

Article

A Network Analysis Approach to Detecting Social Issues with Web-Based Data

Seunghyun Lee ¹, Jiho Lee ¹, Jae-Min Lee ², Hong-Woo Chun ² and Janghyeok Yoon ^{1,*}

¹ Department of Industrial Engineering, Konkuk University, Seoul 05029, Republic of Korea; ant6836@konkuk.ac.kr (S.L.); jiholee255@konkuk.ac.kr (J.L.)

² Division of Data Analysis, Korea Institute of Science and Technology Information, Seoul 02456, Republic of Korea; jmlee@kisti.re.kr (J.-M.L.); hw.chun@kisti.re.kr (H.-W.C.)

* Correspondence: janghyoon@konkuk.ac.kr

Abstract: Social issues refer to topics that occur and become increasingly focused in various areas of society. Because of the evolutionary pattern of issues, detecting social issues requires monitoring various stories formed by members of society over time. Various studies related to issue detection have been preceded, but it is necessary to supplement in two aspects: presenting the time when issues occurred and prioritizing issues by urgency. As a remedy, the purpose of this study is to propose a new approach to detecting social issues from web-based data through network analysis. Since stories that form social issues are composed of various keywords and topics, this study detects social issues by monitoring keyword co-occurrence networks constructed with web-based data. Specifically, this approach uses network structure entropy to identify a time period at which social issues occur. Next, a community detection algorithm is used to extract social issue candidates in the identified time period. Finally, social issues are detected by deriving the priority of social issue candidates through the centrality index of keywords constituting the candidates. This study detected South Korean social issue topics that attract people's attention among the various topics of society. The proposed approach contributes to the existing literature by identifying when social issues occurred quantitatively based on the characteristics of issues. In addition, since the proposed approach detects urgent issues to be dealt with priority, it can support timely responses to social issues.

Keywords: social issue; web-based data; network analysis; network structure entropy; community detection algorithm



Citation: Lee, S.; Lee, J.; Lee, J.-M.; Chun, H.-W.; Yoon, J. A Network Analysis Approach to Detecting Social Issues with Web-Based Data. *Appl. Sci.* **2023**, *13*, 8516. <https://doi.org/10.3390/app13148516>

Academic Editor: Tao Zhou

Received: 30 June 2023

Revised: 17 July 2023

Accepted: 21 July 2023

Published: 23 July 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

1.1. Background

People create issues from problems or topics they are interested in [1], so an issue is defined as a specific topic or problem that is formed from the diverse stories of different individuals or groups and becomes increasingly focused [2]. As modern society develops rapidly, numerous issues arise in various fields of society such as economy, politics, and industry, and analyzing these issues has long been regarded as a core task in society [3]. If a social issue is not adequately addressed, it is possible it will develop into a national challenge, which will cause enormous economic losses, such as climate change, food crisis, cybersecurity, etc. Since an issue has organically evolving patterns like appearing for a brief period of time or consistently becoming worse over time [4], in order to detect social issues, continuous monitoring of the overall picture of society is required over time.

To detect social issues, research on event detection using web-based data has been actively conducted and is divided into three categories: feature-based approach, topic-modeling-based approach, and incremental approach [5]. The feature-based approach detects issues by utilizing various features that can be extracted from documents [6,7], and the topic-modeling-based approach uses probabilistic models to detect issues [8,9]. These approaches require parameters to be determined in advance, such as the number of issues

and the period to be monitored. Since the dynamic patterns of social issues, it is hard to define the parameters mentioned above. This was addressed by suggesting an incremental approach that detects both existing issues and new ones as they arise [10,11].

1.2. Motivation

Despite the continuous improvement of the above studies, there are some parts to be supplemented. First of all, it is challenging to describe the time when the issues occurred quantitatively. The majority of prior studies that took dynamic patterns of social issues into account classified a group of similar documents as an issue. If a group of similar documents is defined as an issue, it is recognized as a significant issue when sufficient documents are accumulated over time. In order to illustrate the time when the social issue occurred, by quantitatively assessing the moment when society suddenly changed, it is required to use the concept that an issue is a topic or problem that suddenly captures people's attention. Second, it can be hard to rank issues by urgency or importance. Prior studies either stopped at identifying various issues or presented events with a higher frequency of occurrence as important issues. When some documents were merged into a new issue, it was unclear whether the issue was created by concentration from people or by a minor social event. With this method, multiple issues must be continuously reviewed, making it hard to detect urgent issues in a timely manner. In addition, according to the concept of an issue mentioned above, issues that are frequently encountered or have a lot of relevant documents are not always the highest priorities. As a result, it is necessary to rank the issues that people suddenly became focused on at a particular time while taking the definition of an issue into consideration.

In light of the above, this study proposes an approach for detecting social issues from web-based data using network analysis, considering the definition and characteristics of issues. The first step of this approach is constructing co-occurrence networks of keywords extracted from web-based data by time period. As mentioned earlier, an issue is a specific topic or problem that is formed from a variety of stories. A point to note here is that a story includes several topics, and each topic is composed of various combinations of keywords [12]. In other words, various keywords gather to form a topic, various topics gather to form a story, and issues are created from the stories. Therefore, in this approach, the keyword co-occurrence networks are constructed to consider combinations of various keywords forming topics and stories. At this time, since the issue has organic evolution patterns over time, a keyword co-occurrence network is constructed for each time period. Second, in order to identify the time period in which social issues occurred, the network structure is quantified for each time period and the structural change is monitored. In a keyword co-occurrence network, the influence of keywords and the relationship between keywords can be quantitatively presented, such as a centrality index, based on the co-occurrence frequency of the keywords [13]. After calculating a network structure entropy using the centrality indices of the keywords constituting the networks [14], the time period in which the entropy value changes rapidly is identified. Third, a community detection algorithm that detects a group more densely clustered than other groups within a network is applied to extract social issue candidates from the network in the identified time period. And finally, the most urgent social issues are detected by deriving the priority of the extracted social issue candidates using the centrality index of the keywords utilized in the process of calculating the network structure entropy. Through the amount of change in the keyword centrality index at the time period when the network structure entropy drastically changed, it is possible to detect which topic the members of society suddenly focused on.

1.3. Contributions

This study has several contributions in that it solves the parts that need to be supplemented of prior studies which detected social issues. First, this study identifies the time period when social issues occurred quantitatively by keeping an eye on the structural

changes of the network prior to examining the contents of the issue. As a result, since the concept of an issue is that it suddenly captures people's attention, the proposed approach provides objective results while meeting the definition and characteristics of the issue. Second, it is possible to support timely responses to social issues by suggesting the priority of issues as well as the issues occurred time together. This study prioritizes the social issues based on the degree to which people suddenly became interested in, not the rank of issues that have shown up frequently. By utilizing the results presented in this study, it is expected that when social issues occur, urgent issues can be responded to sequentially or selectively. Furthermore, since the proposed approach is wide and unbiased, it can support decision-making in various fields of society.

1.4. Organization

The remainder of this paper is organized as follows. Section 2 reviews the related literature. Section 3 describes the proposed approach, which is illustrated by a case study in Section 4. Lastly, Section 5 concludes with our limitations and proposes future research directions.

2. Literature Review

2.1. Web-Based Social Data Analysis

Web-based data, mainly classified by social media, online discussion groups, and online publications, is one of the open data types that is easy to collect and very fast to update compared to other open data types, such as commercial material or public government data [15]. Web-based data plays an increasingly important role in our society as more information and services are delivered over the Internet. Accordingly, many studies have analyzed web-based data in the various fields of our society for analyzing voice of the customer, identifying business opportunities, monitoring the overall stream of society, etc. (Table 1).

First, some previous studies analyzed social media data to identify customer needs or measure their satisfaction. Social media is a group of Internet-based applications that allow individuals or groups to create and discuss content freely [16], so customers can freely write their experiences and opinions about products or services they have used in social media. A study analyzed app review data from Google Play for investigating determinants of customer satisfaction in music streaming services [17]. The study provided a comprehensive understanding of music streaming services from the customer's point of view. Another study analyzed online travel reviews from TripAdvisor to examine travelers' choice behavior for green hotels [18]. The results of the study could support travelers in their decision-making process and provide insights for hotel managers and green policy makers. In addition, a study analyzed user-generated content on Twitter to compare user emotions before and after the launch by sentiment analysis [19]. Using the approach proposed in the study, businesses could obtain real-time feedback on new product experiences and adjust their new product development processes accordingly.

Second, some studies have analyzed business opportunities of the target product or service using the data of online discussion groups, which refer to platforms that form communities of users on various topics and exchange opinions on topics by creating posts and comments. A study analyzed Reddit data generated by customers of a target product for identifying time-evolving product opportunities via the aging theory-based event detection and tracking algorithm [20]. The study contributed to identifying time-evolving product opportunities from large-scale web-based data and monitoring customer needs in real time. In another study that analyzed Reddit data, an opportunity mining approach was proposed to identify product opportunities based on topic modeling and sentiment analysis [21]. The proposed approach contributed to the systematic identification of product opportunities for product planning.

