goal is to capture content-level similarity or difference between documents, it tends to exaggerate distances and thus becomes less informative.

the number of latent variables equals twice the number of news outlets in the data set. Thus, we adopted a Bayesian technique that has greater flexibility in estimating large set of latent variables and its uncertainty.

In order to partially pool each outlet's unobserved tendencies, rather than separately estimate each, we assume the following hierarchical structure where each outlet's inbound dependence and outbound contribution share the same prior:

$$\begin{aligned}
 a'_i &\sim N(0, \sigma'_a) \\
 s'_j &\sim N(0, \sigma'_s)
 \end{aligned}$$

Lastly, priors are assumed to be *half-N*(0,2) for σ'_a and σ'_s , and *N*(0,5) for a' . These priors are relatively neutral, meaning that they do not place substantial prior weight on any particular regions of the parameter space. Further, approximately 4,000 dyadic relationships inform model updates for each news topic, making it highly likely that the data-driven information dominates the prior assumptions. For estimation, we use Stan (Stan Development Team, 2024), a widely adopted probabilistic programming language.

RESULTS

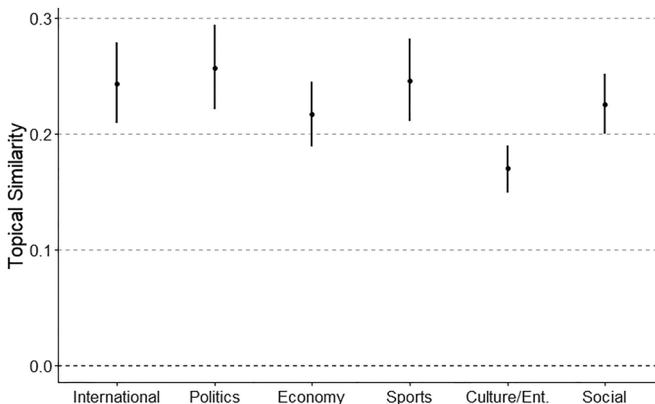
Distribution of Information Interrelations

Figure 2 presents the estimated baseline topical similarity, represented by the intercepts for each news topic. The x-axis displays the news topics, while the y-axis represents the estimated level of topical similarity.

The results indicate that international, politics, and sports exhibit relatively high intrinsic content similarity. This suggests that coverage in these domains is shaped more by topic-driven constraints than by selective editorial choices. In contrast, the culture/entertainment category shows a notably lower similarity. This implies that outlets exercise greater discretion in selecting and curating stories within this domain.

Table 1 reports the estimated contribution by news topic. News agencies such as Yonhap News, News1, and Newsis consistently rank among the top three in across most topics, except for culture/entertainment. Beyond the top three, rankings vary. In international news, KBS and SBS show high contribution, though this likely reflects the frequency with which their content is cited by others rather than a significantly greater reporting

Figure 2. *The Estimated Topical Similarities (α')*



capacity, as they do not deploy significantly more foreign correspondents than national newspapers (Korea Press Foundation, 2023).

Economic newspapers such as *AsiaEconomy*, *HeroldBiz* and *MoneyToday* also rank high not only in economy, but also in politics and society—likely due to their high article volume, which exceeds that of national dailies by 1.5 to 2 times. In contrast, in sports and culture/entertainment, this dominance fades. National dailies such as *Seoul Shinmun*, *Chosun Ilbo*, *JoongAng Ilbo*, and *Kookmin Ilbo* rise in sports, while *YTN* replaces *Newsis* in the culture/entertainment top three. Also, it should be noted that overall scores are lower for the culture/entertainment topic compared to other categories.

Table 1 also reports the ten outlets with the lowest contribution. No distinct pattern emerges, but it is important to note that low or negative scores do not imply high dependence. For instance, the finding that other outlets rarely depend on EBS for international, political, economic, or social news is consistent with common expectations. Yet this does not necessarily mean that EBS relies heavily on other sources as it is specialized in educational programming. Low contribution should not be equated with high dependence overall.

