Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Framing Community Happiness in Korean National and Local Newspapers: A Longitudinal Text-Mining Analysis, 2016–2025

Ji-Hyun Jang

jhjang@smu.ac.kr

Sangmyung University

Research Article

Keywords: community well-being, media discourse, local journalism, COVID-19, text mining, South Korea

Posted Date: September 17th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10759394/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Abstract

This study examines how Korean national and local newspapers framed community happiness before, during, and after the COVID-19 crisis. Articles were retrieved directly from the BIGKinds news database by searching titles and full texts for the Korean equivalents of “community” AND “happiness” between January 1, 2016 and December 31, 2025. The portal returned 65,648 records; 63,214 were exported in six period-by-media Excel files, and 62,738 remained after removing 476 records flagged as duplicates or exceptions. The analytic corpus comprised 7,337 articles from 12 national dailies and 55,401 from up to 45 local dailies. Document frequency, TF-IDF, keyword co-occurrence networks, centrality and spectral clustering, and lexicon-based sentiment analysis were applied. National newspapers disproportionately emphasized expansion, children, the environment, facilities, the economy, and institutional support. Local newspapers more often linked happiness to gratitude, neighbors, care, older people, volunteering, and place-based relations. During 2020–2022, COVID-19 became the dominant network hub and degree centralization increased to .820 in national and .669 in local coverage. Mean sentiment fell sharply before recovering after 2023; local coverage remained more positive, although between-media effects were small. A common-window sensitivity analysis produced nearly identical keyword-gap rankings. The findings show that community well-being is publicly constructed through two persistent but complementary frames: an institutional-national frame and a relational-local frame. They also demonstrate crisis-induced agenda compression followed by post-crisis diversification.

1 Introduction

Community well-being is not reducible to the average happiness of individuals living within a geographic boundary. It concerns the shared material conditions, public services, social relationships, opportunities for participation, and place-based attachments through which people experience a locality as safe, capable, inclusive, and worth inhabiting. Subjective evaluations of life matter, but they are shaped by collective conditions such as health care, care work, education, housing, employment, the environment, trust, and the capacity to act together (Diener, 1984; Sirgy et al., 2000, 2010; Wiseman & Brasher, 2008). This multidimensionality has encouraged community well-being scholarship to examine both measurable conditions and the meanings through which communities understand their circumstances.

Public discourse is one of the mechanisms through which those meanings are produced. News organizations repeatedly select some community conditions for attention, connect problems to particular causes and actors, and make certain remedies appear reasonable. Agenda-setting research explains how persistent coverage influences the perceived importance of issues, while framing research shows how news defines problems, assigns responsibility, and organizes solutions (Entman, 1993; McCombs & Shaw, 1972). A story that connects community happiness to corporate donations, for example, implies a different distribution of agency from a story centered on resident participation, universal public services, or neighborhood solidarity. News coverage does not simply mirror community well-being; it supplies a vocabulary through which governments, organizations, and residents may recognize and contest it.

The geographical scale of news production is therefore theoretically relevant. National newspapers tend to organize local matters through nationwide policy, large institutions, corporations, expert sources, and generalizable indicators. Local newspapers are institutionally closer to municipal governments, neighborhood organizations, civic groups, and recurring local events. They may therefore document everyday relations and named places with greater granularity. Yet local journalism may also depend heavily on official press releases and ceremonial event coverage. The contrast should not be interpreted as a simple hierarchy of journalistic quality. The empirical question is which dimensions, actors, and emotional tones each media type makes visible when it represents community happiness.

The decade from 2016 through 2025 provides a consequential period for observing such representation. COVID-19 simultaneously affected health, social isolation, education, care, employment, and local economies. It also elevated vulnerable groups and emergency support within public discourse. In the subsequent transition, concerns over aging, population decline, care infrastructures, social recovery, and local vitality became increasingly salient. Community well-being research has emphasized that resilience depends not only on formal emergency capacity but also on social capital, trust, and networks able to reach vulnerable residents (Aldrich & Meyer, 2015; Bagnall et al., 2017). Whether national and local news compressed these multiple domains into the same crisis frame, and whether earlier relational and participatory themes re-emerged afterward, are important questions for understanding how collective well-being is publicly imagined.

Existing research in South Korea has analyzed community satisfaction, local government performance, environmental perception, social capital, and quality of life (Cho & Lee, 2021; Lee & Kim, 2018; Lee & Park, 2022; Woo & Jung, 2022). Computational studies have also demonstrated the value of topic modeling and network analysis for examining Korean news discourse in policy fields such as health inequality and nursing (Kim, 2019; Lee et al., 2024; Park et al., 2024). However, large-scale longitudinal comparisons of national and local media discourse specifically linking community and happiness remain limited. This matters because a corpus drawn only from national newspapers cannot show whether local journalism frames the same idea through different actors, relationships, and emotional registers.

This study addresses that gap with 62,738 articles collected directly from the BIGKinds Korean news database. It compares national and local daily newspapers across a pre-pandemic period (2016–2019), a crisis-response period (2020–2022), and a post-crisis transition period (2023–2025). The analysis combines document frequency, TF-IDF, co-occurrence networks, centrality and clustering, and lexicon-based sentiment. Rather than treating raw article counts as comparable across unequal corpora, it relies on within-group prevalence, normalized network measures, and effect sizes. It also explicitly evaluates a platform export limit that affected the local-newspaper corpus in Period 3.

The study asks four questions. RQ1: How did the volume and composition of community-happiness coverage vary over time and between national and local newspapers? RQ2: Which keywords and distinctive terms characterized each period and media type? RQ3: How did keyword-network centrality, concentration, and clustering change across periods and media types? RQ4: How did the sentiment of

coverage vary across periods and between national and local newspapers? These questions support a broader interpretation of how institutional-national and relational-local frames coexist, how crisis conditions compress a multi-domain well-being agenda, and how media discourse may inform community-oriented policy communication.

2 Conceptual Background

2.1 Community Well-Being as Material, Relational, and Participatory

Subjective well-being conventionally includes cognitive judgments of life satisfaction and the balance of positive and negative affect (Andrews & Withey, 1976; Diener, 1984). Community well-being changes the unit of attention from the isolated individual to the shared contexts within which those judgments become possible. Sirgy et al. (2000, 2010) conceptualized community well-being through residents' assessments of services and the perceived effects of community conditions across life domains. Wiseman and Brasher (2008) integrated social, economic, environmental, and governance conditions with community capacity. Theodori (2001) showed that community satisfaction and attachment have consequences for individual well-being. Together, these approaches indicate that community well-being combines objective opportunities, subjective evaluations, collective relationships, and institutional arrangements, while remaining consistent with broader guidance on multidimensional well-being measurement (OECD, 2025).

This synthesis can be organized into three interdependent domains. The material domain includes employment, income, housing, transport, safety, health, education, culture, and environmental quality. The relational domain concerns trust, neighbors, belonging, reciprocity, care, and social support. The participatory domain concerns voice, collective efficacy, local governance, and the ability of residents to shape decisions affecting their lives. Social capital links the latter two domains by facilitating information, cooperation, and mutual aid (Putnam, 2000). A locality can therefore possess economic assets but remain low in community well-being if services are inaccessible, relationships are exclusionary, or residents lack meaningful agency.

The South Korean literature similarly cautions against equating economic performance with community well-being. Lee and Kim (2018) found that economic conditions alone do not purchase community well-being, while local environmental perceptions, institutional satisfaction, and social factors make independent contributions. Lee and Park (2022) connected perceptions of the community environment and satisfaction with local government to quality of life. Cho and Lee (2021) showed the relevance of local government to community satisfaction among younger rural residents. These studies support a view of happiness as locally embedded and institutionally mediated.

Media analysis adds a distinct layer to this literature. It does not measure whether residents are actually satisfied, healthy, or socially connected. Instead, it identifies the public representations through which

community well-being is made visible. This distinction is essential because a rise in positive stories may reflect promotional activity, ceremonial language, or editorial routines rather than an improvement in residents' lived conditions. News discourse should therefore be interpreted as a communicative indicator that can be compared with, but not substituted for, surveys and objective community indicators. Passey et al. (2024) likewise caution that community well-being research must be attentive to the gap between its rhetoric and the realities captured by its methods.

2.2 Media Framing, Journalistic Scale, and Community Agency

Framing is especially useful for analyzing community well-being because the concept is broad enough to support different accounts of what matters and who should act. When support, facilities, corporations, and expansion dominate, happiness may be framed as the outcome of institutional provision. When participation, residents, neighbors, and cooperation dominate, it may be framed as a co-produced social accomplishment. When illness, isolation, and vulnerable groups dominate, it becomes a problem of protection and resilience. These frames may overlap, but their relative prominence changes the implied theory of community agency.

National and local journalism operate at different scales of source access, audience expectation, and geographic specificity. National outlets can connect local concerns to central-government programs, corporate social responsibility, macroeconomic conditions, and nationwide comparisons. Such coverage may reveal structural resources and policy diffusion, but it can abstract communities into examples of a national issue. Local outlets repeatedly encounter municipal departments, schools, welfare centers, civic organizations, volunteers, and residents. Their proximity can preserve the texture of everyday social relations and place-based vulnerability. It can also normalize event-centered stories that celebrate donations, gratitude, and harmony without investigating service gaps or unequal power.

The distinction is therefore best understood as a difference in dominant scale rather than a mutually exclusive binary. National outlets may publish intimate local stories, while local outlets may scrutinize national policy. Nevertheless, aggregate vocabularies can reveal systematic tendencies. A media system that combines national structural visibility with local relational visibility may offer a fuller representation of community well-being than either scale alone. The policy relevance lies in identifying where the frames complement one another and where they leave blind spots—for example, when national provision overlooks lived accessibility, or when local celebration overlooks structural inequality.

2.3 Crisis-Induced Agenda Compression and Post-Crisis Diversification

Crises reorganize public attention. They can elevate one event label into a master frame that connects previously separate domains and narrows the range of legitimate priorities. During COVID-19, health protection, welfare delivery, school disruption, isolation, employment, and community volunteering were all interpreted through the pandemic. Drew and Kim (2025) show that community well-being and public

management theory can illuminate variations in pandemic response, while Narayan (2020) emphasizes the growing role of information infrastructures in community-level well-being. From a discourse perspective, the critical question is whether crisis attention merely adds a new topic or structurally centralizes the network around a single hub.

Such compression may be necessary for coordinated response, but it carries costs. Education, culture, participation, mental health, and non-pandemic care can be relegated to the margins when emergency protection becomes the overarching criterion. Conversely, a crisis can make normally invisible interdependencies legible: the connection between public health and care work, between local business and social stability, or between neighbors and the reach of formal services. Post-crisis diversification therefore does not imply a return to the pre-crisis agenda. It may produce a reconfigured field in which health, welfare, care, and resilience remain more tightly connected than before.

Network centralization offers a direct way to evaluate this process. If crisis-induced compression occurred, a pandemic-related node should connect to many otherwise distinct themes and whole-network degree centralization should rise. If the agenda later diversified, centralization should fall and multiple hubs should reappear. Sentiment should also decline during the crisis and recover afterward, although the depth and speed of change may differ between national and local coverage because of their source structures and narrative conventions.

2.4 Computational Analysis of News Discourse

Document frequency, TF-IDF, co-occurrence networks, and sentiment analysis capture complementary properties of a corpus. Document frequency measures how widely a term is distributed across articles and is less sensitive than raw word counts to repetition within press-release-derived stories. TF-IDF emphasizes terms that are frequent within a document or group but less ubiquitous across the corpus, thereby distinguishing shared high-frequency vocabulary from temporally or institutionally distinctive language (Salton & Buckley, 1988).

Co-occurrence networks represent terms as nodes and their joint appearance within articles as edges. Degree centrality identifies terms directly connected to many others, weighted strength reflects the accumulated intensity of those connections, and betweenness centrality identifies terms bridging otherwise separated clusters (Freeman, 1978; Wasserman & Faust, 1994). Whole-network centralization evaluates the extent to which the network is organized around a small number of dominant hubs. Spectral clustering can then identify recurrent subdiscourses within an association matrix (Ng et al., 2002).