Finally, numerous studies have analyzed trends, patterns, or events in various fields to monitor the stream of our society by utilizing web news data from online publications.

The web news is public information that is more sophisticated and reliable than other web-based data, such as blogs or customer reviews [22]. A study utilized web news data focusing on emerging technologies together with patent data to identify the development trends of emerging technologies [23]. The study provided a new research perspective by adopting web news data for analyzing emerging technology trends. Similarly, another study analyzed patent data and news data together to investigate the long-run relationship and coevolution patterns between technology and society [24]. The study contributed to increasing the explanatory power of coevolution patterns between society and technology, and establishing insightful strategies for technological advancement. Additionally, there were various studies that utilize web-based data to detect issues or important events in society. The next section describes the prior studies for detecting social issues or events using web-based data.

Table 1. Various studies related to web-based social data analysis.

Data	Methods	Purpose	Reference
Music streaming app reviews	Topic model, Text regression	Investigation of determinants for customer satisfaction in music streaming services	[17]
Online hotel reviews	K-means clustering, TOPSIS, Classification and Regression Trees	Examination of green hotel selection behavior of travelers	[18]
Pizza, car, smart phone tweets	Text mining, Sentiment analysis	Comparison of users' sentiments before and after the launch of three new products	[19]
Smart speaker reviews	Event detection and tracking algorithm, Opportunity algorithm, Sentiment analysis	Identification of time-evolving product opportunities	[20]
Smart phone reviews	Opportunity algorithm, Topic model, Sentiment analysis	Identification of product opportunities	[21]
Web news, Patent	Topic model, Sentiment analysis	Identification of emerging technologies development trends	[23]
Web news, Patent	Text mining, Vector autoregressive model	Investigation of long-run relationship and coevolution patterns of technology and society	[24]

2.2. Social Issue Detection

The studies detect social issues or important events by utilizing web-based data mainly proposed an event detection approach, and as mentioned in the introduction, they can be divided into three types: feature-based approach, topic-modeling-based approach, and incremental approach. A summary of the related literature using various web-based data and data analysis methods can be found in Table 2.

First, the feature-based approach uses various features that can be extracted from documents or posts for detecting events in our society. For example, some researchers collected data from the web and used keywords extracted from the text of the data to detect events [6,7,25], and others have used keywords along with other information such as hashtags, timestamps, and locations of web posts [26,27]. The second approach is the topic-modeling-based approach which utilizes probabilistic models to detect issues or events. The topic-modeling-based approach assumes that latent topics are always present in the documents being processed [11]. A document is made up of a mixture of topics, each topic having a probability distribution for the keywords included in that document. And each topic distribution is considered an event. Latent Dirichlet Allocation (LDA) is a probabilistic topic model that assumes the topic distribution has a Dirichlet prior [28]. The LDA is the most widely used topic model due to its high performance in handling large documents and interpreting the identified latent topics. Therefore, many studies proposed a topic-modeling-based approach based on LDA. The Evolutionary Context-Aware Sequential Model (ECSM), which uses an LDA-like topic layer to capture the context-aware se-

mantic coherences, was proposed to track the topic evolution [29]. In addition, a retweeting behavior-based topic model called RL-LDA [30], which leverages retweet behavior to handle event evolution, and a topic model for microblogs called Bursty Event dEtection (BEE+) [31], which can detect bursty events from short-text and model the temporal information, were proposed. And many other topic-modeling-based approaches have also been suggested [8,32,33]. The feature-based and topic-modeling-based approaches require parameters to be determined, such as the number of issues and the period to be monitored before applying the approaches. The dynamic patterns of social issues make it difficult to define the parameters in advance.

To overcome the limitations mentioned above, the incremental approach was proposed. The incremental approach is a method to detect not only events, but also upcoming events, and has been used in various studies until recently. A study proposed an evolutionary model for event detection to capture the dynamics and evolving behavior of events [5]. The study defined an event as a collection of documents from social media that have common content. Another study proposed an event detection system called TwitterNews+ to identify both major and minor newsworthy events from Twitter [11]. In the study, similar tweets were grouped and defined as an event. Additionally, based on document representation using word embedding, an adaptive, online clustering method was proposed for online news event detection using time slicing and event merging [10]. This study also defined a group of news documents with similar content as an event. Most of the studies that used the incremental approach to consider dynamic patterns of social issues grouped similar documents and defined them as an issue or event. When a few documents become a new event, it is not clear whether the event was created by people's concentration or a meaningless event. To explain the time when social issues or important events occurred, it is necessary to use the concept toward an issue or topic that is suddenly attracting people's attention. Furthermore, it can be difficult to rank issues by degree of urgency. Most studies either detect social issues or simply suggest that more frequent issues are more important. If a group of similar documents is defined as an issue, it can be recognized as a significant or urgent issue when enough documents are accumulated over time. This method requires continuous monitoring of numerous issues, making it difficult to detect urgent issues in a timely manner.

Table 2. Summary of literature review for social issue detection using web-based data.

Approach	Data	Methods	Reference
Feature-based approach	Twitter data	Mention anomaly, Network analysis	[6]
	Twitter data	Text mining, Keyword entropy	[7]
	Twitter data	Text mining, Soft frequent pattern mining	[27]
Topic-modeling-based approach	Twitter data, Academic paper data	Evolutionary clustering, LSTM model	[29]
	Twitter data	Topic model, Dynamic parameter update strategy	[30]
	Weibo data	Topic model, Parameter estimating process	[31]
Incremental approach	Twitter data	Matrix decomposition method, Incremental clustering	[5]
	News data	Word embedding, Online clustering	[10]
	Twitter data	Text mining, Incremental clustering	[11]

3. Proposed Approach

In this study, an approach to detecting social issues that considers the definition and characteristics of the issues is proposed. Figure 1 shows the overall process of the proposed approach. First, web-based data is collected, and valid keywords are selected within the data through pre-processing. Second, by extracting co-occurrence information between the selected valid keywords, keyword co-occurrence networks are constructed by time

periods. Third, a network structure entropy is used to monitor the structural changes of the networks, which ultimately leads to the identification of a time period in which social issues occurred. Finally, social issue candidates are defined for the time period identified in the previous step, and the most urgent social issue is detected based on the priority of the social issue candidates.

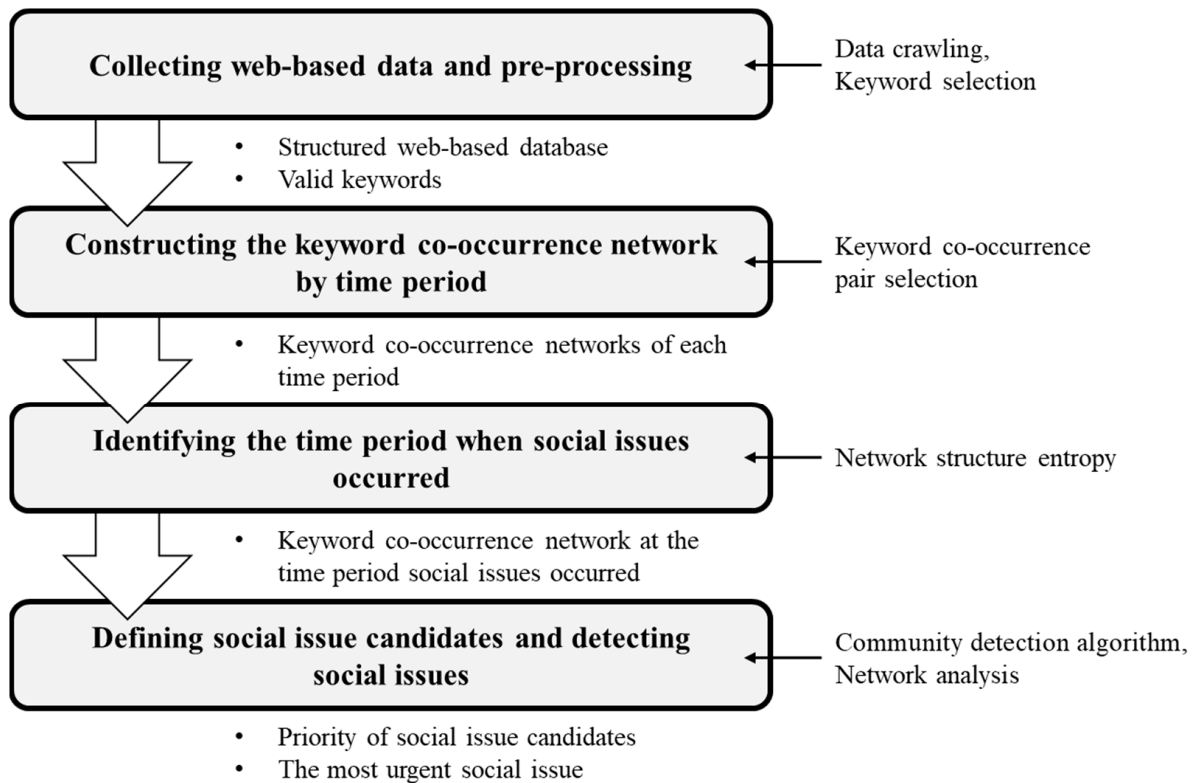


Figure 1. Overall process of the proposed approach.