Table 2 presents the estimated dependence by news topic. Compared to contribution, these results are less intuitive to interpret. A high dependence score suggests that an outlet's reporting in a given topic draws—at least partially—on content previously published by other media. Major outlets like *Sisa Journal*, *KHAN* (*Kyunghyang Shinmun*), and *JoongAng Ilbo* frequently appear among the top-ranked, which may seem counterintuitive at first. However, high dependence does not imply undesirable practices such as copying. Rather, it may reflect legitimate editorial practices, such as synthesizing existing reports to offer broader coverage of ongoing issues. In this sense, dependence can be interpreted neutrally—as

an indicator of how closely an outlet's reporting remains embedded within the broader media discourse.

Conversely, outlets with low dependence tend to produce content that is more disconnected from prior coverage. In cases like *Yonhap News*, the largest news agency in Korea, this likely reflects its role as an original content provider. In others, such as EBS, an educational TV channel, it may indicate a thematic detachment from the dominant news agenda altogether.

The findings so far suggest that contribution and dependence, while both describing inter-organizational information relationships, reflect distinct dynamics—ranging from passive reproduction to active participation in the broader information ecosystem. To clarify the various positions on the interdependence network that news organizations assume, Figure 3 plots media outlets on a quadrant graph generated by the two dimensions of information interrelations. This empirical visualization builds upon the conceptual framework presented in Figure 1, illustrating how individual outlets are positioned as sources, dependents, mediators, or isolates within the ecosystem.

The figure highlights several key patterns. The news agencies, *Yonhap News*, *Newsis*, and *News1*, consistently act as core information sources as visualized by their location on the fourth quadrant (low dependence/high contribution). By contrast, outlets such as *Jeonju MBC*, *Gangwon Domin Ilbo*, and EBS consistently occupy the third quadrant across all topics, reflecting their relatively isolated roles within the broader ecosystem. National newspapers are primarily located in the first quadrant as mediators, absorbing and reprocessing mainstream information. Meanwhile, many regional broadcasters and cable PPs are positioned in the second quadrant as dependents, likely reflecting resource limitations.

By topic, more outlets appear as dependents in international and economic news, whereas

political news sees more mediators. It suggests that the former relies more on simple transmission of earlier reporting by well-resourced outlets, while the latter involves active framing and perspective-building. In culture/entertainment, most outlets cluster near the origin, implying lower levels of both contribution and dependence, and thus more independent coverage across organizations.

Another way to examine the relationship between contribution and dependence is through net contribution, calculated by subtracting dependence from contribution. This metric highlights the extent to which an outlet offers original information with less dependence on others. While less suited for capturing complex interrelations, it serves as a clear indicator for

Table 1. *Estimated Contribution by News Topic*

Rank	International	Politics	Economy	Social	Sports	Culture/Ent.
Highest						
1	YP .25	Newsis .20	News1 .18	Newsis .19	YP .20	YP .12
2	News1 .15	YP .20	Newsis .17	News1 .18	News1 .17	YTN .10
3	Newsis .14	Newsis .17	YP .17	YP .17	Newsis .13	News1 .10
4	KBS .13	Joongang .12	Asia Economy .15	Asia Economy .12	Hani21 .13	eDaily .10
5	SBS .13	Segye .11	eDaily .14	Segye .11	eDaily .11	Newsis .09
6	HaroldBiz .13	Asia Economy .11	Money Today .13	NoCut .09	Seoul Shinmun .10	JTBC .08
7	Asia Economy .11	eDaily .11	HaroldBiz .13	HaroldBiz .09	Chosun .10	MBN .08
8	Joongang .11	HaroldBiz .11	ChosunBiz .12	Kookmin .08	Joongang .10	Kookmin .08
9	YTN .10	KBS .10	MoneyS .11	Money Today .08	Kookmin .08	MoneyS .08
10	Seoul Shinmun .09	Kookmin .10	Seoul Daily .10	Seoul Shinmun .08	MoneyS .08	TheFact .06
Lowest						
10	ZDNet -.10	Joongang Sday -.12	L'Officiel -.11	Shin Donga -.09	ZDNet -.07	KD Ilbo -.06
9	Daejon -.11	JIBS -.12	KBC -.11	ZDNet -.10	CJB -.09	JIBS -.06
8	Nongmin -.12	BizWatch -.14	IT Donga -.11	L'Officiel -.10	Economy Chosun -.12	Donga -.07
7	KTV -.13	KTV -.14	Weekly Khan -.12	The Scoop -.11	Donga -.18	KBC -.07
6	JIBS -.13	KD Ilbo -.15	KD Ilbo -.13	CJB -.11	JIBS -.18	Weekly Khan -.08
5	CJB -.16	Weekly Khan -.19	Dailian -.16	Weekly Khan -.13	EBS -.19	CJB -.09
4	Weekly Khan -.18	Dailian -.19	CJB -.16	SisaIN -.18	SisaIN -.20	SisaIN -.13
3	KD Ilbo -.20	SisaIN -.22	SisaIN -.19	Dailian -.19	KD Ilbo -.21	EBS -.14
2	EBS -.22	Jeonju MBC -.26	EBS -.20	EBS -.19	Dailian -.24	Jeonju MBC -.17
1	Dailian -.25	EBS -.27	Jeonju MBC -.22	Jeonju MBC -.23	Jeonju MBC -.25	Dailian -.17