Lexicon-based sentiment analysis estimates the positive and negative orientation of language from predefined terms (Pang & Lee, 2008; Taboada et al., 2011). It is transparent and reproducible but less capable of resolving irony, quotation, negation scope, and context-dependent meanings than supervised language models. Because the corpus was selected with the term “happiness,” uncorrected sentiment would be mechanically positive. The present design therefore excludes the search terms from sentiment

scoring, uses an issue-adapted lexicon, and interprets mean differences together with category shares, annual trends, and standardized effect sizes.

Table 1
Research questions and corresponding analytical strategies

Research question	Empirical focus	Primary measures
RQ1	Change in corpus volume and composition by period and media type	Article counts; outlet and category distributions
RQ2	Common, distinctive, and period-specific vocabulary	Document prevalence; prevalence gaps; TF-IDF
RQ3	Change in the relational organization of issues	Co-occurrence; centrality; centralization; spectral clusters
RQ4	Change in emotional tone by period and media type	Sentiment scores; category shares; Welch tests; Cohen's d

3 Methods

3.1 Design and Periodization

The research design crosses time with media scale. Period 1 (P1, 2016–2019) represents the pre-pandemic baseline. Period 2 (P2, 2020–2022) represents crisis and organized response. Period 3 (P3, 2023–2025) represents a post-crisis transition rather than a claim that the social consequences of COVID-19 had ended. For each period, national and local daily newspapers were analyzed separately. The sequence of analysis proceeded from corpus composition to keyword prevalence and TF-IDF, then to co-occurrence structure and sentiment. Findings were integrated only when multiple measures supported a common interpretation.

The article was the basic unit of comparison. The same keyword was counted once per article for document-frequency analyses regardless of within-article repetition. This decision reduces the influence of formulaic repetition in promotional and press-release-based stories. Because the national and local corpora were highly unequal in size, raw frequencies were not treated as direct evidence of relative issue importance. The primary comparative measures were within-group prevalence, normalized association strength, standardized sentiment, and network centrality.

3.2 Data Source, Search Strategy, and Corpus Construction

The data were collected directly by the researcher from BIGKinds, the news big-data service operated by the Korea Press Foundation (Korea Press Foundation, 2026). Searches covered article titles and full texts published from January 1, 2016 through December 31, 2025. The Boolean query joined the exact Korean terms transliterated as “jiyeok sahoe” (community) AND “haengbok” (happiness). The deliberately concise query was selected for conceptual transparency and replicability. The portal was

queried separately by period and media type, producing six Excel exports: three for 12 national daily newspapers and three for local daily newspapers, whose coverage expanded from 41 outlets in P1 to 43 in P2 and 45 in P3.

The portal reported 65,648 search results. The six exported files contained 63,214 rows and 19 fields, including article identifier, date, outlet, title, category, persons, organizations, locations, keywords, extracted features, full text, URL, and BIGKinds' analysis-exclusion flag. Records were excluded only when BIGKinds marked them as duplicates or exceptions. Specifically, 222 were marked as duplicates, 252 as exceptions, and two as both, for 476 exclusions. No missing article identifiers, titles, full texts, or keyword fields remained in the final data, and no duplicate article identifiers remained. The final analytic corpus contained 62,738 articles: 7,337 from national dailies and 55,401 from local dailies.

The difference between portal hits and exported rows requires explicit explanation. For local newspapers in P3, BIGKinds reported 22,434 results, but its one-time Excel export limit was 20,000 records. Consequently, 2,434 search results were not included in the downloaded file. These were not researcher-defined exclusions and their content could not be inspected. The saved file covered May 17, 2023 through December 31, 2025; after removing 49 duplicates and 96 exceptions, it yielded 19,855 analyzable articles. This truncation affects completeness at the beginning of P3 and is addressed through a common-window sensitivity analysis.

Table 2

Corpus construction by period and media type

Period	Media	Portal hits	Exported	Export-limit loss	Flagged exclusions	Analyzed
P1 2016–2019	National	2,493	2,493	0	21	2,472
P1 2016–2019	Local	18,964	18,964	0	178	18,786
P2 2020–2022	National	2,115	2,115	0	22	2,093
P2 2020–2022	Local	16,858	16,858	0	98	16,760
P3 2023–2025	National	2,784	2,784	0	12	2,772
P3 2023–2025	Local	22,434	20,000	2,434	145	19,855
Total	All	65,648	63,214	2,434	476	62,738

Note. The 2,434 local P3 records absent from the Excel file reflect the BIGKinds one-time export ceiling, not researcher exclusion. Flagged exclusions comprise duplicates and/or exceptions identified by BIGKinds.

3.3 Preprocessing

The six files were harmonized to a common schema and assigned period and media-type indicators from their filenames. Dates were converted from YYYYMMDD strings to date objects. Only rows with an empty analysis-exclusion field were retained. The keyword field was tokenized on commas; whitespace and minor orthographic variants were normalized. Document-frequency and TF-IDF analyses used the BIGKinds keyword field, while network analysis used its ranked top-50 feature-extraction field because that field provides a more selective representation of article content.

Procedural terms such as “photo provided,” “online news team,” and “reporting team,” along with outlet names and named entities lacking conceptual relevance, were removed. Terms central to the research construct—including support, participation, education, culture, health, environment, neighbors, residents, children, welfare, care/love, and gratitude—were protected even when they overlapped with named-entity lists. Terms were deduplicated within articles for document-frequency calculations. No article-level sampling or automated summarization was used; all 62,738 retained records entered the analysis.

3.4 Document Frequency and TF-IDF

For term t, document frequency DF(t) was the number of articles in which t appeared at least once. Prevalence was DF(t) divided by the number of articles in the corresponding period-media group. The national-local prevalence gap was calculated in percentage points. A log2 prevalence ratio was also examined to distinguish high-frequency shared terms from terms disproportionately associated with one media type.

TF-IDF was estimated for terms appearing in at least 20 articles and no more than 75% of all articles. Sublinear term frequency and L2 normalization were applied. Article-level scores were averaged within each period-media group and multiplied by 1,000 for reporting. The method identifies vocabulary that is relatively informative for a particular group even when it is not among the largest corpus-wide frequencies.

3.5 Co-Occurrence Networks, Centrality, and Clustering

For each of the six period-media groups, the 36 highest-frequency eligible feature terms were used as nodes. Two terms were connected when they appeared in the same article. Edge weight was association strength, calculated as the joint-document count divided by the geometric mean of the two terms' document frequencies. To prevent the much larger local corpus from receiving connections merely because of sample size, the minimum joint occurrence was the larger of eight articles or 0.1% of the group's article count. Up to 160 highest-strength edges were retained, while the strongest edge of an otherwise isolated node was preserved.

Degree centrality, weighted strength, betweenness centrality, and eigenvector centrality were computed for each node. Network-level statistics included density, mean degree, and Freeman degree centralization. Centralization was the main indicator of crisis-induced agenda compression. Spectral

clustering was applied to the association matrix with four clusters per network. Cluster labels were assigned interpretively from the terms with the highest within-cluster weighted strength; the labels are analytical summaries, not model-generated ground truth.

3.6 Sentiment Analysis

An issue-adapted Korean lexicon included 70 positive and 68 negative terms related to policy, community relations, safety, health, decline, exclusion, and crisis. Repeated appearances of the same sentiment term within an article were counted once. Severe harm terms such as death, suicide, and disaster received a weight of two. The article score was (positive terms − weighted negative terms)/(positive terms + weighted negative terms). Scores of .15 or higher were classified as positive, scores of − .15 or lower as negative, and intermediate scores as neutral. The query terms “happiness” and “community” were excluded to reduce selection-induced positivity.

Media differences in mean article sentiment were evaluated with Welch's t test, which does not assume equal variance, and Cohen's d. Because very large samples can make substantively small differences statistically significant, effect sizes, annual trajectories, and the proportions of positive, neutral, and negative articles were emphasized over p-values alone.

3.7 Validity, Sensitivity, and Interpretive Boundaries

Validity was strengthened in four ways. First, portal hits, exported rows, platform-flagged exclusions, and analyzed records were reported separately. Second, document prevalence, TF-IDF, network structure, and sentiment were triangulated rather than allowing one technique to determine the interpretation. Third, normalized rates and effect sizes were used because the corpora differed substantially in size. Fourth, news sentiment was explicitly distinguished from residents' well-being.

The P3 export ceiling was not assumed to produce random missingness. Because the local file began on May 17, 2023, the principal sensitivity analysis restricted the national P3 corpus to the same starting date and re-estimated national-local keyword gaps and sentiment. Similarity was assessed with Spearman rank correlation and top-gap overlap. The exercise tests directional robustness under a shared observation window; it cannot recover or characterize the unexported 2,434 local records.

An OpenAI language model was used to translate the Korean manuscript into English and to support linguistic and submission-format editing. The empirical analyses, numerical results, interpretations, and responsibility for the final submitted text remain with the author. In accordance with journal guidance, the tool is not listed as an author.

4 Results

4.1 Corpus Profile and Temporal Distribution

The analyzed corpus comprised 62,738 articles, of which 7,337 (11.69%) were from national and 55,401 (88.31%) from local dailies. P1 contained 2,472 national and 18,786 local articles; P2 contained 2,093 and 16,760; and P3 contained 2,772 and 19,855. The local corpus was therefore larger in every period, making normalized measures essential. Twelve national outlets were represented throughout. Local coverage expanded from 41 outlets in P1 to 43 in P2 and 45 in P3.

National coverage increased from 471 articles in 2016 to 855 in 2019, declined to 717 in 2020 and 672 in 2021, and then rose to 1,047 in 2025. Local coverage rose from 3,425 articles in 2016 to 6,028 in 2019, remained near 5,600 annually during P2, and reached 7,116 in 2024 and 8,540 in 2025. The low local count in 2023 should not be read as substantive decline because the exported P3 file begins in mid-May. Figure 1 displays the annual series in separate panels because the scales differ substantially.

Outlet contributions were unequal. Asia Today supplied 2,067 national articles (28.17%), followed by Kukmin Ilbo (923) and Seoul Shinmun (765). Joongdo Ilbo contributed 5,907 local articles (10.66%), followed by several Chungcheong-region newspapers. Results therefore describe the discourse of the collected article populations, not an equally weighted “average outlet.” The automated top-level category distribution also differed: “regional” articles constituted 69.90% of local but 46.57% of national coverage. National coverage was more dispersed across social (22.46%), cultural (11.03%), and economic (9.95%) categories than local coverage (16.02%, 6.05%, and 3.27%, respectively). This distribution is consistent with a more place-specific local corpus but also reflects BIGKinds' automated classification and outlet editorial practices.

4.2 Keyword Prevalence and Media Differences

Across all articles, support was the most widely distributed theme, appearing in 52.47% of documents. It was followed by participation (39.94%), welfare (33.72%), education (30.42%), culture (29.21%), health (27.06%), neighbors (26.95%), children (26.25%), residents (25.77%), and development (25.40%). Community happiness was thus represented less as a free-standing emotion than as a combination of institutional resources, participatory action, service domains, social relations, and local development.

Figure 2 shows the largest national-local prevalence gaps. National newspapers used expansion 9.35 percentage points more often than local papers, children 9.32 points more often, support 7.95, environment 7.28, facilities 7.02, economy 6.77, improvement 6.48, growth 5.02, education 4.06, family 4.05, and participation 3.99. These differences cluster around institutional provision, program scale, infrastructure, target populations, and development outcomes. Local newspapers used gratitude 7.00 points more often than national papers, neighbors 4.34, care/love 2.63, older people 2.11, volunteering 2.07, village 2.07, and empathy/mind 1.67. These terms point toward interpersonal recognition, everyday care, civic contribution, aging, and place-based belonging.

Table 3

Most prevalent keywords and largest national-local gaps

Overall rank	Keyword	All articles (%)	National minus local (pp)	Interpretive emphasis
1	Support	52.47	+ 7.95	Institutional and material provision
2	Participation	39.94	+ 3.99	Collective action and governance
3	Welfare	33.72	−0.37	Basic services and protection
4	Education	30.42	+ 4.06	Capacity, schools, and children
5	Culture	29.21	+ 2.18	Shared experience and leisure
6	Health	27.06	+ 1.41	Health and safety
7	Neighbors	26.95	−4.34	Relational and place-based ties
8	Children	26.25	+ 9.32	Policy target and future generation
—	Expansion	—	+ 9.35	Scale and program growth
—	Gratitude	—	−7.00	Recognition and ceremonial relation

Note. Prevalence is the percentage of articles containing a term at least once. A positive gap indicates higher national prevalence; values are rounded to two decimal places.