3.1. Collecting Web-Based Data and Pre-Processing

Various open data generated in our society can be utilized to monitor the stream of society, from public data such as government reports to web-based data including web news, social media, and commercial reviews. This study utilizes web-based data among them. Since web-based data is created in real-time, it is relatively fast to update compared to other types of open data, and collection of the data is easy because the majority of web platforms offer open application programming interface (API) services [15]. Even if open API service does not exist, the data can be systematically collected through web crawling. Therefore, a vast number of stories about various phenomena occurring in our society can be continuously monitored by using web-based data.

In this regard, the text of web-based data, which was created from members of society, is collected, and keywords are extracted. Once a list of keywords is obtained from the text data, keyword selection should be conducted to filter out some keywords that are irrelevant or too generic for textual analysis. Specifically, stopwords, which are not appropriate for monitoring society, such as emoticons or onomatopoeic words, are excluded from the keyword list. As a result, valid keywords constituting the text of the web-based data are defined.

3.2. Constructing the Keyword Co-Occurrence Network by Time Period

In this step, after setting various time periods, a keyword co-occurrence network is constructed for each time period by utilizing the valid keywords defined in the previous step. Keywords extracted from web-based data are elements that construct various stories and topics in our society, so by utilizing the co-occurrence information between the

keywords, the stream of our society can be investigated. In addition, since an issue is a specific topic that suddenly attracts people's attention, in order to detect social issues, it is necessary to monitor when and what content members of society are paying attention to. Therefore, this study identifies co-occurrence pairs of keywords for each time period from the extracted valid keywords and constructs multiple networks, rather than constructing a single co-occurrence network from the entire data.

At this point, keyword co-occurrence pairs can be selectively used according to the characteristics of web-based data that were used to detect social issues. For example, for relatively short web documents, such as microblogs, all combinations of keywords that appear together in a document can be defined as keyword co-occurrence pairs. On the other hand, for long documents such as web news data, extracting all keyword co-occurrence pairs in a single document can result in a vast number of pairs, which can actually make it difficult to interpret the analysis results. So, in such cases, combinations of keywords that appear together in a paragraph or sentence, instead of a document, can be defined as keyword co-occurrence pairs to examine relationships between more relevant keywords. By the process described earlier, keyword co-occurrence pairs are identified for each time period considering the characteristics of the utilized data, and then keyword co-occurrence networks are constructed for each time period. The node of the constructed network is a keyword, the link between two nodes indicates whether the two keywords co-occur, and the weight of the link indicates the number of co-occurrences between the two keywords.

3.3. Identifying the Time Period When Social Issues Occurred

As mentioned in the previous step, in order to detect social issues, it necessary to keep an eye on when and what content people pay attention to. The aim of this step is to identify the time period in which social issues occurred among the various time periods defined above. Since this study presents various stories generated in our society by constructing keyword co-occurrence networks for each time period, structural changes of the keyword co-occurrence networks are monitored to identify the time period when social issues occurred. Specifically, since a phenomenon in which a link between specific keywords is suddenly formed in a network or the weight of the link is rapidly increased means that the structure of the network is rapidly changed; this study aims to identify a time period in which this phenomenon occurs.

This study utilizes network structure entropy based on information entropy concept to quantify and monitor the structure of a keyword co-occurrence network. The information entropy concept is used to represent the degree of order in a system by measuring the quantity and diversity of information that exists in the system [34]. An increase in information entropy means that the system is increasingly chaotic and disordered due to high-information diversity, while a decrease of the entropy means that the system is in order due to low information diversity. In this perspective, network structure entropy indicates whether the network is structurally ordered or disordered. If the structure entropy of the keyword co-occurrence network rapidly decreases in a specific time period, it means that a specific topic (keyword combination) is being mentioned a lot in the corresponding time period, even though various topics were mentioned evenly in the past. As shown in Table 3, the network structure entropy is calculated using the centrality indices of nodes and links constituting the network [14]. In equations of the table, the degree centrality of a node is the number of connections attached to other nodes, and the betweenness centrality of a node measures how many times the node appears in the shortest path between all pairs of nodes in the network [13]. The weight of a link is the number of co-occurrences of the two nodes that make up the link, and the betweenness centrality of a link is the frequency of the link that lies on the shortest paths between all pairs of nodes [35,36]. As a result, the time period in which the network structure entropy rapidly decreases is identified as the time period when social issues occurred.

Table 3. Equations of structure entropy.

Structure Entropy		Equation
Node	$N'_q = - \sum_{i=1}^N \frac{(p_i^{q_i}) - p_i}{1 - q_i + \epsilon}$	$p_i = \frac{DegreeCentrality(Node_x)}{\sum_{x=1}^N DegreeCentrality(Node_x)}$ $q_i = 1 - (\max(v(i)) - v(i)),$ $v(i) = BetweennessCentrality(Node_i),$ $\epsilon > 0$
Link	$L'_q = - \sum_{j=1}^N \frac{(p_j^{q_j}) - p_j}{1 - q_j + \epsilon}$	$p_j = \frac{Weight(Link_y)}{\sum_{y=1}^N Weight(Link_y)}$ $q_j = 1 - (\max(l(j)) - l(j)),$ $l(j) = BetweennessCentrality(Link_j),$ $\epsilon > 0$
Network		$S'_q = N'_q + L'_q$

3.4. Defining Social Issue Candidates and Detecting Social Issues

First, in this step, this study defines social issue candidates in the time period which was identified in the previous step as social issues occurred. At this point, the community detection algorithm is utilized for defining social issue candidates from the keyword co-occurrence network of the identified time period. A community in a network is a group of nodes that are densely connected to each other [37]. At the same time, the nodes that constitute the community are weakly connected to nodes of other communities (Figure 2). In the case of a keyword co-occurrence network, since a community is formed by gathering links between keywords that frequently appear simultaneously, the community means a topic composed of keywords related to each other. In this study, various communities were extracted by applying the community detection algorithm to the keyword co-occurrence network in the identified time period as social issues occurred. Since the extracted communities refer to topics mentioned by people in the time period in which social issues occurred, they can be defined as social issue candidates.

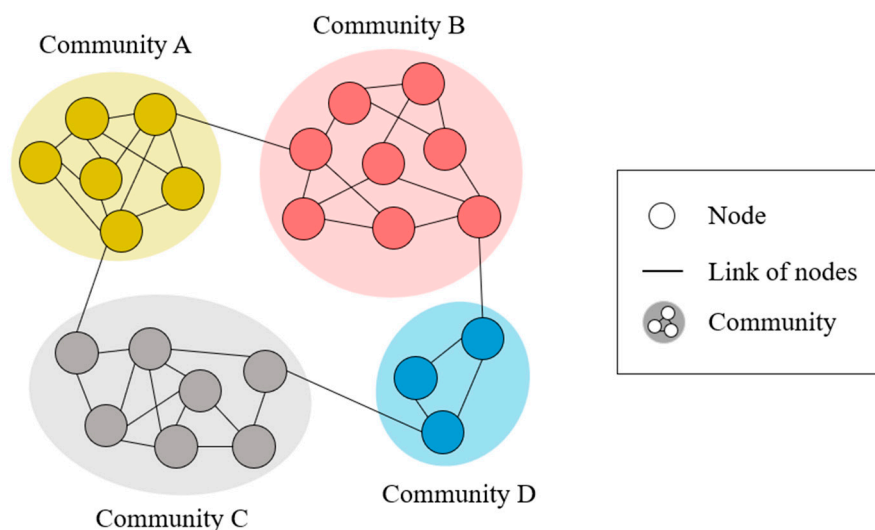


Figure 2. An example of a network with community structure.