identifying predominant suppliers. These results are illustrated in Figure 4.

A positive net contribution indicates a source role, while a negative score suggests reliance on others' reporting. As in previous analyses, Yonhap News, Newsis, and News1 consistently function

as information sources, though their influence is weaker in culture/entertainment—where overall net contribution is lower. In addition, many economic newspapers rank as sources in politics, economy, and society topics. This appears related to their stable financial structures, business-backed

Table 2. *Estimated Dependence by News Topic*

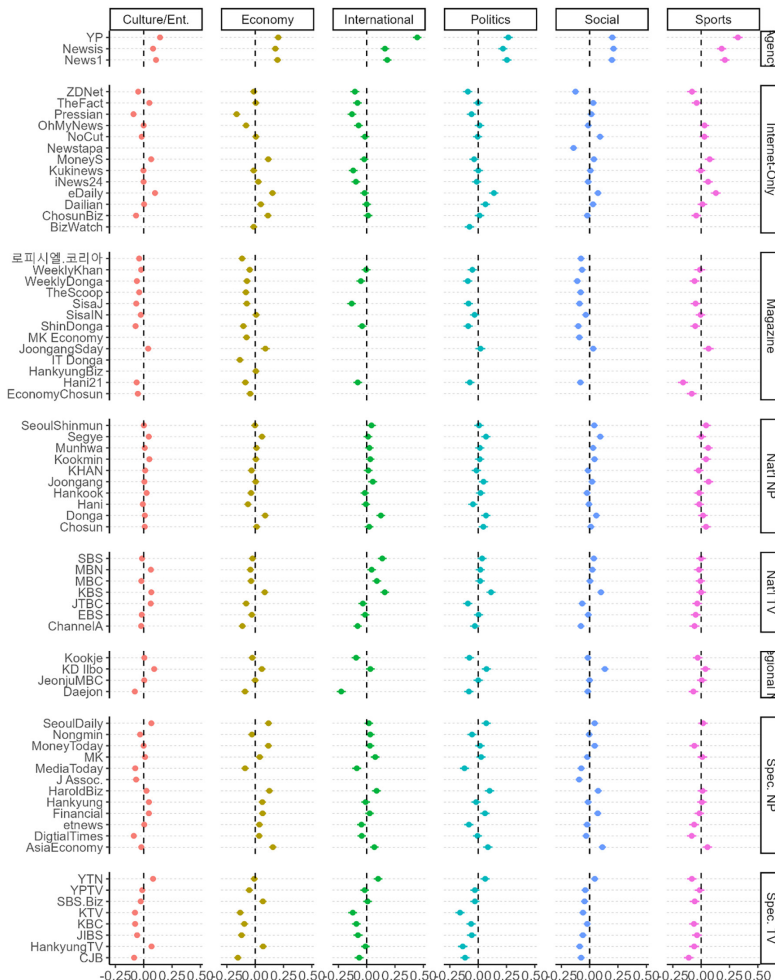
Rank	International	Politics	Economy	Social	Sports	Culture/Ent.
Highest						
1	Daejon .11	SisaJ .13	Pressian .08	SisaJ .09	Hani21 .28	Economy Chosun .07
2	Pressian .11	KHAN .10	KHAN .07	Han kyungTV .07	Digital Times .13	Pressian .06
3	SisaJ .09	Pressian .10	Economy Chosun .06	KHAN .06	Daejon .10	Hani21 .06
4	Kookje .07	Kookmin .09	SisaJ .06	ChosunBiz .06	ChosunBiz .10	NoCut .05
5	Hankook .07	Digital Times .09	Hani .06	Joongang .06	KHAN .08	Media Today .05
6	KHAN .06	Hani .09	Digital Times .05	Hankook .06	Kookje .08	Joongang .05
7	OhMy News .06	Kookje .08	ShinDonga .05	Newstapa .06	Money Today .08	J Assoc. .05
8	Joongang .06	Daejon .08	Joongang .05	Digital Times .05	TheFact .07	Digital Times .04
9	Kukinews .06	Joongang .08	ChannelA .05	Chosun .05	Hani .07	Money Today .04
10	Hani .06	NoCut .07	Hankook .05	Hani .05	Chosun .06	Seoul Shinmun .04
Lowest						
10	MK -.05	YP -.07	Nongmin -.04	KBS -.03	KBS -.06	Joongang Sday -.05
9	JIBS -.05	JIBS -.07	KBS -.04	CJB -.04	Joongang Sday -.09	SeoulDaily -.05
8	CJB -.09	Donga -.13	Weekly Khan -.07	Weekly Khan -.06	YP -.12	Han kyungTV -.06