The period-media heatmap in Fig. 3 clarifies change within these persistent frames. Support remained high in national coverage (57.69% in P1, 57.76% in P2, and 62.41% in P3) and increased in local coverage (47.40%, 52.06%, and 55.02%). Education was most prevalent in P1 and declined during the crisis in both media types. Health increased markedly in P3, reaching 31.28% of national and 31.50% of local articles. Welfare also rose in P3 to 34.96% and 37.10%, respectively. The post-crisis transition therefore did not erase support; it reconnected support with welfare, health, and cooperation.

Relational language changed more strongly in the local corpus. Gratitude increased from 16.90% in local P1 to 25.67% in P2 and 29.36% in P3. During P2, neighbors appeared in 29.77% of local articles compared with 22.17% of national articles. This pattern suggests that crisis response was translated through different narrative grammars: national reporting more often emphasized institutional and corporate provision, while local reporting more often situated response within neighbors, welfare, love/care, older residents, and volunteering.

4.3 TF-IDF and Distinctive Vocabulary

TF-IDF distinguished vocabulary that was especially informative within each period-media group (Table 4; Fig. 4). In P1 national coverage, the leading terms were children (29.7), support (28.9), education (26.1), participation (22.0), and corporations (20.8). Local P1 coverage was led by education (28.1), support (27.6), neighbors (24.4), older people (24.3), participation (23.0), care/love (22.8), welfare (22.3), and volunteering (22.3). The national pattern foregrounded children, corporations, and institutional contribution, whereas the local pattern connected education and support to neighbors, aging, care, and civic activity.

In P2, COVID-19 ranked first in both corpora, with a mean TF-IDF of 39.99 in national and 45.56 in local coverage. National coverage then emphasized support (31.90), children (27.48), corporations (20.71), and education (20.69). Local coverage emphasized support (31.32), neighbors (30.65), welfare (24.70), education (23.26), and care/love (22.88). The same crisis label organized both discourses, but its nearest substantive companions differed by media scale.

In P3, the national leaders were support (36.47), children (30.02), participation (23.60), welfare (22.81), education (21.15), older people (20.78), and health (20.23). The local leaders were support (33.46), welfare (28.34), older people (27.97), neighbors (27.33), children (26.59), participation (25.21), and health (23.27). With the pandemic label no longer dominant, both corpora diversified. National coverage returned more strongly to participation and institutional domains, while local coverage retained a pronounced care-and-aging orientation.

Table 4

Leading mean TF-IDF terms by period and media type

Period	National dailies	Local dailies
P1	children 29.7; support 28.9; education 26.1; participation 22.0; corporations 20.8	education 28.1; support 27.6; neighbors 24.4; older people 24.3; participation 23.0
P2	COVID-19 40.0; support 31.9; children 27.5; corporations 20.7; education 20.7	COVID-19 45.6; support 31.3; neighbors 30.7; welfare 24.7; education 23.3
P3	support 36.5; children 30.0; participation 23.6; welfare 22.8; education 21.2	support 33.5; welfare 28.3; older people 28.0; neighbors 27.3; children 26.6

Note. Values are group mean TF-IDF scores multiplied by 1,000. Entries are ordered from highest to lower score.

4.4 Keyword Networks: Crisis Compression and Media-Specific Hubs

The network results provide the strongest evidence of crisis-induced agenda compression (Figs. 5 and 6; Table 5). In P1, degree centralization was .457 in national and .306 in local coverage. National hubs included employees, children, people with disabilities, community, and volunteering, producing a visible corporate-social-contribution and policy-target structure. Local hubs emphasized neighbors,

volunteering, care/love, older people, and support, giving the network a more relational and care-oriented center.

In P2, degree centralization rose sharply to .820 in the national network and .669 in the local network. COVID-19 reached degree centrality of 1.000 in national coverage, meaning that it connected directly to all 35 other core terms. Its weighted strength was 4.849 and betweenness .662. In local coverage, COVID-19 had degree centrality .886, weighted strength 5.073, and betweenness .550. The pandemic therefore did not operate as one topic among many; it became the bridge linking vulnerable groups, neighbors, welfare, children, education, the local economy, volunteering, and support. Higher national centralization indicates that national reporting arranged diverse community concerns under a single crisis frame even more strongly than local reporting.

In P3, degree centralization declined to .395 in both media types. National hubs included employees, neighbors, volunteering, people with disabilities, and vulnerable groups. Local hubs included neighbors, welfare, support, vulnerable groups, and older people. The disappearance of a singular pandemic hub restored a more polycentric structure, but social contribution and support for vulnerable groups remained prominent. Spectral clusters repeatedly combined participation and community, corporate contribution and volunteering, welfare and care, and education-cultural-life domains. In P2, these clusters were reorganized around crisis, support, and solidarity. In P3 local coverage, the most cohesive cluster connected neighbors, welfare, support, vulnerability, older residents, and volunteering.

Table 5

Network structure and dominant hubs

Period	Media	Degree centralization	Dominant hub	Hub degree	Hub weighted strength	Hub betweenness
P1	National	.457	Employees / children	—	—	—
P1	Local	.306	Neighbors / volunteering	—	—	—
P2	National	.820	COVID-19	1.000	4.849	.662
P2	Local	.669	COVID-19	.886	5.073	.550
P3	National	.395	Employees / neighbors	—	—	—
P3	Local	.395	Neighbors / welfare	—	—	—

Note. Dashes indicate that the qualitative hub summary, rather than a single comparable node, is reported for that network.

4.5 Sentiment Across Periods and Media Types

Sentiment followed the expected crisis trajectory (Table 6; Figs. 7 and 8). In P1, mean sentiment was .646 in national and .721 in local coverage. Positive articles constituted 90.53% and 92.44%, respectively; negative articles constituted 3.03% and 1.64%. In P2, mean sentiment fell to .438 nationally and .519 locally. The positive share fell to 75.97% and 80.98%, while the negative share rose to 10.75% and 6.04%. P3 brought substantial recovery: mean sentiment reached .638 in national and .692 in local reporting, and positive shares returned to 90.12% and 92.49%.

The annual series shows that the largest break occurred in 2020, when mean sentiment fell to .342 in national and .464 in local coverage. It recovered gradually through 2021 and 2022 and returned close to pre-pandemic levels after 2023. By 2025, mean sentiment was .667 in national and .710 in local newspapers. The period pattern is therefore not an artifact of an arbitrary three-period average; the discontinuity and recovery are visible at annual resolution.

Local sentiment was more positive in every period and overall. The overall mean was .584 in national and .649 in local coverage. Welch tests were statistically significant, as expected with tens of thousands of articles, but standardized effects were small: Cohen's d was − .216 in P1, − .194 in P2, − .158 in P3, and − .174 overall. These differences are consistent with the stronger local prevalence of gratitude, care/love, volunteering, and community-event language. They should not be interpreted as evidence that local communities were objectively happier or that local journalism was less attentive to hardship.

Table 6
Sentiment by period and media type

Period	Media	Mean score	Positive (%)	Neutral (%)	Negative (%)	Cohen's d (N − L)
P1	National	.646	90.53	6.43	3.03	−.216
P1	Local	.721	92.44	5.92	1.64	
P2	National	.438	75.97	13.28	10.75	−.194
P2	Local	.519	80.98	12.98	6.04	
P3	National	.638	90.12	7.25	2.63	−.158
P3	Local	.692	92.49	5.79	1.72	
Overall	National	.584	—	—	—	−.174
Overall	Local	.649	—	—	—	

Note. Sentiment ranges from − 1 to 1. Cohen's d is calculated as national minus local; negative values indicate more positive local coverage. All media differences are statistically significant, but effect sizes are small.

4.6 Common-Window Sensitivity Analysis

The common-window sensitivity analysis restricted the national P3 corpus to May 17, 2023–December 31, 2025, matching the observed local P3 window. This reduced the national P3 sample from 2,772 to 2,514 articles. National mean sentiment changed only from .638 to .645; the local mean remained .692. The rank ordering of national-local keyword gaps was nearly unchanged (Spearman ρ = .991), and nine of the ten largest absolute gaps appeared in both the full-period and common-window lists. The sensitivity results indicate that the principal P3 comparison is not driven by the first four and a half months of national 2023 coverage.

This robustness check does not eliminate the data limitation. The missing local records cannot be assumed to resemble the observed records, and the file's selection order is a platform property rather than a research sampling design. The analysis therefore supports the direction of the reported media differences under a matched date window, not complete representation of all local P3 search results.

Table 7
Sensitivity analysis for the Period 3 export ceiling

Diagnostic	Full-period estimate	Common-window estimate	Interpretation
National P3 articles	2,772	2,514	National data restricted to local start date
National mean sentiment	.638	.645	Minimal change
Local mean sentiment	.692	.692	Unchanged corpus
Keyword-gap rank correlation	—	Spearman ρ = .991	Nearly identical ranking
Top-10 absolute-gap overlap	—	9 of 10	Substantively stable leading gaps

Note. The common window is May 17, 2023–December 31, 2025. The check evaluates robustness but cannot recover the 2,434 unexported local records.

5 Discussion

5.1 Two Complementary Frames of Community Well-Being

The central finding is the persistence of two complementary representations of community happiness. National newspapers more often foregrounded expansion, children, the environment, facilities, the economy, improvement, corporations, and support. This institutional-national frame represents well-being through programs, infrastructure, policy target groups, organizational action, and measurable growth. It is capable of making structural resources visible and connecting local concerns to national policy, but it tends to position communities as sites in which institutions deliver interventions.

Local newspapers more often foregrounded gratitude, neighbors, care/love, older people, volunteering, village, empathy, welfare, and support. This relational-local frame represents well-being through recognition, reciprocity, proximate care, civic activity, and place-based social ties. It gives empirical substance to theories that treat social capital, belonging, and community capacity as constitutive of well-being (Aldrich & Meyer, 2015; Kee & Nam, 2016; Putnam, 2000). Yet gratitude and harmony can also be ritualized. A celebratory story about a donation may make civic solidarity visible while leaving the causes of unmet need or the adequacy of public services unexamined.

The frames are therefore neither opposites nor rankings. Material and institutional resources are required for relational capacity, and relational networks influence whether resources are accessible, trusted, and sustained. A more complete public account of community well-being would combine national journalism's structural and comparative reach with local journalism's attention to lived access and everyday social infrastructure. The empirical gap between media scales identifies not only different emphases but potential opportunities for complementary reporting.

5.2 Crisis-Induced Agenda Compression

The COVID-19 period produced the clearest temporal transformation. The pandemic term ranked first in TF-IDF for both media types, sentiment declined sharply, and network centralization rose. In the national network, COVID-19 connected directly to every other core term and achieved exceptionally high betweenness. This is stronger evidence than a simple frequency increase: it shows that one crisis concept became the relational center through which welfare, education, children, vulnerable groups, neighbors, economic support, and volunteering were interpreted.

The pattern supports the concept of crisis-induced agenda compression. A multi-domain construct was temporarily organized around emergency protection and response. National compression was stronger, perhaps because national outlets aggregated local consequences through central-government announcements, nationwide restrictions, and large organizational responses. Local compression was also substantial, but relational terms remained more central, reflecting the practical delivery of food, masks, care, welfare, and information through proximate institutions and neighbors.

The post-crisis period shows diversification rather than simple restoration. Centralization fell to the same .395 in both media types, and sentiment recovered, but welfare, health, vulnerability, aging, and support remained prominent. The crisis appears to have altered the connectivity of community well-being: health and care were not returned to a separate policy silo but remained linked to local relationships and institutional provision. This finding aligns with community resilience scholarship, which treats formal capacity and social infrastructure as mutually dependent.

5.3 Theoretical and Methodological Contributions

The study contributes to community well-being research in three ways. First, it treats news discourse as a meso-level layer between community conditions and individual evaluations. Existing measures commonly combine objective indicators and resident surveys. Media discourse adds evidence about

which conditions are publicly prioritized, which actors are assigned responsibility, and which solutions circulate. This layer can reveal a rhetorical-policy environment that affects attention and legitimacy even when it does not track actual well-being.

Second, the national-local comparison shows that measurement at one communicative scale can miss important dimensions. A national-only corpus would overrepresent institutional provision, target groups, and corporations relative to the full media system. A local-only corpus could overrepresent event-based solidarity and place relations. Community well-being scholarship benefits from scale-sensitive designs that identify where structural and relational meanings are produced.