Second, once the social issue candidates are defined, this study derives a priority of the social issue candidates to detect the most urgent social issues among them. The priority of the social issue candidates is derived using the centrality index of the keywords utilized in the previous step for calculating the network structure entropy. According to the equations of network structure entropy in Table 3, if the degree centrality of a specific keyword is

low and suddenly increases, the network structure entropy also changes rapidly under the influence of the centrality change of the keyword. Therefore, an impact score of each social issue candidate for the change in network structure entropy is computed using the degree centrality change of the keywords constituting a social issue candidate. The impact score of a social issue candidate c at time period t is defined as follows:

$$\text{ImpactScore}_t(c) = \frac{\sum_{i \in K_t(c)} f(i)}{\sum_{i \in K_t(c)} g(i)} \quad (1)$$

$$K_t(c) = \text{Set of keywords constituting community } c \text{ at time period } t \quad (2)$$

$$f(i) = \begin{cases} DC_t(i) - DC_{t-1}(i), & \text{if keyword } i \text{ exists at both time periods } t \text{ and } t-1 \\ 0, & \text{otherwise} \end{cases} \quad (3)$$

$$DC_t(i) = \text{Degree centrality of keyword } i \text{ at time period } t \quad (4)$$

$$g(i) = \begin{cases} 1, & \text{if keyword } i \text{ exists at both time periods } t \text{ and } t-1 \\ 0, & \text{otherwise} \end{cases} \quad (5)$$

Finally, since social issue candidates with high impact scores are the topics that members of society suddenly focused on, this study detects the most urgent social issues by identifying the candidates with high impact scores.

4. Illustrative Example: Diesel Exhaust Fluid Issue in South Korea

This section illustrates the social issue detection process of the proposed approach using web-based data. Specifically, to illustrate whether a social issue that occurred in the past is appropriately detected from web-based data generated at that time, this study selected the diesel exhaust fluid (DEF) issue in South Korea as a case of a social issue.

In late 2021, South Korea confronted an issue about shortage of DEF, which is an essential liquid used in modern diesel engines to reduce emissions of harmful nitrogen oxides. The issue was caused by several factors such as difficulties in importing DEF, a surge in demand due to stricter environmental regulations, supply chain disruptions caused by the COVID-19 pandemic, etc. According to the South Korean Ministry of Environment, there were approximately 9.8 million diesel vehicles registered in South Korea at the end of 2021. Among them, the vehicles requiring DEF were about 2.1 million, including passenger cars, cargo trucks, and public transportation buses [38]. Since vehicles that require DEF cannot operate without it, the DEF issue caused damages in various fields of society, such as trucks being unable to transport cargo and the commuting of workers being restricted. As a result, the DEF issue developed into the transport, logistics and public service crisis of South Korea. This section illustrates the process of detecting the DEF issue from web-based data generated at that time.

4.1. Experimental Results

This study utilized web news articles among web-based data to monitor social streams. Because a large volume of web news articles covering a wide range of topics, including economics, technology, and business, are generated in real time, web news data is suitable for monitoring our society over time [39].

This study collected Korean web news data through BIGKinds (www.bigkinds.or.kr (accessed on 20 July 2023)), which is a news big data analytics service run by the Korea Press Foundation. Since 1990, BIGKinds has accumulated news articles published by 54 Korean media organizations, including newspapers and broadcasting, and provides about 70 million articles. To evaluate the effectiveness of the proposed social issue detection approach in detecting the DEF issue that occurred in late 2021, this study first collected news arti-

cle in the late 2021 from BIGKinds. Specifically, since social issues appear without any restrictions, such as field or region, this study collected news article data from 11 national daily newspapers, excluding regional daily newspapers and specialty newspapers. As a result, information such as publication date, media outlet, and title were collected for 129,747 news articles from 1 September 2021 to 31 November 2021. In addition, since BIGKinds only provides partial text of news articles, this study collected the full text of the articles through web scraping.

To roughly identify when the social issue related to DEF occurred, this study measured the number of news articles in which the keyword, ‘요소수’ (Diesel exhaust fluid, DEF), was mentioned in the text by publication date, and found that the number increased sharply from the beginning of November 2021 (Figure 3). Therefore, out of the 129,747 news articles data collected earlier, 23,925 articles from all fields of society published between 25 October and 14 November were selectively used. Next, 234,975 unique keywords were extracted from the keyword data of the selected news articles. BIGKinds provides keyword data for each news article, which is a list of noun keywords that appear in the title and body of the article using a morpheme analyzer and a Korean dictionary. BIGKinds’ keyword data is suitable for keyword analysis, such as topic model analysis or network analysis [40,41]. Since there are many stopwords and meaningless keywords among the extracted keywords, pre-processing was performed with the following conditions (Table 4) to select 155,609 valid keywords for constructing keyword co-occurrence networks.

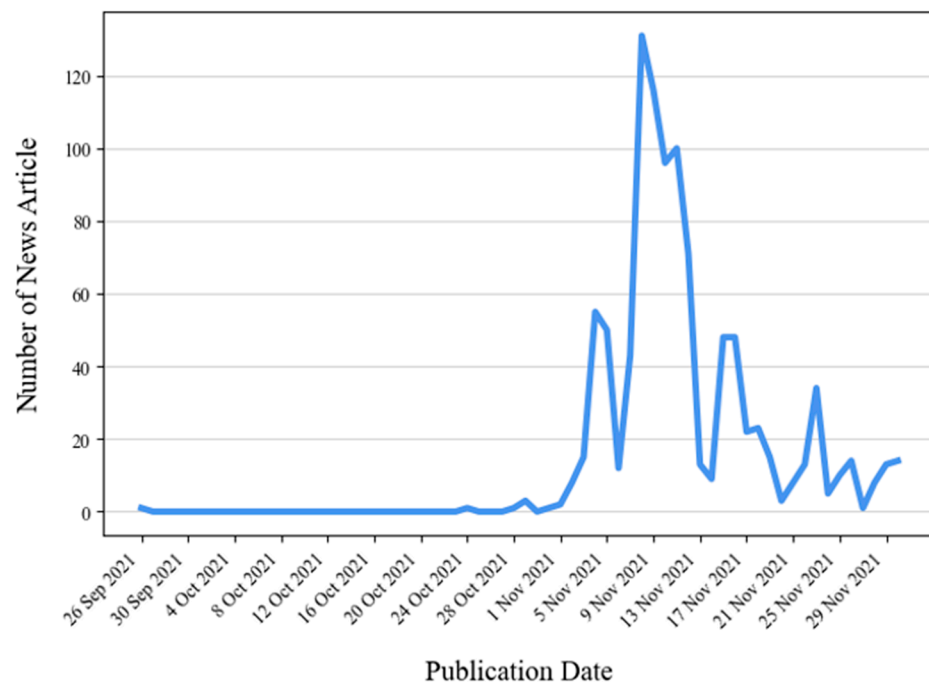


Figure 3. Number of news articles in which DEF was mentioned in the text.

Table 4. Stopword removal conditions for data pre-processing.

Condition	Examples of Stopword
Keywords with a length of 1	“A”, “B”, “끝” (Finish)
Keywords starting with a number	“1181조” (1181 trillion), “282.8%”, “632곳” (632 places)
Keywords in foreign languages other than Korean or English	“ざるうどん”, “水火”, “εκκλησία”
Keywords that are meaningless or have a broad meaning	“B씨” (Anonymous B), “내년” (Next year), “지난달” (Last month)

Before constructing the keyword co-occurrence network using the selected valid keywords, time periods were set for monitoring. When the number of news articles per day was computed based on the publication date, there were significantly fewer news articles on weekends compared to weekdays, as shown in Figure 4. Therefore, given the pattern of the number of news articles per day, this study set each time period to 7 days so that one time period always includes Saturday and Sunday. As a result, a total of eight time periods were defined from 25 October 2021 to 14 November 2021, with each time period overlapping (Table 5).

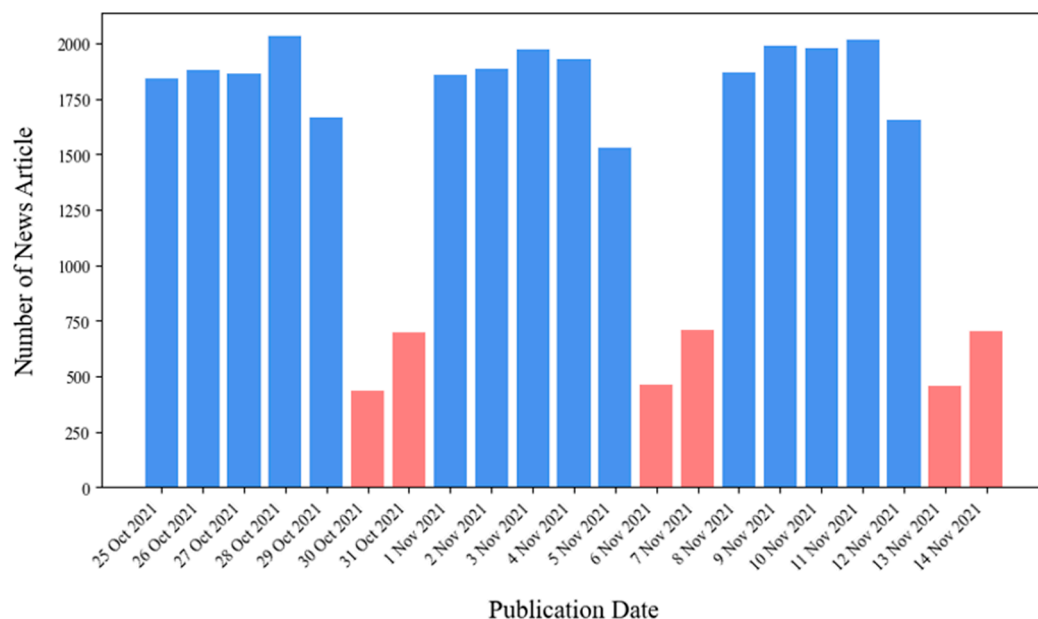


Figure 4. Number of news articles used in this study by publication date (Blue bars mean weekdays and red bars mean weekends).