7	Nongmin -.15	Joongang Sday -.14	Donga -.10	Nongmin -.06	JIBS -.14	Weekly Khan -.06
6	Weekly Khan -.17	Weekly Khan -.14	Joongang Sday -.12	Donga -.07	EBS -.14	Donga -.08
5	Donga -.17	SisaIN -.18	EBS -.17	SisaIN -.14	SisaIN -.20	SisaIN -.11
4	YP -.20	KD Ilbo -.22	KD Ilbo -.19	KD Ilbo -.17	Donga -.20	EBS -.12
3	EBS -.21	Jeonju MBC -.26	SisaIN -.20	EBS -.18	Dailian -.25	KD Ilbo -.15
2	KD Ilbo -.23	Dailian -.26	Dailian -.21	Dailian -.22	KD Ilbo -.25	Dailian -.17
1	Dailian -.25	EBS -.27	Jeonju MBC -.22	Jeonju MBC -.23	Jeonju MBC -.26	Jeonju MBC -.18

($R^2 = 0.625$), outperforming models using dependence or contribution. This suggests that net contribution is a more valid indicator of strategic positioning within the portal-based news ecosystem. Notably, contribution is more strongly associated with organizational characteristics than dependence, reflecting an asymmetry: while reliance on others appears structurally concentrated around a few strong sources,

information provision varies more by outlet-level strategies and capacities.

Age showed no consistent effect across the models, indicating that legacy status alone has limited explanatory power. By contrast, media type is more informative: magazines and regional broadcasters/PPs tend to show higher dependence, while national newspapers and news agencies exhibit greater contribution. Some

Figure 4. *Estimated Net Contribution by Media Type and News Topic*



outlet types, such as specialized media, display both high contribution and high dependence scores, confirming that the two dimensions remain analytically distinct—even when summarized through net contribution. Although

net contribution simplifies analysis, it should be interpreted with caution, as it compresses two underlying dimensions.