Third, the integration of prevalence, TF-IDF, network centralization, clustering, and sentiment provides a more defensible interpretation than any one method alone. Frequency established the broad agenda, TF-IDF identified distinctive vocabulary, networks demonstrated structural compression, and sentiment captured crisis-related tone. The matched-window sensitivity analysis further shows how a data-platform limitation can be acknowledged and tested rather than hidden. This transparency is especially important as scholarship increasingly relies on commercial or public data portals whose export settings shape observable corpora.

5.4 Implications for Policy and Journalism

The findings suggest that community well-being monitoring should use multiple layers of evidence. Objective indicators of services, health, economy, and environment should be combined with resident surveys and recurring analysis of national and local discourse. Divergence is informative: positive event coverage may coexist with deteriorating access, while negative crisis coverage may reflect heightened recognition rather than worsening conditions alone. A quarterly dashboard could therefore track topic prevalence, neglected domains, resident experience, and distributional outcomes together.

National policy communication should move beyond the scale of support, facilities, and program expansion to show whether services reach residents, reduce blind spots, incorporate resident voice, and produce durable changes in daily life. Local communication should retain the value of relational language while connecting gratitude, volunteering, and harmony to structural conditions such as care staffing, transport, housing, employment, environmental quality, and universal service responsibility. Civic contribution should be presented as a complement to, not a substitute for, public obligation.

Disaster planning should also include a community well-being impact assessment. When a health emergency compresses attention, continuity in education, culture, mental health, participation, care, and local livelihoods can be marginalized. Emergency indicators should therefore include isolation, learning disruption, care interruption, participation, and local economic stability alongside infection and mortality. Media monitoring can help identify which domains disappear from public attention during a crisis.

Finally, source diversity should be strengthened at both scales. National journalism can include more residents, small organizations, and local service users; local journalism can supplement event and municipal press-release coverage with data, independent expertise, and critical evaluation. Collaborative

reporting and shared local data infrastructures could connect nationwide structural comparisons to the specificity of lived community experience.

5.5 Limitations and Future Research

Several limitations qualify the findings. First, the concise query maximizes transparency but does not capture every semantic equivalent of community well-being. Articles using “quality of life,” “resident satisfaction,” “local community,” or “well-being” without both Korean search terms may be absent. Conversely, the two words may occasionally occur in unrelated passages. Future work should compare a core query with an expanded dictionary and evaluate precision and recall through manually coded samples.

Second, outlet contributions were unequal. The results describe two aggregated article populations rather than equally weighted outlets. Outlet fixed effects, equal-outlet weighting, regional stratification, and source-type coding could distinguish a media-scale effect from the practices of high-volume newspapers. The stronger representation of Chungcheong-region outlets in the local corpus is especially relevant to generalization.

Third, the P3 local export is incomplete. BIGKinds returned 22,434 results but saved only 20,000 because of the one-time service ceiling, leaving 2,434 unobserved. The retained file begins on May 17, 2023 and cannot represent early 2023. The common-window results are reassuring but do not establish random missingness. Future data collection should split high-volume searches by month or quarter, export all segments, and deduplicate by article identifier.

Fourth, lexicon-based sentiment cannot fully interpret negation, irony, quotation, or context. Event language such as support, love, and gratitude may raise scores even in stories that contain criticism or structural hardship. Stratified human coding with multiple coders and Cohen's kappa, followed by comparison with sentence-level Korean language models, would improve validation.

Fifth, the network results depend on the 36-node and 160-edge display and comparison criteria. Edge limits can constrain density, and spectral-cluster labels involve researcher judgment. Sensitivity analyses varying node count, edge thresholds, association measures, and community-detection algorithms would clarify structural robustness.

Finally, news volume, prevalence, networks, and sentiment do not measure residents' actual happiness. They reflect editorial selection, publicity flows, platform classification, and news-production capacity. Future studies should link discourse measures to regional survey data, community health indicators, public-service access, local economic conditions, and demographic change. Monthly interrupted time-series or event-study designs could also separate pandemic-related shifts from contemporaneous political and economic events. The present analysis is descriptive and observational; it does not identify COVID-19 as the sole causal source of change.

6 Conclusion

Korean news discourse on community happiness moved from a broad pre-pandemic combination of education, participation, welfare, and relationships to a crisis-centered configuration organized around COVID-19 and support, and then to a more diversified post-crisis structure emphasizing health, welfare, care, and cooperation. Across this common temporal arc, national newspapers more often represented well-being through institutions, corporations, expansion, facilities, policy targets, and economic-development language. Local newspapers more often represented it through neighbors, gratitude, care, older residents, volunteering, and place-based relationships.

The practical objective is not to increase the number of positive “happiness” stories. It is to connect structural resources with lived accessibility, private and civic support with public responsibility, and community relations with meaningful resident agency. National and local journalism provide different pieces of that account. Their combination can make community well-being more multidimensional, but only when celebratory support narratives are linked to service adequacy, inequality, and long-term outcomes.

The study's 62,738 collected articles make these patterns visible across a decade, while the 2,434 unexported local records in P3 demonstrate how platform design can constrain large-scale public-data research. Transparent reporting of retrieval, export, exclusion, and analysis counts is therefore a methodological contribution as well as a limitation. Future community well-being research should integrate communicative, perceptual, and objective indicators and should treat the scale of public discourse as part of the phenomenon being studied.

Declarations

Author Contribution

The sole author was responsible for conceptualization, data collection, methodology, analysis, interpretation, visualization, and preparation of the manuscript.

Acknowledgement

Not applicable

Data Availability

The underlying news texts were retrieved from BIGKinds and are subject to the platform’s access and redistribution conditions. The six exported source files cannot be redistributed publicly by the author. An anonymized minimal dataset of derived aggregate tables, dictionaries, and analysis code can be made available for peer review through the journal office and, after publication, upon reasonable request, subject to BIGKinds conditions.

Clinical trial number: not applicable.

References

1. Aldrich, D. P., & Meyer, M. A. (2015). Social capital and community resilience. *American Behavioral Scientist*, *59*(2), 254–269. https://doi.org/10.1177/0002764214550299
2. Andrews, F. M., & Withey, S. B. (1976). *Social indicators of well-being: Americans' perceptions of life quality*. Plenum.
3. Bagnall, A. M., South, J., Di Martino, S., Southby, K., Pilkington, G., Mitchell, B., Pennington, A., & Corcoran, R. (2017). *A systematic review of interventions to boost social relations through improvements in community infrastructure*. What Works Centre for Wellbeing.
4. Cho, Y., & Lee, K. Y. (2021). Does local government affect community satisfaction of the younger generation in rural areas? The case of Jeonbuk, South Korea. *Asian Journal for Public Opinion Research*, *9*(3), 214–239. https://doi.org/10.15206/ajpor.2021.9.3.214
5. Diener, E. (1984). Subjective well-being. *Psychological Bulletin*, *95*(3), 542–575. https://doi.org/10.1037/0033-2909.95.3.542
6. Drew, J., & Kim, Y. (2025). Community wellbeing and public management theory: How theory can explain pandemic responses. *International Journal of Community Well-Being*, *8*, 43–59. https://doi.org/10.1007/s42413-025-00235-6
7. Entman, R. M. (1993). Framing: Toward clarification of a fractured paradigm. *Journal of Communication*, *43*(4), 51–58. https://doi.org/10.1111/j.1460-2466.1993.tb01304.x
8. Freeman, L. C. (1978). Centrality in social networks: Conceptual clarification. *Social Networks*, *1*(3), 215–239. https://doi.org/10.1016/0378-8733(78)90021-7
9. Kee, Y., & Nam, C. (2016). Does sense of community matter in community well-being? In Y. Kee, S. J. Lee, & R. Phillips (Eds.), *Social factors and community well-being* (pp. 39–56). Springer.
10. Kim, J. H. (2019). An exploratory study of the health inequality discourse using Korean newspaper articles: A topic modeling approach. *Journal of Preventive Medicine and Public Health*, *52*(6), 384–392. https://doi.org/10.3961/jpmph.19.221
11. Korea Press Foundation (2026). BIGKinds: News big data analysis service. https://www.bigkinds.or.kr/
12. Lee, J., Chang, H. E., Cho, J., Yoo, S., & Hyeon, J. (2024). Analysis of issues related to nursing law: Examination of news articles using topic modeling. *PLOS ONE*, *19*(8), e0308065. https://doi.org/10.1371/journal.pone.0308065
13. Lee, K. Y., & Park, K. (2022). Perception of community environment, satisfaction with local government, and quality of life: The case of Gyeonggi. *Korea Social Sciences*, *11*(9), 394. https://doi.org/10.3390/socsci11090394
14. Lee, S. J., & Kim, Y. (2018). Economy doesn't buy community wellbeing: A study of factors shaping community wellbeing in South Korea. *International Journal of Community Well-Being*, *1*(1), 33–44.

https://doi.org/10.1007/s42413-018-0004-2

15. McCombs, M. E., & Shaw, D. L. (1972). The agenda-setting function of mass media. *Public Opinion Quarterly*, *36*(2), 176–187. https://doi.org/10.1086/267990
16. Narayan, B. (2020). Leveraging digital intelligence for community well-being. *International Journal of Community Well-Being*, *3*, 539–558. https://doi.org/10.1007/s42413-020-00085-4
17. Ng, A. Y., Jordan, M. I., & Weiss, Y. (2002). On spectral clustering: Analysis and an algorithm. *Advances in neural information processing systems 14* (pp. 849–856). MIT Press.
18. OECD (2025). OECD guidelines on measuring subjective well-being: 2025 update. OECD Publishing. https://doi.org/10.1787/9203632a-en
19. Pang, B., & Lee, L. (2008). Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval*, *2*(1–2), 1–135. https://doi.org/10.1561/1500000011
20. Park, D., Kim, D., & Park, A. H. (2024). Agendas on nursing in South Korea media: Natural language processing and network analysis of news from 2005 to 2022. *Journal of Medical Internet Research*, *26*, e50518. https://doi.org/10.2196/50518
21. Passey, M., Davies, K., & Evans, S. (2024). The gap between rhetoric and reality in community wellbeing research: Knowledge, perspectives, and priorities. *International Journal of Community Well-Being*, *7*, 427–446. https://doi.org/10.1007/s42413-024-00216-1
22. Putnam, R. D. (2000). *Bowling alone: The collapse and revival of American community*. Simon & Schuster.
23. Salton, G., & Buckley, C. (1988). Term-weighting approaches in automatic text retrieval. *Information Processing & Management*, *24*(5), 513–523. https://doi.org/10.1016/0306-4573(88)90021-0
24. Sirgy, M. J., Rahtz, D. R., Cicic, M., & Underwood, R. (2000). A method for assessing residents' satisfaction with community-based services: A quality-of-life perspective. *Social Indicators Research*, *49*(3), 279–316. https://doi.org/10.1023/A:1006990718673
25. Sirgy, M. J., Widgery, R. N., Lee, D. J., & Yu, G. B. (2010). Developing a measure of community well-being based on perceptions of impact in various life domains. *Social Indicators Research*, *96*, 295–311. https://doi.org/10.1007/s11205-009-9479-9
26. Taboada, M., Brooke, J., Tofiloski, M., Voll, K., & Stede, M. (2011). Lexicon-based methods for sentiment analysis. *Computational Linguistics*, *37*(2), 267–307. https://doi.org/10.1162/COLI_a_00049
27. Theodori, G. L. (2001). Examining the effects of community satisfaction and attachment on individual well-being. *Rural Sociology*, *66*(4), 618–628. https://doi.org/10.1111/j.1549-0831.2001.tb00087.x
28. Wasserman, S., & Faust, K. (1994). *Social network analysis: Methods and applications*. Cambridge University Press.
29. Wiseman, J., & Brasher, K. (2008). Community wellbeing in an unwell world: Trends, challenges, and possibilities. *Journal of Public Health Policy*, *29*(3), 353–366. https://doi.org/10.1057/jphp.2008.16

30. Woo, C., & Jung, H. (2022). The impact of social enterprises on individual wellbeing in South Korea: The moderating roles of social capital in multilevel analysis. *Social Indicators Research*, *159*, 433–454. https://doi.org/10.1007/s11205-021-02759-8