Table 5. Number of selected keyword co-occurrence pairs for each time period.

Time Period	Number of Selected Keyword Co-Occurrence Pairs
25 Oct 2021~31 Oct 2021	27,387
27 Oct 2021~2 Nov 2021	25,796
29 Oct 2021~4 Nov 2021	25,564
31 Oct 2021~6 Nov 2021	25,779
2 Nov 2021~8 Nov 2021	26,143
4 Nov 2021~10 Nov 2021	26,533
6 Nov 2021~12 Nov 2021	27,081
8 Nov 2021~14 Nov 2021	27,065

Next, in order to construct a keyword co-occurrence network for each defined time period, this study derived keyword co-occurrence pairs using the previously selected valid keywords. Keyword data of BIGKinds provides an average of 99.2 keywords per one news article. If all keywords occurring together in a single news article were defined as co-occurrence pairs, the analysis process and interpretation of the results would be difficult due to the large amount of data. Therefore, based on the n-gram concept, which represents n consecutive words, this study defined co-occurrence pairs as keywords that appeared together within a certain distance (N), considering the existing position (Index) of the keywords within a sentence. The specific method is shown in Figure 5. First, the text of the news article is divided into sentences, the sentences are split by spaces, and the valid keywords selected from BIGKinds keyword data are matched. Next, among the matched keywords, two keywords with an index value difference of N or less are defined

as a co-occurrence pair. In this study, co-occurrence pairs were derived by setting N to 5 based on repeated experiments.

Example for the Korean version

Sentence	최근 국내에서 요소수 품귀 현상이 나타나고 있다.						
BIGKinds keyword data	요소수, 품귀, 현상						
Sentence split by spaces	최근	국내에서	요소수	품귀	현상이	나타나고	있다.
Index	0	1	2	3	4	5	6
Matched keywords	-	-	요소수	품귀	현상	-	-
Co-occurrence	N = 1	(요소수, 품귀), (품귀, 현상)					
pairs	N = 2	(요소수, 품귀), (요소수, 현상), (품귀, 현상)					

Example for the English version

Sentence	Recently, there has been a DEF shortage phenomenon in Korea.									
BIGKinds keyword data	DEF, shortage, phenomenon									
Sentence split by spaces	Recently,	there	has	been	a	DEF	shortage	phenomenon	in	Korea
Index	0	1	2	3	4	5	6	7	8	9
Matched keywords	-	-	-	-	-	DEF	shortage	phenomenon	-	-
Co-occurrence	N = 1	(DEF, shortage), (shortage, phenomenon)								
pairs	N = 2	(DEF, shortage), (DEF, phenomenon), (shortage, phenomenon)								

Figure 5. Defining keyword co-occurrence pairs.

In addition, this study applied the Pareto principle (also known as 80–20 rule) to selectively use defined keyword co-occurrence pairs. If a co-occurrence pair has too few co-occurrences, the pair is unlikely to be a meaningful co-occurrence. Therefore, this study selected co-occurrence pairs that were in the top 20% (3 or more) by number of co-occurrences within each time period. Finally, an average of 26,418.5 keyword co-occurrence pairs were selected for each time period (Table 5), and keyword co-occurrence networks were constructed for each time period by using the selected pairs.

This study identified the time period when social issues occurred by monitoring structural changes in the keyword co-occurrence networks. First, to quantify the structure of the keyword co-occurrence networks constructed earlier, this study calculated the structure entropy for each the network using the equations in Table 3. The structure entropy of the networks by time period is shown in Table 6.

Table 6. Network structure entropy for each time period.

Time Period	Node Structure Entropy	Link Structure Entropy	Network Structure Entropy
25 Oct 2021~31 Oct 2021	-9.1626	-11.9696	-21.1322
27 Oct 2021~2 Nov 2021	-9.2683	-11.9396	-21.2078
29 Oct 2021~4 Nov 2021	-9.1651	-11.9442	-21.1093
31 Oct 2021~6 Nov 2021	-9.0344	-11.9087	-20.9431
2 Nov 2021~8 Nov 2021	-9.3711	-11.9329	-21.3039
4 Nov 2021~10 Nov 2021	-9.3991	-11.9853	-21.3843
6 Nov 2021~12 Nov 2021	-9.3639	-11.9960	-21.3599
8 Nov 2021~14 Nov 2021	-9.3361	-11.9731	-21.3092

Next, the calculated network structure entropy was monitored to identify a time period in which the entropy value rapidly decreased. As explained in Section 3.3, a sharp decrease in the structure entropy of a network in a particular time period indicates that a particular topic in the network is mentioned a lot, especially in the corresponding time period. Therefore, this study monitored the change of network structure entropy over time by visualizing the value of entropy, and it was found that the entropy decreased rapidly in the fifth time period “2 Nov 2021~8 Nov 2021” (Figure 6). As a result, this study defined the identified time period “2 Nov 2021~8 Nov 2021” as the time period when the social issue occurred.

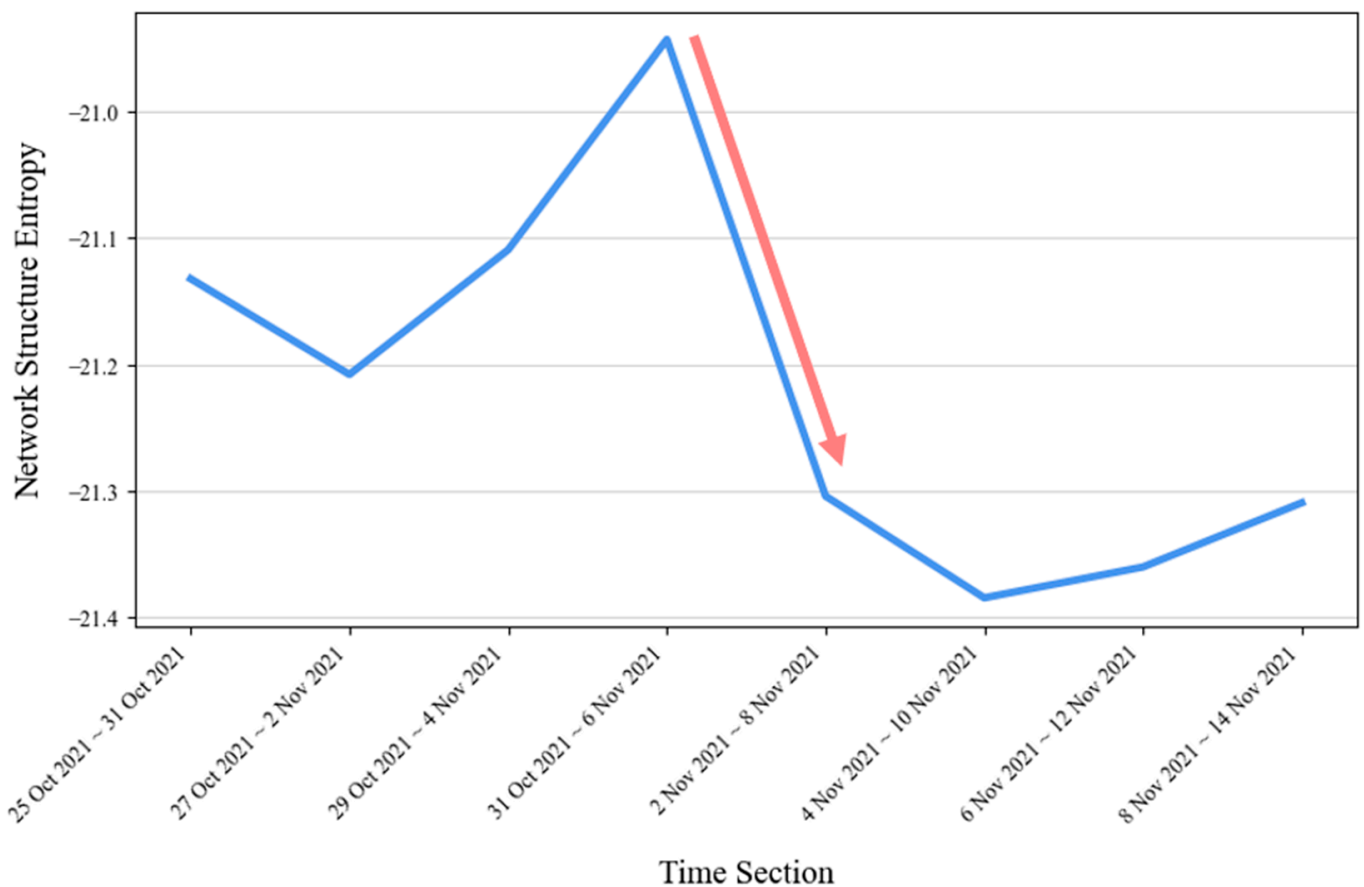


Figure 6. Monitoring network structure entropy over time (Red line represents the rapid decrease in entropy).