The larger the reporting organization, the greater the total number of articles, and the higher the

Table 3. Relationships Between Dependence, Contribution, Net Contribution, and Explanatory Variables

	Contribution	Dependence	Net Contribution
(Intercept)	-0.037** (0.014)	-0.001 (0.015)	-0.036*** (0.010)
# journalists (Topicwise)	0.000** (0.000)	-0.000 (0.000)	0.000*** (0.000)
# articles	0.000 (0.000)	-0.000 (0.000)	0.000* (0.000)
# articles per journalists	0.001*** (0.000)	0.000 (0.000)	0.001*** (0.000)
Age	-0.000 (0.000)	0.000 (0.000)	-0.000 (0.000)
Topic-International	0.010 (0.013)	-0.000 (0.014)	0.010 (0.010)
Topic: Culture/Entertainment	0.009 (0.013)	-0.001 (0.014)	0.010 (0.009)
Topic: Social	-0.015 (0.013)	0.005 (0.014)	-0.020* (0.009)
Topic: Sports	0.019 (0.013)	-0.003 (0.014)	0.022* (0.010)
Topic: Politics	0.010 (0.013)	0.002 (0.014)	0.008 (0.009)
Type: Magazine	-0.029 (0.015)	-0.005 (0.016)	-0.024* (0.011)
Type: National NP	0.054*** (0.016)	0.026 (0.017)	0.028* (0.011)
Type: Specialized NP	0.030* (0.013)	0.021 (0.014)	0.009 (0.010)
Type: National TV	-0.013 (0.014)	-0.020 (0.015)	0.007 (0.010)
Type: Specialized TV	-0.039** (0.015)	0.000 (0.015)	-0.039*** (0.011)
Type: Regional NP	-0.069*** (0.019)	-0.081*** (0.020)	0.012 (0.014)
Type: News Agency	0.062* (0.026)	-0.022 (0.028)	0.083*** (0.019)
R ²	0.461	0.138	0.625
Adj. R ²	0.435	0.096	0.606
Num. obs.	344	344	344

* $p < .05$, ** $p < .01$, *** $p < .001$.

average number of articles per reporter, the higher the net contribution—indicating more original content production. However, these factors had limited impact on dependence, suggesting that reliance is shaped less by individual outlet strategies and more by broader structural dynamics within the news ecosystem.

DISCUSSION

Implications of the Study

This study examined information interrelations to provide a structural analysis in the Korean news ecosystem. By applying a posterior similarity measure that accounts for publication sequence, it revealed strong systemic reliance on news agencies such as Yonhap News, News 1, and Newsis. Even if this dependence arise from contracts, its prevalence—especially among small outlets—suggests that more media does not automatically lead to greater content diversity. At the same time, it would be inappropriate to unconditionally criticize individual media outlets for adopting problematic reporting practices based solely on this observation, since this dependence is more likely to be driven within a broader structure.

To address limitations of previous studies, the study decomposed news similarity into three components: baseline topical similarity, contribution, and dependence. Results confirmed that contribution and dependence are not inversely related but distinct: outlets like national newspapers exhibited both high contribution and high dependence scores, while regional outlets showed low scores on both. This allowed for a nuanced classification of outlets into sources, mediators, dependents, and isolates. A topic-level comparison revealed that more outlets were classified as dependents in international and economic news, while political news saw more mediators. In the culture/entertainment domain, coverage tended to be more independent,

with neither contribution nor dependence strongly pronounced. These findings suggest that media organizations allocate reporting resources differently by topic, shaping content diversity across domains. In hard news areas—especially economy and international relations—high reporting costs and professional demands appear to create structural constraints, limiting diversity by concentrating dependence on a few key sources. In contrast, soft news fields like entertainment and sports showed lower levels of dependence. These results imply that, as profitability declines, only diversity in soft news may grow disproportionately, while diversity in hard news areas remains stagnant.

This study also examined how media outlets' information relate to organizational characteristics. The findings show that net contribution—measured as contribution minus dependence—exhibited strong correlations with these characteristics, indicating high construct validity. Among the two components, contribution was notably more strongly linked to organizational traits than dependence. This suggests that organizational features, rather than directly shaping an outlet's own contribution to diversity, more strongly affect the likelihood that other outlets reference its reporting, thereby influencing the broader structure of information dependence.