Figures

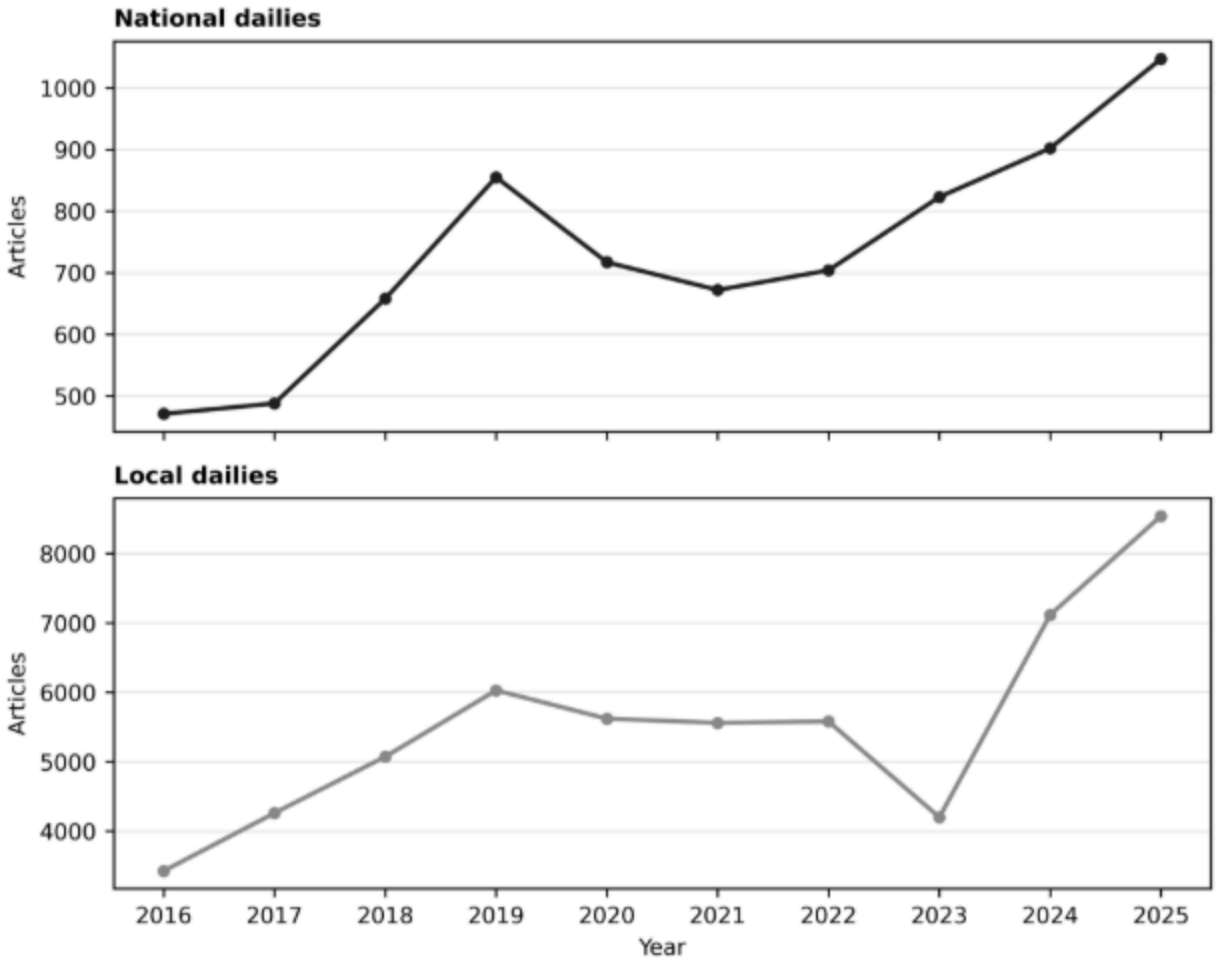


Figure 1

Annual number of analyzed articles by media type, 2016–2025.

Note. The local-newspaper series for 2023 begins on May 17 because of the export ceiling and is not a complete annual count.

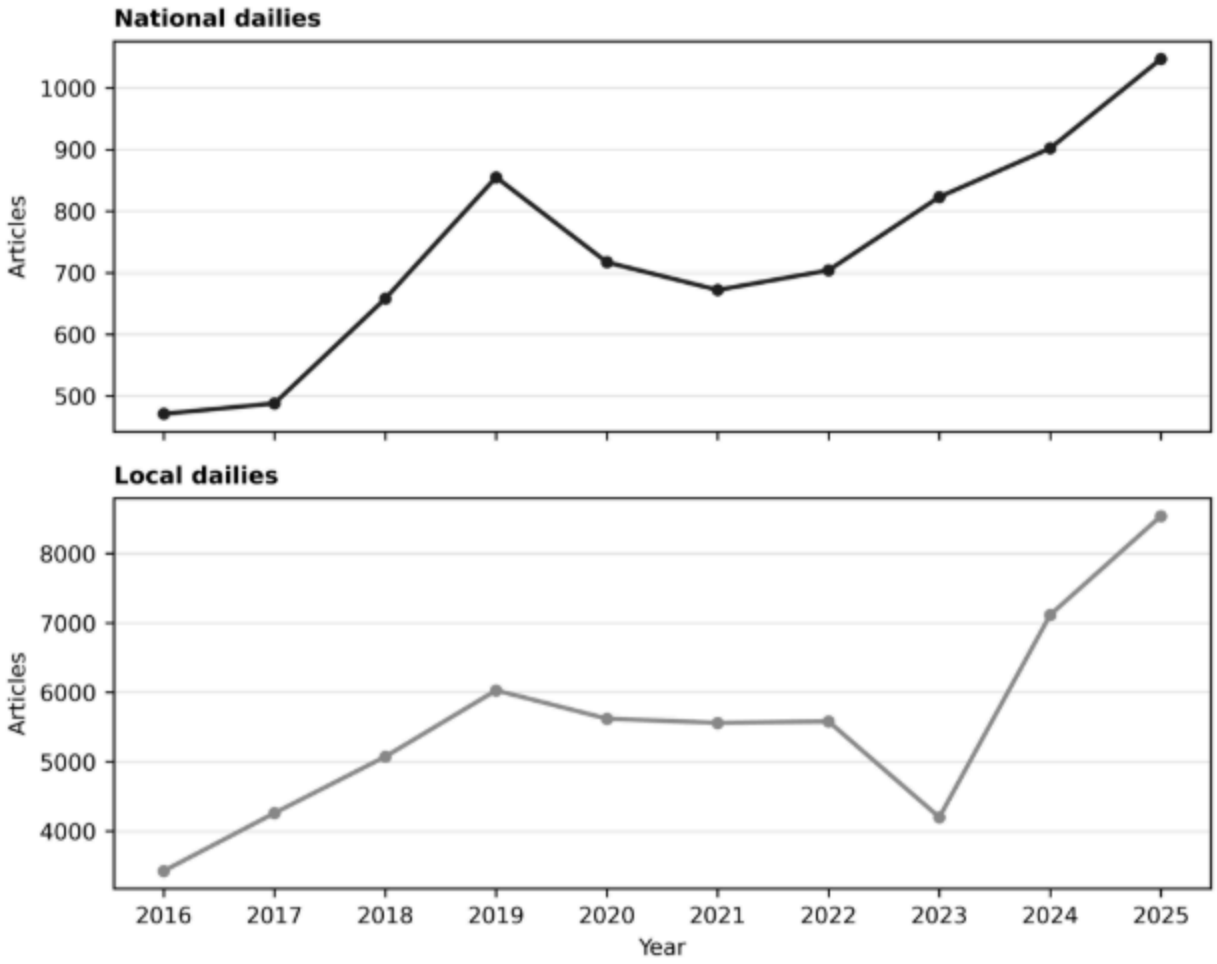


Figure 1

Annual number of analyzed articles by media type, 2016–2025.

Note. The local-newspaper series for 2023 begins on May 17 because of the export ceiling and is not a complete annual count.

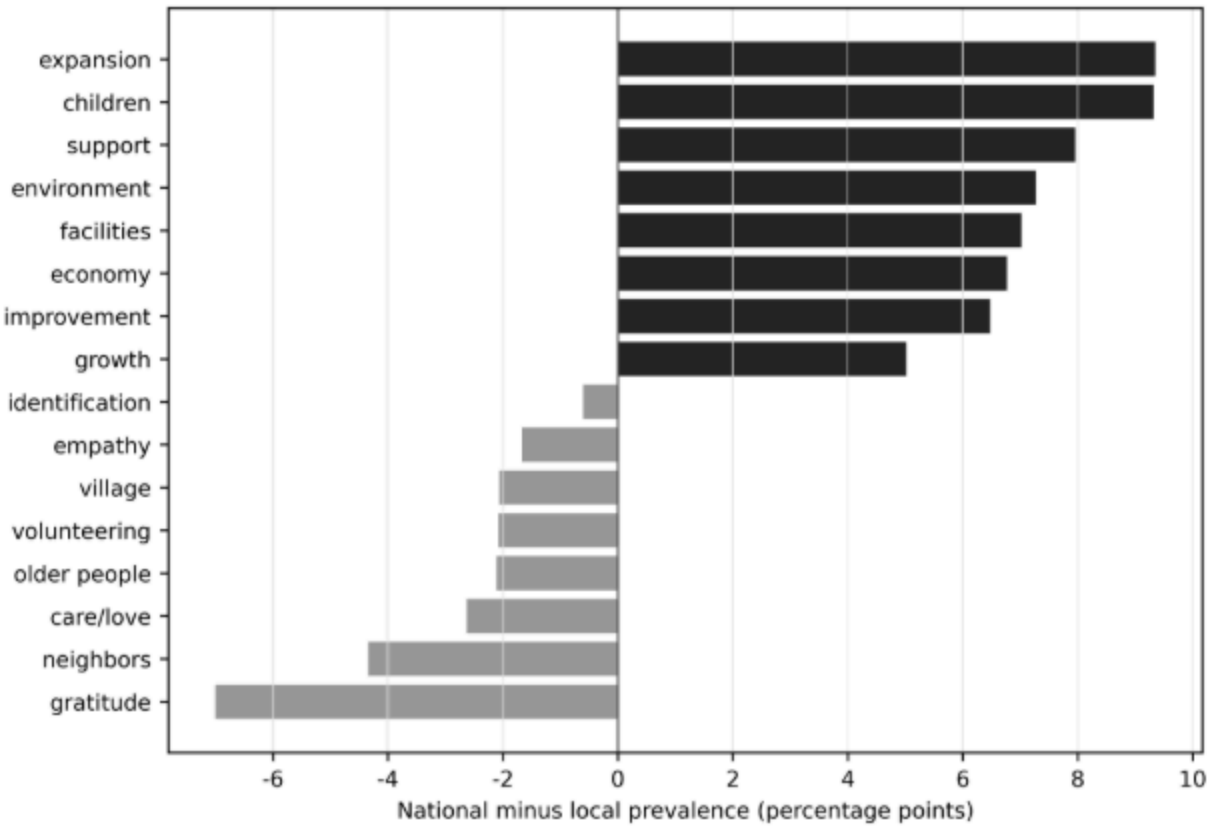


Figure 2

Largest national-local differences in article-level keyword prevalence.

Note. Positive values indicate greater prevalence in national dailies; negative values indicate greater prevalence in local dailies.