A community detection algorithm was applied to define social issue candidates from the keyword co-occurrence network in the time period (2 Nov 2021~8 Nov 2021) identified earlier as the point when social issues occurred. This study used a Louvain method among various algorithms that can detect communities in the network structure. The Louvain method is an algorithm that quickly and accurately detects high-modularity community structures from large networks [42]. The modularity is a measure of the density of links within a community, compared to links between communities [43]. The modularity of weighted networks is defined as Equation (6) [44].

$$Q = \frac{1}{2m} \sum_{i,j} \left[A_{ij} - \frac{k_i k_j}{2m} \right] \delta(c_i, c_j) \tag{6}$$

here, A_{ij} is the weight of the link between node i and node j , $k_i = \sum_j A_{ij}$ is the sum of the weights of the links incident to node i , c_i is the community to which node i is assigned, the function $\delta(u, v)$ is 1 if $u = v$ and 0 otherwise and $m = \frac{1}{2} \sum_{ij} A_{ij}$. The process of the Louvain method can be divided two steps. In the first step, each node in a network is initially assigned to its own community, and the change in modularity when a node i is removed from the original community and reassigned to the neighboring community C is measured (Equation (7)) [42].

$$\Delta Q = \left[\frac{\sum_{in} + 2k_{i, in}}{2m} - \left(\frac{\sum_{tot} + k_i}{2m} \right)^2 \right] - \left[\frac{\sum_{in}}{2m} - \left(\frac{\sum_{tot}}{2m} \right)^2 - \left(\frac{k_i}{2m} \right)^2 \right] \tag{7}$$

here, \sum_{in} is the sum of the weights of the links inside C , \sum_{tot} is the sum of the weights of the links incident to nodes in C , k_i is the sum of the weights of the links incident to node i , $k_{i,in}$ is the sum of the weights of the links from i to nodes in C and m is the sum of the weights of all links in the network. This step compares measurements; the node i moves, and places the node i in the community with the largest increase in modularity until the modularity change does not improve. And at the second step, the algorithm aggregates the nodes within each community and builds a new network where each community is represented as a single node. The links between the new nodes are weighted based on the sum of the weights of the links between nodes in the original communities. The Louvain method repeats the first step on the aggregated network, treating the communities as nodes. This step iteratively merges the most connected communities to optimize the modularity of the network at a global level. This process is repeated until no further improvement in modularity can be achieved.

Communities were extracted from the keyword co-occurrence network at the time period of the social issues' occurrence (Table 7), and the communities were defined as social issue candidates that occurred in the time period. Next, to detect the most pressing social issues among the defined social issue candidates, the impact score was calculated for each social issue candidate to derive a prioritization. A higher impact score means that the community has contributed more to the rapid change in network structural entropy. The social issue candidate with the highest impact score was community #86 (Tables 8 and 9). Therefore, community #86 was detected as the most urgent social issue (Figure 7).

Table 7. Community detection results for defining social issue candidates.

Number of Keywords in the Network	Number of Detected Communities	Statistics on the Number of Keywords for Each Community			
		Mean	Median	Max	Min
19,152	342	56	22	3341	10

Table 8. Top 3 social issue candidates with high impact score.

Community	Number of Keywords	Impact Score	Part of Keywords
#86	46	194.1	‘요소수’ (DEF), ‘고속버스’ (Highway bus), ‘디젤차량’ (Diesel vehicle), ‘화물차’ (Lorry), ‘천정부지’ (Sky-rocketing), ...
#42	13	81.1	‘내성균’ (Resistant bacteria), ‘비인체’ (Non-human), ‘세균’ (Bacteria), ‘카바페넴’ (Carbapenem), ‘항생제’ (Antibiotic), ...
#276	84	22.0	‘진단서’ (Medical certificate), ‘진통제’ (Painkiller), ‘처방전’ (Prescription), ‘치료목적’ (Therapeutic purpose), ‘펜타닐’ (Fentanyl), ...

Table 9. Impact score for community #86.

Keyword	Degree Centrality at 4th Time Period	Degree Centrality at 5th Time Period	Change in Degree Centrality	Impact Score
요소수 (DEF)	5116	11,039	5923	194.1
울주소방서 (Ulju fire station)	0	54	54	
주입 (Injection)	128	167	39	
홍정기 (Jeong-kee Hong)	0	34	34	
출입문 (Gate)	39	66	27	
완주 (Wanju)	42	66	24	
접수처 (Reception)	28	47	19	
아톤산업 (Aton Industry)	3	20	17	
소요량 (Requirement)	0	13	13	
기름 (Oil)	51	63	12	
익산 (Iksan)	24	34	10	
적재량 (Carrying capacity)	0	9	9	
화물차 (Lorry)	11	15	4	
화물연대 (Cargo Truckers Solidarity)	0	4	4	
...	
집중단속 (Crackdown)	30	7	-23	

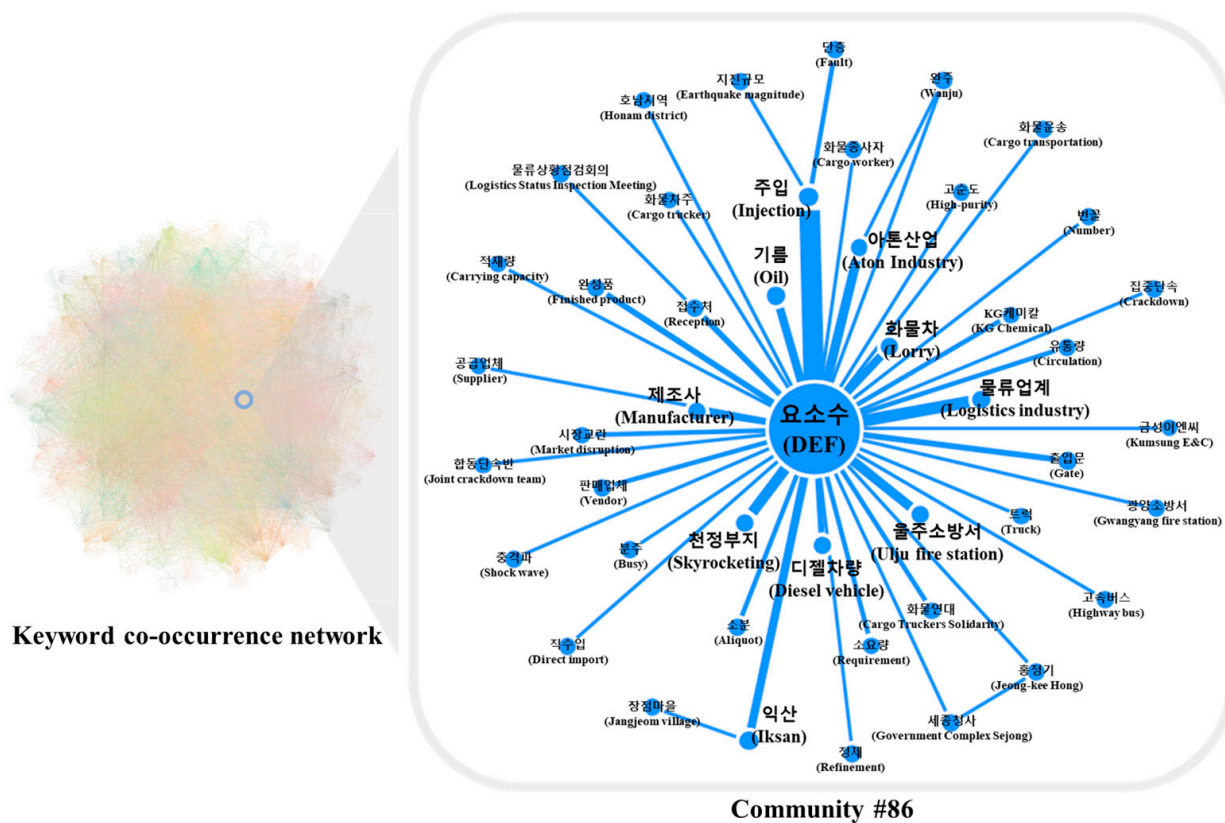


Figure 7. Most urgent social issue community #86.