Synthesizing these results, it appears that news agencies, economic newspapers, and national dailies, by producing vast volumes of original content and acting as discourse mediators, play central roles in shaping the supply of diverse information. In contrast, small and medium-sized outlets—the “long tail” of the ecosystem—do not primarily supply diversity through distinct specialization strategies; rather, they collectively rely on the leading outlets. A notable exception is the culture/entertainment domain, where neither the contribution of major outlets nor the dependence of smaller ones was strongly pronounced. Thus, from a supply perspective, the current news ecosystem can be understood

as centered on the dominance of major outlets in fields such as economy and international news, while independent contributions to diversity emerge primarily from smaller, internet-based media in culture and entertainment.

These findings suggest that the lack of exposure diversity often seen in online environments may already originate at the supply stage, reinforced by inter-organizational symbiosis among media outlets and the dominant role of a few major news agencies. Promoting content diversity thus requires systemic interventions that incentivize original reporting and support alternative information sources. One possible strategy is to integrate diversity-aware mechanisms into algorithmic recommendation systems, complementing personalization with incentives for novel and distinctive reporting.

Contributions and Limitations

This study contributes to news industry research by emphasizing that content diversity is not solely a function of individual outlets, but of the broader ecosystem shaped by inter-organizational information relations. Moving beyond the traditional focus on niche discovery or technological adaptation, it highlights the role of symbiotic relationships among media outlets as a structural factor influencing content diversity. By empirically analyzing news articles, the study demonstrates that such ecosystemic interrelations is not merely theoretical, but observable in media behavior.

Methodologically, this study advances existing approaches to measuring news content through text similarity in two key ways. First, it generalizes similarity beyond detecting duplicate news, using it to assess the amount of added information in articles relative to prior reports. Second, it incorporates publication timing, allowing similarity to reflect the directional flow of information between media outlets—capturing both contribution and dependence.

This approach preserves document-level textual richness while integrating time-series insights, offering a potential complement or alternative to Vector Autoregressive (VAR) models commonly used in intermedia agenda-setting research.

Additionally, the study introduces a framework for decomposing posterior similarity scores. Recognizing that similarity measures relational overlap rather than frequency, the study builds on the HITS algorithm's distinction between inflow and outflow, extending it through a gravity model to account for topical similarity. This yields novel measures of contribution, dependence, and net contribution. Potential applications include evaluating media influence or designing platform curation systems that reward outlets based on original information interrelations. Beyond these platform-level implications, this framework also offers a policy-relevant approach for assessing inter-organizational imitation within the news ecosystem. In particular, the contribution and dependence measures—derived from Bayesian modeling of content-copying networks—could serve as quantitative weights in regulatory initiatives such as media concentration assessments, enabling a fairer evaluation of outlets' *network power* that goes beyond simple counts of published or viewed articles.

Despite these contributions, the study has several limitations concerning data coverage, temporal scope, and the exclusion of audience impact. The dataset was limited to news articles from Daum, excluding other major platforms such as Naver or You Tube, and thus may not fully represent the Korean news ecosystem. As Daum differs from Naver in editorial algorithms and partnership structures, future studies could apply the same framework across multiple platforms to compare how platform governance and visibility hierarchies shape informational interrelations. Small or independent outlets not supplying to Daum were also underrepresented, restricting insights into their potential contributions to diversity. In addition, because the dataset covers

only December 2022, seasonal biases may affect topic-specific reporting patterns (e.g., political coverage slowdown, year-end entertainment peaks). While such effects cannot be fully controlled, the large-scale nature of the data and topic-level decomposition help mitigate this limitation. Finally, the study focused on supply diversity without addressing how such contributions affect audiences or public discourse.

The policy implications proposed here remain exploratory. Developing actionable strategies would require collaboration with platform operators and regulators. Future research should expand the analysis temporally and cross-platform, combine quantitative modeling with qualitative inquiry into newsroom practices, and extend comparative studies to international contexts. Such efforts would further test the generalizability of the contribution-dependence framework and clarify how institutional environments mediate information flows within platform-driven news ecosystems.