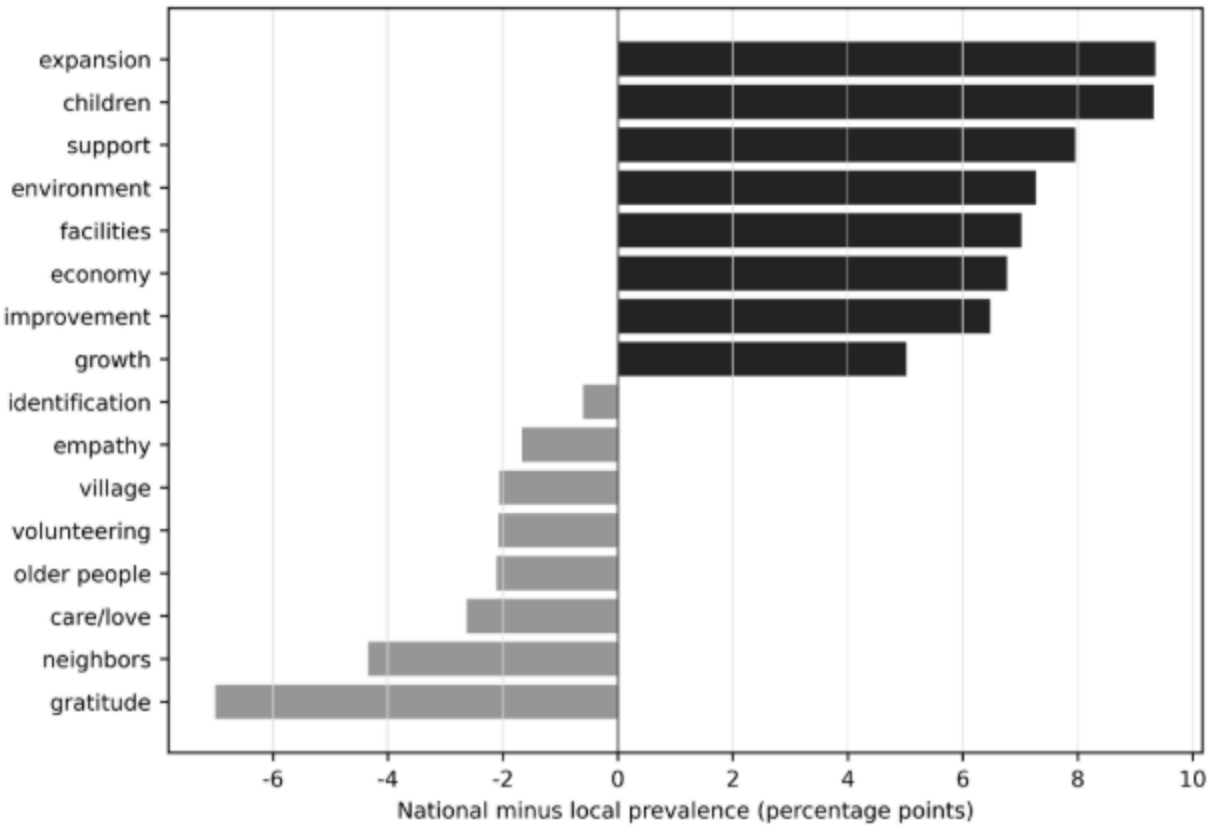


Figure 2

Largest national-local differences in article-level keyword prevalence.

Note. Positive values indicate greater prevalence in national dailies; negative values indicate greater prevalence in local dailies.

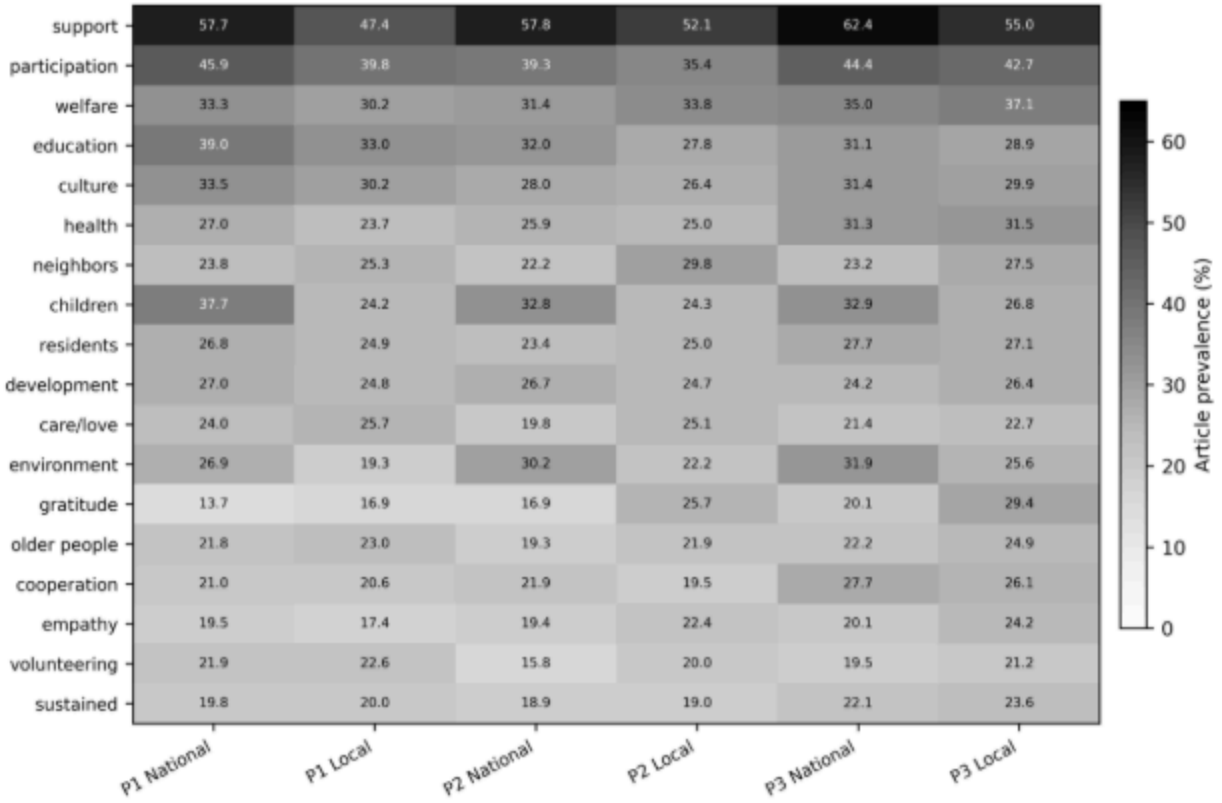


Figure 3

Article-level prevalence of major keywords by period and media type.

Note. Cells report the percentage of articles in each group containing the keyword at least once.

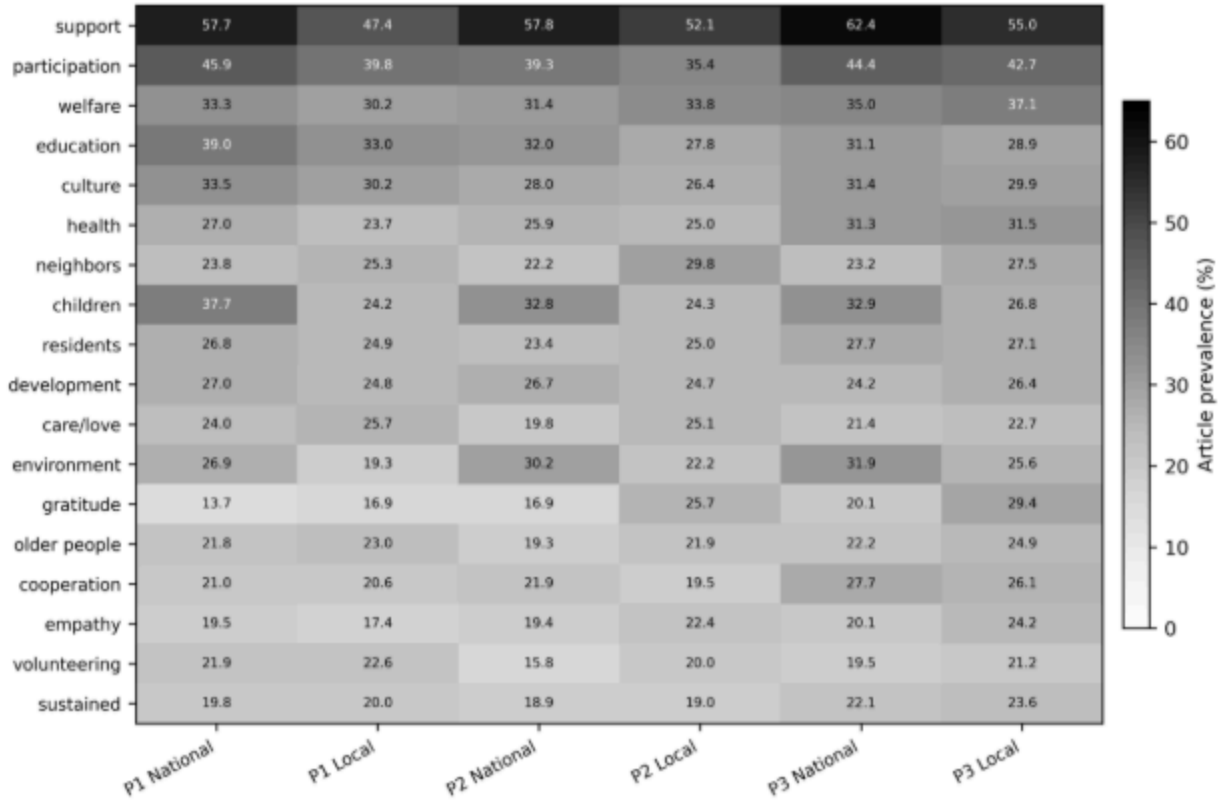


Figure 3

Article-level prevalence of major keywords by period and media type.

Note. Cells report the percentage of articles in each group containing the keyword at least once.

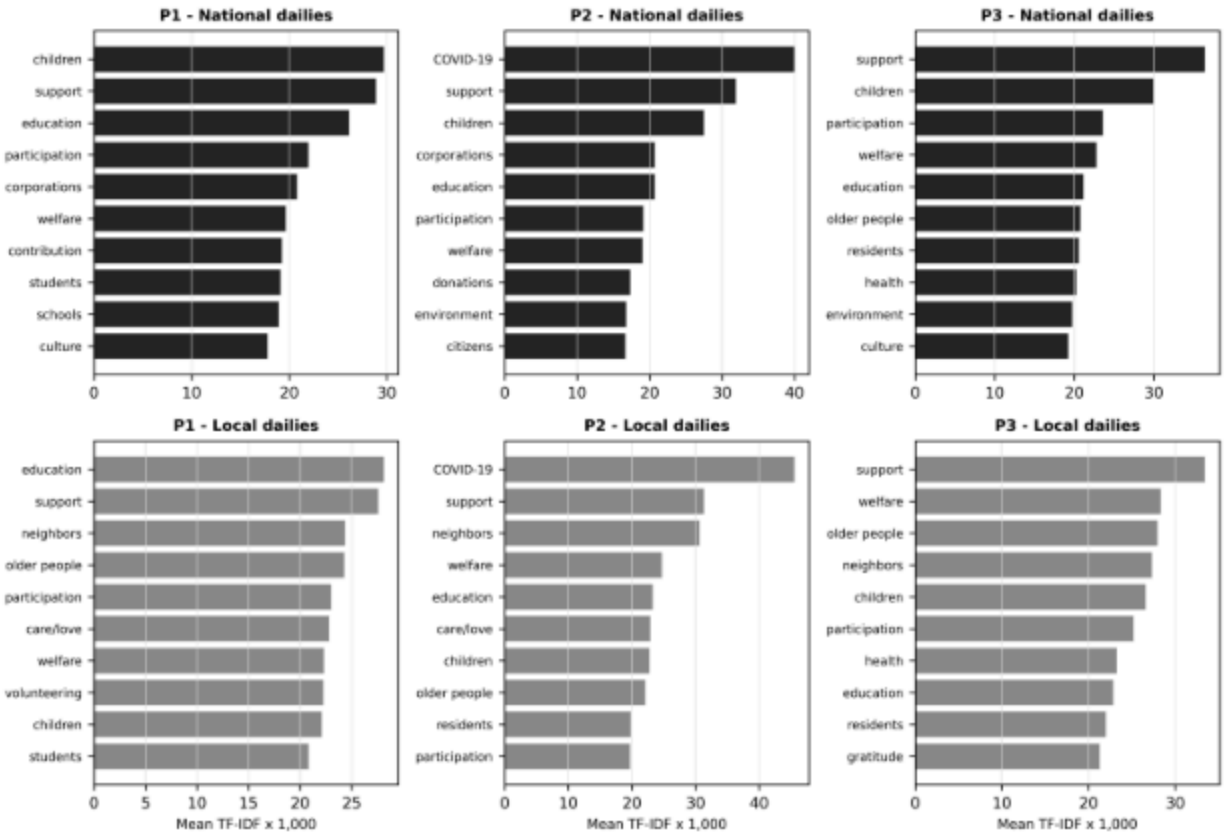


Figure 4

Highest mean TF-IDF terms by period and media type.

Note. TF-IDF scores are group means multiplied by 1,000.