4.2. Analyses

The community #86 consisted of 46 keywords, including the high-degree centrality keyword ‘요소수’ (DEF), and the keywords ‘고속버스’ (Highway bus), ‘디젤차량’ (Diesel vehicle), ‘화물차’ (Lorry), ‘천정부지’ (Skyrocketing), ‘시장교란’ (Market disruption). In fact, the transportation of many people and national industries was largely dependent on vehicles that require DEF injection, so shutting down vehicles such as highway buses and lorries, due to the shortage and skyrocketing prices of DEF, was the biggest problem with the issue. Among the keywords that make up community #86, the keywords that showed an increase in degree centrality compared to the previous time period were ‘요소수’ (DEF), ‘주입’ (Injection), ‘완주’ (Wanju), ‘아톤산업’ (Aton Industry), ‘익산’ (Iksan), etc. South Korea was experiencing a nationwide DEF shortage, causing the price of DEF to skyrocket, and threatening to shut down many of the vehicles around us, including fire trucks, delivery trucks, and agricultural equipment. At that time, the Aton Industry, a DEF manufacturing company based in Iksan-si, South Korea, signed a convention with local governments, including Wanju-gun, to prioritize the supply of DEF to the region. Even with a limited amount per day, Aton Industry sold DEF to local residents on a daily basis at a reasonable price until the DEF issue was stabilized. In addition, the companies involved in the convention were provided with sufficient quantities to ensure that their industrial activities were not constrained. Aton Industry was praised for its efforts to stabilize the local industry, even though it could have sold at a higher price and made a large profit during the DEF shortage. As a result, Aton Industry and the local governments, such as Iksan-si and Wanju-gun, attracted attention at the time.

The social issue candidates with the next highest impact score were community #276 and community #42. The main keywords in the community #276 were ‘내성균’ (Resistant bacteria), ‘비인체’ (Non-human), ‘세균’ (Bacteria), ‘카바페넴’ (Carbapenem) and ‘항생제’ (Antibiotic). The issue was related to the ‘Second National Antimicrobial Resistance Management Plan’, which was established by the South Korean Ministry of Health and Welfare

at the time to manage antibiotic resistance. Antimicrobial resistance refers to the ability of bacteria to resist and survive certain antibiotics. Antibiotic-resistant bacteria are generated and spread through multiple routes such as people, agriculture, livestock, fisheries, food, and the environment, so a governmental response is necessary. The development of antibiotic-resistant bacteria is closely associated with the use of antibiotics, and South Korea has been found to have the third highest human antibiotic use among OECD countries. Therefore, South Korea has established a new policy to ensure that the right amount and type of antibiotics are used to reduce the development of antibiotic-resistant bacteria.

The main keywords of community #42 were '진단서' (Medical certificate), '진통제' (Painkiller), '처방전' (Prescription), '치료목적' (Therapeutic purpose), and '펜타닐' (Fentanyl). In fact, at the time, doctors who overprescribed fentanyl, a narcotic painkiller, and young men in their 20s who prescribed and took fentanyl were arrested by the police and made headlines. Since then, drug-related crime has been on the rise in South Korea, and even today, the drug-related crime among young people remains a major social problem of South Korea.

This study detected the social issues that people paid attention to. If the most frequently appearing topics were identified based on the frequency as in previous studies, the topics would be serious social problems that most people already know about. In fact, the most mentioned keywords in the news articles used in this study were primarily related to the COVID-19 pandemic, including COVID-19, vaccine, and inoculation. Therefore, this study adequately detected social issue topics that attract people's attention among the various topics that exist in our society.

Additionally, this study intends to suggest ways to apply the proposed approach by using new keywords of the issue. Specifically, the change in degree centrality of the keywords that make up the community was used to calculate the impact score. Here, the degree centrality of newly introduced keywords that did not exist in the previous time period was not reflected in the impact score. Instead, it is possible to infer the near future development of a social issue through new keywords of the community detected as the social issue.

For example, in the case of the DEF issue discussed in Section 4.1, fire station-related keywords such as '울주소방서' (Ulju fire station) and '광양소방서' (Gwangyang fire station) were found among the new keywords that were not used in the impact score calculation process. From this, it was expected that the DEF issue would be linked to events related to fire stations. In fact, as the DEF shortage continued, not only transportation, but also fire trucks and ambulances, which are directly related to people's safety and life, were restricted. To improve this situation, many helpers began appearing across the country, such as anonymous donors who secretly placed DEF in front of the fire station and disappeared. In addition, among the new keywords, the keywords related to cargo transportation were also found, such as '화물연대' (Cargo Truckers Solidarity), '화물차주' (Cargo trucker), and '화물운송' (Cargo transportation). Cargo Truckers Solidarity is a labor union formed by workers in the cargo transportation industry in South Korea. As the DEF issue persists, the price of DEF continued to rise, making it harder to obtain. The Cargo Truckers Solidarity announced a strike, saying that all damages caused by operation suspension are being passed on to the workers. In fact, when the Cargo Truckers Solidarity went on strike twice in one year, it caused an economic loss of 10.4 trillion won, so strikes of solidarity like this case led to huge national economic losses in South Korea. As a result, by monitoring the new keywords in the DEF issue, it could be inferred that the DEF issue could develop into an incident related to Cargo Truckers Solidarity and cause additional economic losses.

5. Concluding Remarks

This study aimed to address the ongoing task of detecting and resolving social issues by proposing a quantitative approach based on web-based data. It focused on identifying the time period when social issues emerged and determining the most urgent issues that require immediate attention. By constructing keyword co-occurrence networks using

valid keywords extracted from web-based data, this study monitored our society by calculating the structural changes in the networks over time to identify the specific time period when social issues occurred. Subsequently, social issue candidates were defined, and their priorities were derived based on their impact on the network's structural changes. As a result, this study detected urgent social issues that our society should pay attention to and respond to first.

This study has major contributions in two aspects. First, from an academic perspective, this study contributes by proposing a new approach to detecting social issues based on the definition and characteristics of an issue, unlike previous studies that mainly focus on the frequency of certain topics to detect issues. Specifically, the proposed approach constructed keyword co-occurrence networks to represent various stories in our society. Next, the structural changes of the networks were monitored using the network structure entropy to capture moments when members of the society are focused on specific topics. As a result, the proposed approach quantitatively provides the time when social issues occurred by identifying the time period when the structure of the network changes rapidly. An issue is a topic that suddenly becomes of interest to people. Thus, the approach proposed in this study can suggest when social issues occurred by identifying when members of society focus on certain topics, which is a new issue detection approach based on the definition and characteristics of an issue.

Next, from a practical perspective, the proposed approach offers valuable support for timely responses to social issues by quantitatively measuring and suggesting their occurrence time period and priority. The prioritization of social issue candidates was not solely based on frequency, but rather on their impact score, which quantifies their influence on structural changes in the network. Candidates with a high impact score have contributed significantly to structural changes in the network, indicating their impact. With limited resources available to address the numerous social issues across various fields such as culture, economy, industry, and technology, it is crucial to efficiently address them in a sequential manner. The proposed approach can support decision-making for organizations, including governments and corporations, by providing priorities for social issue candidates at the time period when social issues occurred. This enables the efficient allocation of resources to address urgent social issues. Moreover, the approach can assist the establishment of future-oriented responses to social issues by providing insights into their developmental directions, as discussed in Section 4.2 of this study.

Despite the contributions, there are several areas for future research to consider. Firstly, although the study proposed a social issue detection approach using quantitative methods, it did not suggest a method for verifying the results of this study using quantitative indicators. Instead, based on the qualitative judgment of the authors, it is judged that the social issues that were of great social concern in South Korea at the time, including the DEF issue, were adequately detected. Future research should focus on conducting quantitative verification of the identified time periods and detected social issues. Moreover, it will be possible to compare issue detection performance with other studies through quantitative verification. Secondly, this study primarily focused on web news data from South Korea. To assess the usefulness of the proposed approach, conducting case studies in other countries is necessary. Additionally, exploring the detection of social issues by incorporating web-based data from multiple countries will help identify shared or global social issues. Lastly, although the study monitored keyword co-occurrence networks over time to identify specific time periods and detect urgent social issues, it did not capture the entire life cycle of social issues. The proposed approach has the advantage of being cost-effective because it does not extract and examine social issue candidates for each time period. Future research could be conducted to monitor the birth-to-death cycle of social issues while meeting the definition and characteristics of an issue. Addressing these aspects in future research will further enhance the robustness, usefulness, and comprehensiveness of the proposed approach for social issue detection.