In conclusion, this study makes both academic and policy contributions by empirically analyzing the structural constraints shaping Korea's news ecosystem and proposing practical measures to promote information diversity. Future research and policy efforts building on these findings can further support the development of a healthier, more sustainable news environment.

REFERENCES

- AD-PR & Media (2023, May 15). 981 Naver news partner outlets in the first half of 2023. *AD-PR & Media*.
<https://www.apnews.kr/news/articleView.html?idxno=3010007>
- Ahn, S. (2021). *The new past and aged future of journalism: The reproductive report and the original report* [Doctoral Dissertation, Korea University].
<http://www.riss.kr/link?id=T15784456>
- An, J., & Kwak, H. (2017). *Multidimensional analysis of the news consumption of different demographic groups on a nationwide scale* [Paper presentation]. International Conference on Social Informatics (pp. 124–142).
- Assmann, K. (2023). Rise of the zombie papers: Infecting Germany's local and regional public media ecosystem. *Media and Communication*, 11(3), 360–370.
- Bell, E. J., Owen, T., Brown, P. D., Hauka, C., & Rashidian, N. (2017). *The platform press: How Silicon Valley reengineered journalism*. Tow Center for Digital Journalism, Columbia University.
- Boczkowski, P. J. (2009). Rethinking hard and soft news production: From common ground to divergent paths. *Journal of Communication*, 59(1), 98–116.
- Boczkowski, P. J., & De Santos, M. (2007). When more media equals less news: Patterns of content homogenization in Argentina's leading print and online newspapers. *Political Communication*, 24(2), 167–180.
- Boumans, J., Trilling, D., Vliegthart, R., & Boomgaarden, H. (2018). The agency makes the (online) news world go round: The impact of news agency content on print and online news. *International Journal of Communication*, 12(2018), 1768–1789.
- Davies, N. (2008). *Flat earth news*. Chatto & Windus.
- Dimmick, J. (2003). *Media competition and coexistence: The theory of the niche*. Lawrence Erlbaum Associates.
- Dimmick, J., & Rothenbuhler, E. (1984). The theory of the niche: Quantifying competition among media industries. *Journal of Communication*, 34(1), 103–119.
- Dwyer, T. (2010). *Media convergence*. McGraw-Hill Education (UK).
- Dwyer, T., & Martin, F. (2017). Sharing news online: Social media news analytics and their implications for media pluralism policies. *Digital Journalism*, 5(8), 1080–1100.