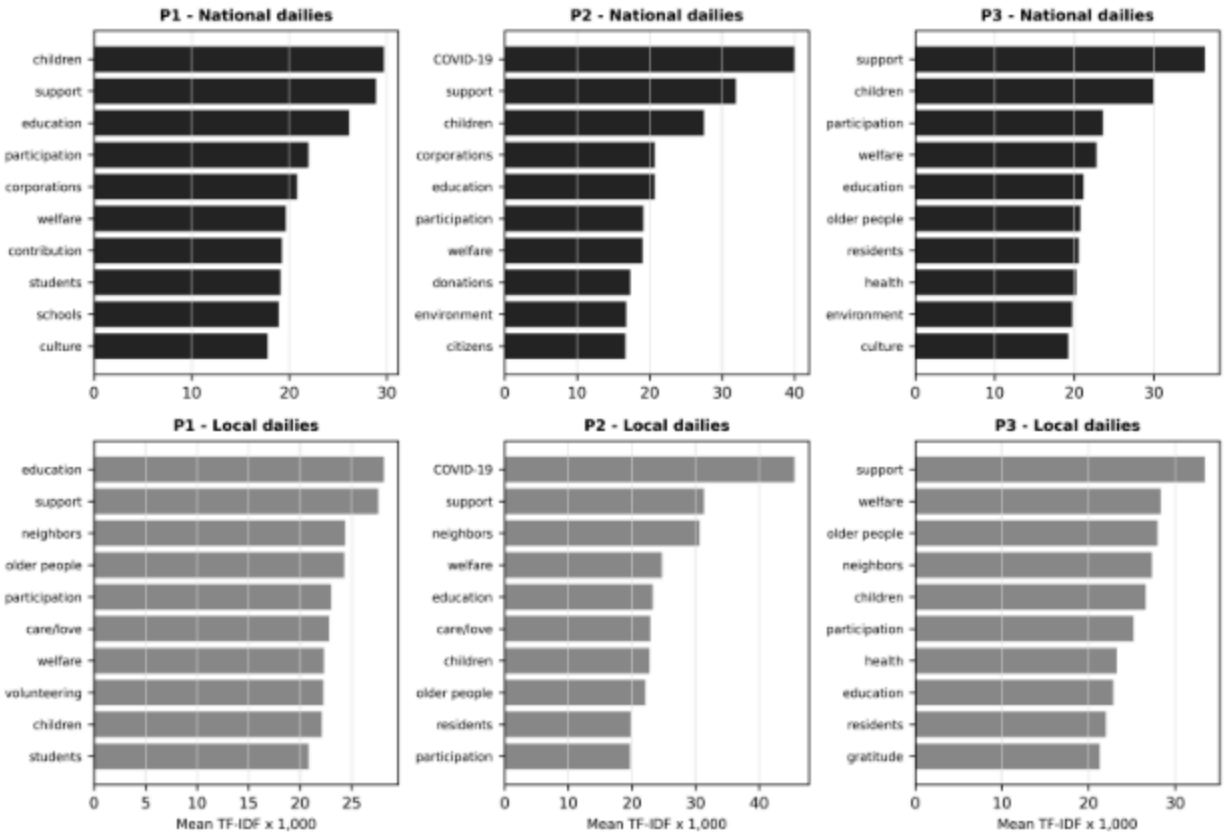


Figure 4

Highest mean TF-IDF terms by period and media type.

Note. TF-IDF scores are group means multiplied by 1,000.

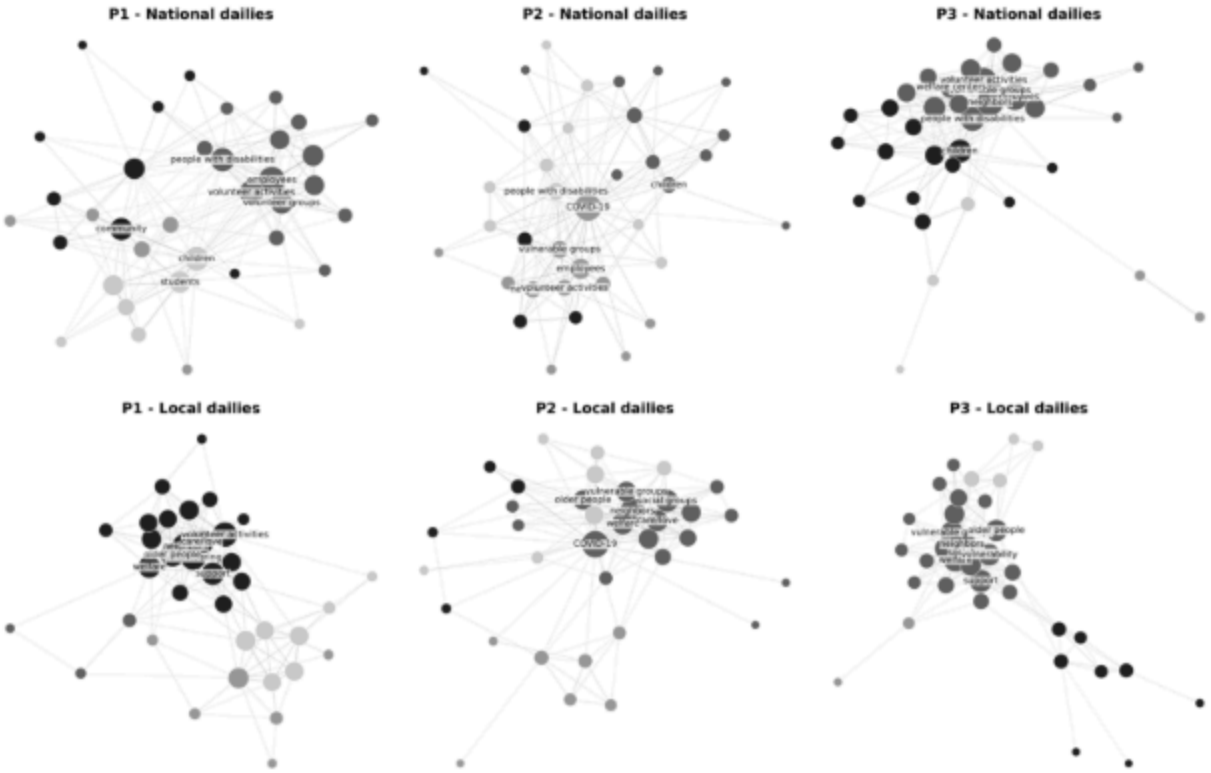


Figure 5

Keyword co-occurrence networks by period and media type.

Note. Node size represents weighted strength, line weight represents association strength, and grayscale shade distinguishes spectral clusters. Labels are shown for the most central nodes.

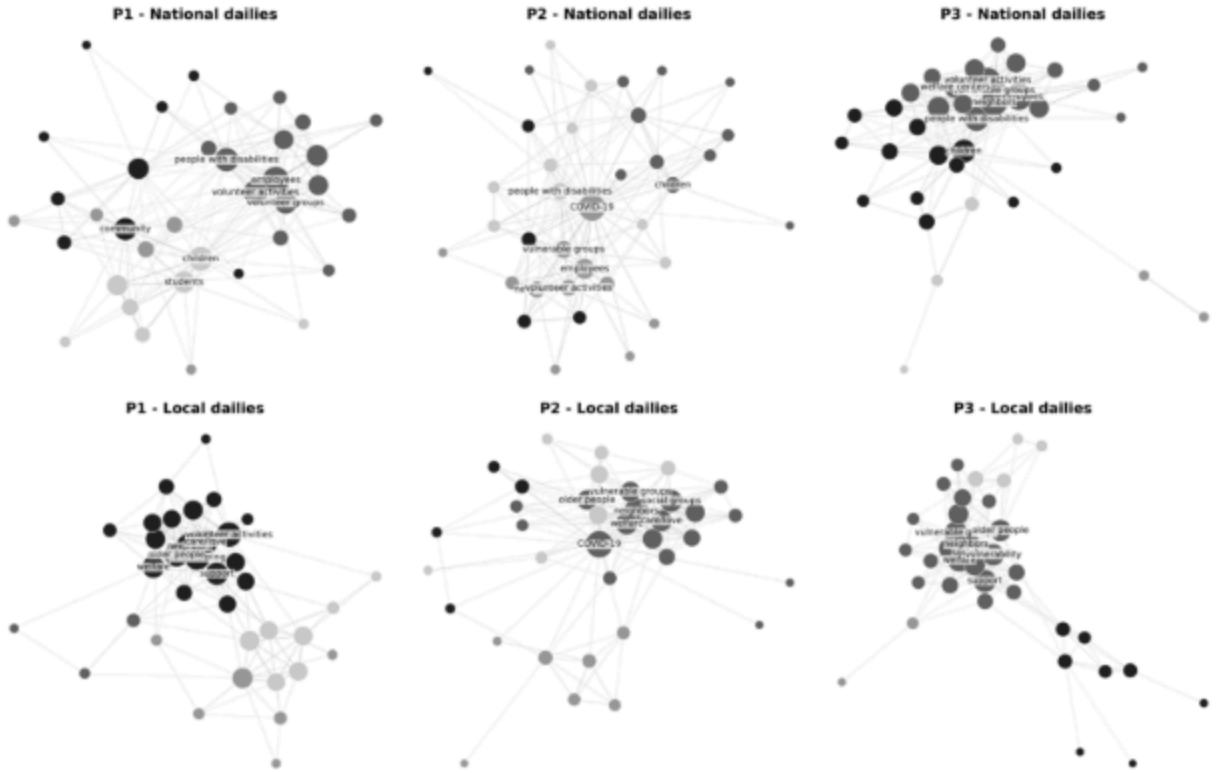


Figure 5

Keyword co-occurrence networks by period and media type.

Note. Node size represents weighted strength, line weight represents association strength, and grayscale shade distinguishes spectral clusters. Labels are shown for the most central nodes.

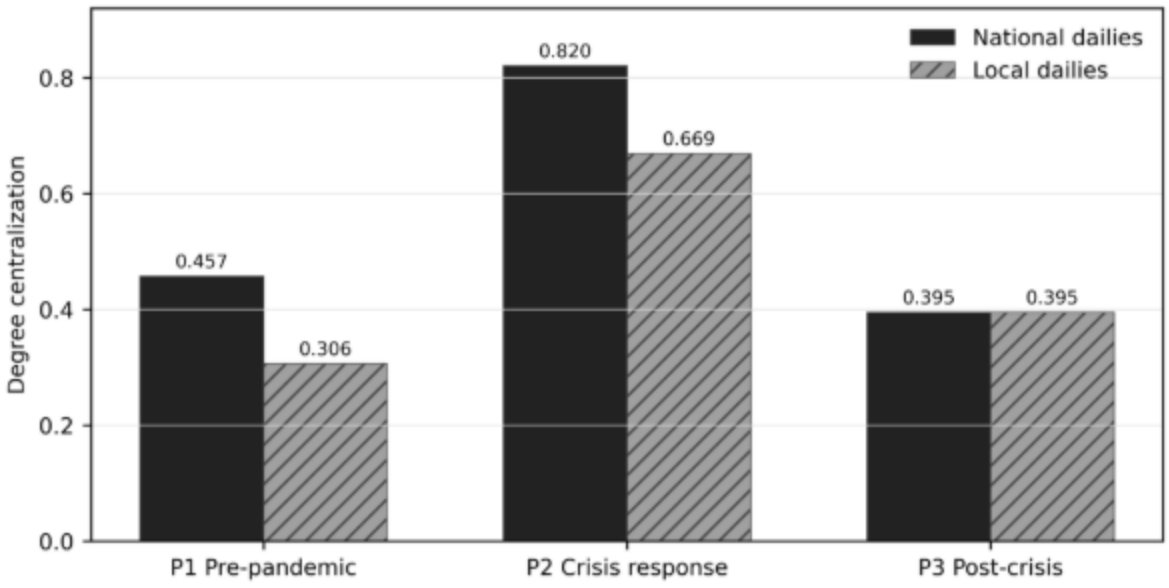


Figure 6

Degree centralization of keyword networks by period and media type.

Note. Higher values indicate that a smaller number of hubs connect a larger share of the network.