- Harcup, T. (2004). *Journalism principles and practice*. London: Sage.
- Jackson, D., & Moloney, K. (2015). Inside churnalism: PR, journalism and power relationships in flux. *Journalism Studies*, 17(6), 763–780.
- Ji, M. (2023, November 24). Daum limits search results to articles from 150 partner news organizations. *Dong-A Ilbo*.
<https://www.donga.com/news/Society/article/all/20231124/122328912/1>
- Johnston, J., & Forde, S. (2017). Churnalism: Revised and revisited. *Digital Journalism*, 5(8), 943–946.
- Jung, D., & Kim, D.-G. (2016). A study on journalism scholar's recognition and evaluation to issues and reform measures about portal sites' news service. *Journal of Communication Science*, 16(4), 254–286.
- Jung, J., & Park, B. (2012). Business news competition for user gratification: A niche analysis of network TV, newspaper, Internet, and SNS. *Korean Journal of Broadcasting and Telecommunication Studies*, 26(2), 205–246.
- Kleinberg, J. (1999). Authoritative sources in a hyperlinked environment. *Journal of the ACM*, 46(5), 604–632.
- Klopčič, A. L., Hojnik, J., Bojnec, Š., & Papler, D. (2020). Global transition to the subscription economy: Literature review on business model changes in the media landscape. *Managing Global Transitions*, 18(4), 323–348.
- Korea Press Foundation. (2023). *Media users in Korea 2023*. Korea Press Foundation.
- Korea Press Foundation Industry Analysis Team. (2022). *The Korea press yearbook 2022*. Korea Press Foundation.
- Lee, J. H. (2021). An exploratory study on news homogenization using document similarity measurement through Doc2Vec. *Journal of Communication Research*, 58(4), 5–48.
- Lewis, J., Williams, A., & Franklin, B. (2008). Four rumours and an explanation: A political economic account of journalists' changing newsgathering and reporting practices. *Journalism Practice*, 2(1), 27–45.
- Lewis, N. P. (2008). Plagiarism antecedents and situational influences. *Journalism & Mass Communication Quarterly*, 85(2), 353–370.
- Lewis, N. P. (2013). Idea plagiarism: Journalism's ultimate heist. *Mass Communication and Society*, 16(5), 738–757.
- Lowrey, W. (2012). Journalism innovation and the ecology of news production: Institutional tendencies. *Journalism & Communication Monographs*, 14(4), 214–287.
- Maier, S. (2010). All the news fit to post? Comparing news content on the web to newspapers, television, and radio. *Journalism & Mass Communication Quarterly*, 87(3-4), 548–562.
- McQuail, D. (1992). *Media performance: Mass communication and the public interest*. Sage.
- Napoli, P. M. (1999). Deconstructing the diversity principle. *Journal of Communication*, 49(4), 7–34.
- Nielsen, R. K., & Fletcher, R. (2022). Concentration of online news traffic and publishers' reliance on platform referrals: Evidence from passive tracking data in the UK. *Journal of Quantitative Description: Digital Media*, 2.
<https://doi.org/10.51685/jqd.2022.015>
- Pak, C. (2019). News organizations' selective link sharing as gatekeeping: A structural topic model approach. *Computational Communication Research*, 1(1), 45–78.
- Pak, C., Cotter, K., & DeCook, J. (2020). Intermedia reliance and sustainability of emergent media: A large-scale analysis of American news outlets' external linking behaviors. *International Journal of Communication*, 14, 23.
- Park, J., Huh, M., & Ahn, S. (2020). *A study on the press beat system and reporting practices*. Korea Press Foundation.
- Phillips, A. (2011). Journalists as unwilling 'sources': Transparency and the new ethics of journalism. In B. Franklin & M. Carlson

- (Eds.), *Journalists, sources and credibility: New perspective* (pp. 49–60). Routledge.
- Picard, R. G. (2011). *The economics and financing of media companies*. Fordham Univ Press.
- Picard, R. G. (2014). Twilight or new dawn of journalism? Evidence from the changing news ecosystem. *Journalism Practice*, 8(5), 488–498.
- Reid, J. (2025). Digitising “The big lie”: Algorithmic curation as an inhibitor of media exposure diversity online. *Communicatio*, 1–21.
- Shin, J. H., & Youn, S. (2025). Theoretical framework for analyzing news homogenization: Proposing the concept of ‘Original Article’ and indicators of ‘Influence’ and ‘Productivity’. *Korean Journal of Journalism and Communication*, 69(3), 5–40.
- Song, Y.-H., & Lee, S.-S. (2007). A study on the influences of participation government’s press policy on practices in beats. *Korean Journal of Communication & Information*, 40, 114–149.
- Stan Development Team. (2024) Stan modeling language users guide and reference manual. Version 2.35.
<https://mc-stan.org>
- Stempel, G. H. (1985). Gatekeeping: The mix of topics and the selection of stories. *Journalism Quarterly*, 62(4), 791–815.
- Vogler, D., Udris, L., & Eisenegger, M. (2020). Measuring media content concentration at a large scale using automated text comparisons. *Journalism Studies*, 21(11), 1459–1478.
- Weber, M. (2012). Newspapers and the long-term implications of hyperlinking. *Journal of Computer-Mediated Communication*, 17(2), 187–201.
- Weber, M., & Monge, P. (2011). The flow of digital news in a network of sources, authorities, and hubs. *Journal of Communication*, 61(6), 1062–1081.