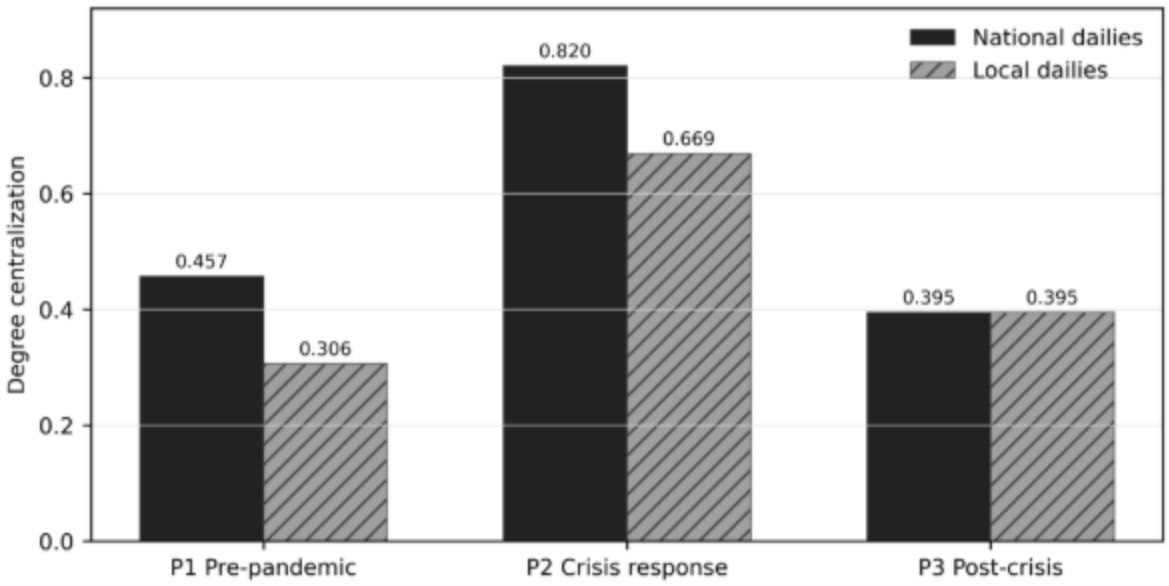


Figure 6

Degree centralization of keyword networks by period and media type.

Note. Higher values indicate that a smaller number of hubs connect a larger share of the network.

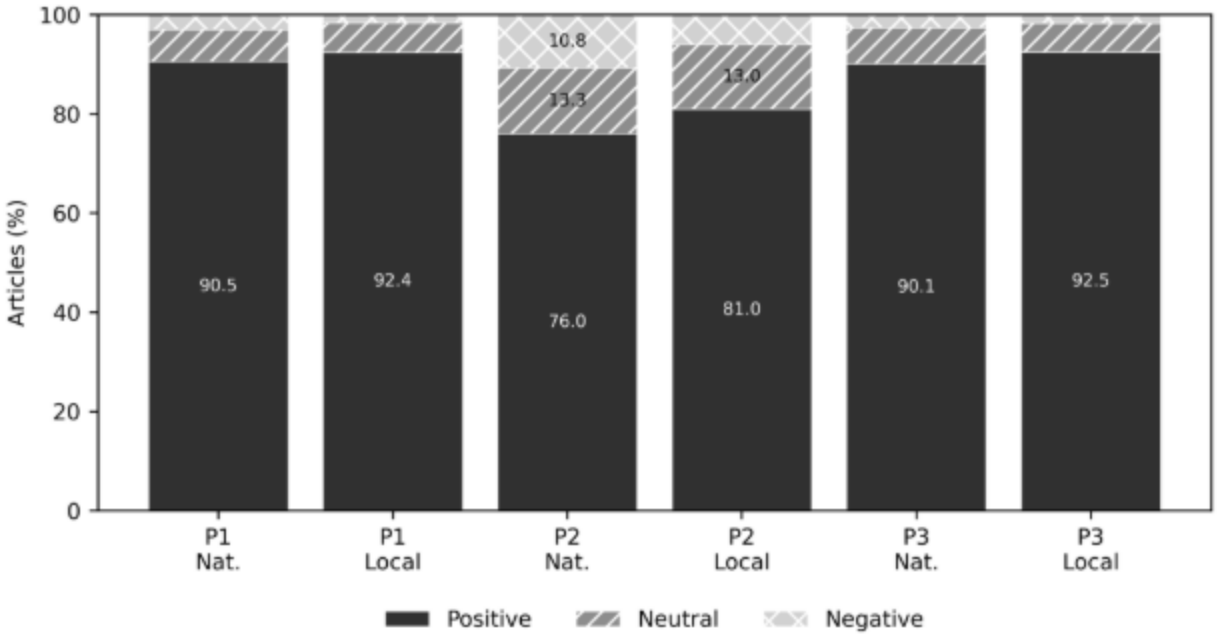


Figure 7

Composition of positive, neutral, and negative articles by period and media type.

Note. Classification thresholds were .15 and −.15. Happiness and community were excluded from the lexicon score.

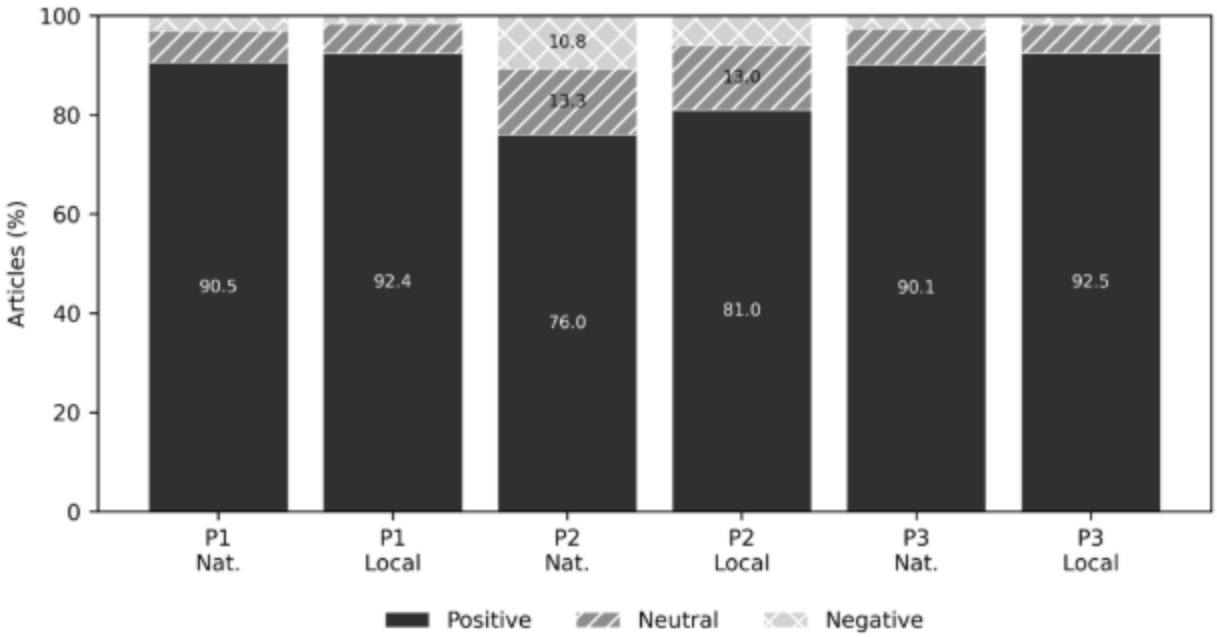


Figure 7

Composition of positive, neutral, and negative articles by period and media type.

Note. Classification thresholds were .15 and −.15. Happiness and community were excluded from the lexicon score.

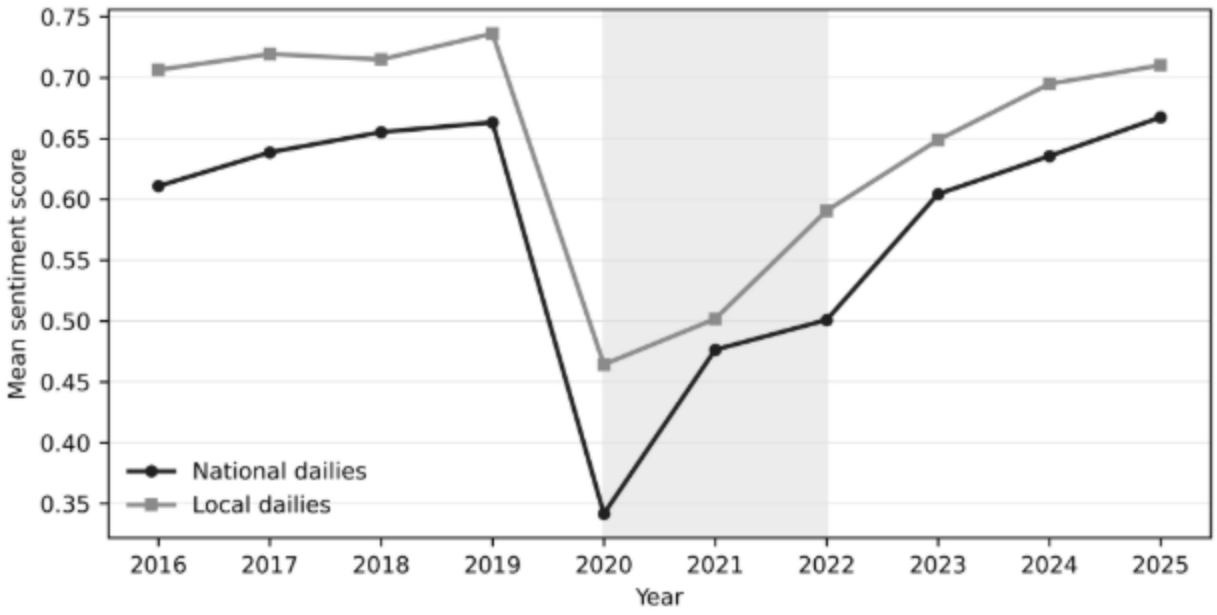


Figure 8

Annual mean sentiment by media type, 2016–2025.

Note. The shaded band indicates the COVID-19 crisis-response period, 2020–2022.

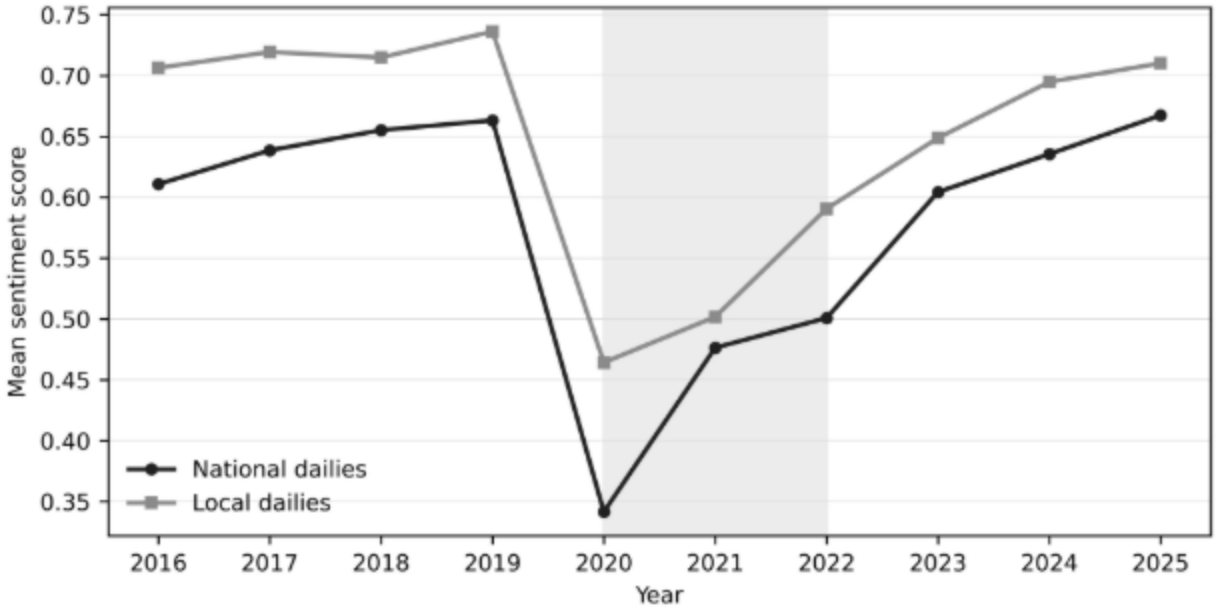


Figure 8

Annual mean sentiment by media type, 2016–2025.

Note. The shaded band indicates the COVID-19 crisis-response period, 2020–2022.