Author Contributions: Conceptualization, S.L., J.L., J.-M.L., H.-W.C. and J.Y.; Methodology, S.L., J.L. and J.Y.; Software, S.L. and J.L.; Validation, J.Y.; Formal analysis, J.L.; Investigation, S.L., J.-M.L., H.-W.C. and J.Y.; Resources, J.-M.L. and H.-W.C.; Data curation, S.L.; Writing—original draft, S.L., J.L. and J.Y.; Writing—review and editing, J.-M.L., H.-W.C. and J.Y.; Visualization, S.L. and J.L.; Supervision, J.Y.; Project administration, S.L. and J.Y.; Funding acquisition, J.-M.L. and H.-W.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the Korea Research Institute of Science and Technology Information (KISTI Project No. K-23-L03-C02-S01); and the National Research Foundation of Korea (NRF) funded by the Ministry of Education (NRF-2021R1A2C1010027).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data used in this study are available from the authors.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Crable, R.E.; Vibbert, S.L. Managing issues and influencing public policy. *Public Relat. Rev.* **1985**, *11*, 3–16. [\[CrossRef\]](#)
2. Huang, C.C.; Liang, W.Y.; Lin, S.H.; Tseng, T.L.; Wang, Y.H.; Wu, K.H. Detection of Potential Controversial Issues for Social Sustainability: Case of Green Energy. *Sustainability* **2020**, *12*, 8057. [\[CrossRef\]](#)
3. Bigelow, B.; Fahey, L.; Mahon, J. A typology of issue evolution. *Bus. Soc.* **1993**, *32*, 18–29. [\[CrossRef\]](#)
4. Dougall, E. Revelations of an ecological perspective: Issues, inertia, and the public opinion environment of organizational populations. *Public Relat. Rev.* **2005**, *31*, 534–543. [\[CrossRef\]](#)
5. Erfanian, P.Y.; Cami, B.R.; Hassanpour, H. An evolutionary event detection model using the Matrix Decomposition Oriented Dirichlet Process. *Expert Syst. Appl.* **2022**, *189*, 116086. [\[CrossRef\]](#)
6. Guille, A.; Favre, C. Event detection, tracking, and visualization in twitter: A mention-anomaly-based approach. *Soc. Netw. Anal. Min.* **2015**, *5*, 18. [\[CrossRef\]](#)
7. Benhardus, J.; Kalita, J. Streaming trend detection in twitter. *Int. J. Web Based Communities* **2013**, *9*, 122–139. [\[CrossRef\]](#)
8. Qian, S.; Zhang, T.; Xu, C.; Shao, J. Multi-modal event topic model for social event analysis. *IEEE Trans. Multimed.* **2015**, *18*, 233–246. [\[CrossRef\]](#)
9. Capdevila, J.; Cerquides, J.; Torres, J. Mining urban events from the tweet stream through a probabilistic mixture model. *Data Min. Knowl. Discov.* **2018**, *32*, 764–786. [\[CrossRef\]](#)
10. Hu, L.; Zhang, B.; Hou, L.; Li, J. Adaptive online event detection in news streams. *Knowl.-Based Syst.* **2017**, *138*, 105–112. [\[CrossRef\]](#)
11. Hasan, M.; Orgun, M.A.; Schwitter, R. Real-time event detection from the Twitter data stream using the TwitterNews+ Framework. *Inf. Process. Manag.* **2019**, *56*, 1146–1165. [\[CrossRef\]](#)
12. Seymore, K.; Rosenfeld, R. Using story topics for language model adaptation. In Proceedings of the 1997 European Conference on Speech Communication and Technology, Rhodes, Greece, 22–25 September 1997; Volume 4, pp. 1987–1990.
13. Freeman, L.C. Centrality in social networks conceptual clarification. *Soc. Netw.* **1978**, *1*, 215–239. [\[CrossRef\]](#)
14. Xu, H.; Luo, R.; Winnink, J.; Wang, C.; Elahi, E. A methodology for identifying breakthrough topics using structural entropy. *Inf. Process. Manag.* **2022**, *59*, 102862. [\[CrossRef\]](#)
15. Choi, J.; Yoon, J.; Chung, J.; Coh, B.Y.; Lee, J.M. Social media analytics and business intelligence research: A systematic review. *Inf. Process. Manag.* **2020**, *57*, 102279. [\[CrossRef\]](#)
16. Kaplan, A.M.; Haenlein, M. Users of the world, unite! The challenges and opportunities of Social Media. *Bus. Horiz.* **2010**, *53*, 59–68. [\[CrossRef\]](#)
17. Chung, J.; Lee, J.; Yoon, J. Understanding music streaming services via text mining of online customer reviews. *Electron. Commer. Res. Appl.* **2022**, *53*, 101145. [\[CrossRef\]](#)
18. Yadegaridehkordi, E.; Nilashi, M.; Nasir, M.H.N.B.M.; Momtazi, S.; Samad, S.; Supriyanto, E.; Ghabban, F. Customers segmentation in eco-friendly hotels using multi-criteria and machine learning techniques. *Technol. Soc.* **2021**, *65*, 101528. [\[CrossRef\]](#)
19. Rathore, A.K.; Ilavarasan, P.V. Pre-and post-launch emotions in new product development: Insights from twitter analytics of three products. *Int. J. Inf. Manag.* **2020**, *50*, 111–127. [\[CrossRef\]](#)
20. Choi, J.; Oh, S.; Yoon, J.; Lee, J.M.; Coh, B.Y. Identification of time-evolving product opportunities via social media mining. *Technol. Forecast. Soc. Chang.* **2020**, *156*, 120045. [\[CrossRef\]](#)
21. Jeong, B.; Yoon, J.; Lee, J.-M. Social media mining for product planning: A product opportunity mining approach based on topic modeling and sentiment analysis. *Int. J. Inf. Manag.* **2019**, *48*, 280–290. [\[CrossRef\]](#)
22. Sayyadi, H.; Salehi, S.; AbolHassani, H. Survey on news mining tasks. In *Innovations and Advanced Techniques in Computer and Information Sciences and Engineering*; Springer: Dordrecht, The Netherlands, 2007.

23. Li, X.; Xie, Q.; Huang, L. Identifying the development trends of emerging technologies using patent analysis and web news data mining: The case of perovskite solar cell technology. *IEEE Trans. Eng. Manag.* **2019**, *69*, 2603–2618. [[CrossRef](#)]
24. Lee, K.; Kim, S.; Yoon, B. A systematic idea generation approach for developing a new technology: Application of a socio-technical transition system. *Technol. Forecast. Soc. Chang.* **2022**, *176*, 121431. [[CrossRef](#)]
25. Zuo, Y.; Zhao, J.; Xu, K. Word network topic model: A simple but general solution for short and imbalanced texts. *Knowl. Inf. Syst.* **2016**, *48*, 379–398. [[CrossRef](#)]
26. Budak, C.; Georgiou, T.; Agrawal, D.; El Abbadi, A. Geoscope: Online detection of geo-correlated information trends in social networks. *Proc. VLDB Endow.* **2013**, *7*, 229–240. [[CrossRef](#)]
27. Gaglio, S.; Re, G.L.; Morana, M. A framework for real-time Twitter data analysis. *Comput. Commun.* **2016**, *73*, 236–242. [[CrossRef](#)]
28. Blei, D.M.; Ng, A.Y.; Jordan, M.I. Latent dirichlet allocation. *J. Mach. Learn. Res.* **2003**, *3*, 993–1022.
29. Lu, Z.; Tan, H.; Li, W. An evolutionary context-aware sequential model for topic evolution of text stream. *Inf. Sci.* **2019**, *473*, 166–177. [[CrossRef](#)]
30. Chen, X.; Zhou, X.; Sellis, T.; Li, X. Social event detection with retweeting behavior correlation. *Expert Syst. Appl.* **2018**, *114*, 516–523. [[CrossRef](#)]
31. Li, J.; Wen, J.; Tai, Z.; Zhang, R.; Yu, W. Bursty event detection from microblog: A distributed and incremental approach. *Concurr. Comput. Pract. Exp.* **2016**, *28*, 3115–3130. [[CrossRef](#)]
32. Huang, G.; He, J.; Zhang, Y.; Zhou, W.; Liu, H.; Zhang, P.; Ding, Z.; You, Y.; Cao, J. Mining streams of short text for analysis of world-wide event evolutions. *World Wide Web* **2015**, *18*, 1201–1217. [[CrossRef](#)]
33. Zhou, X.; Chen, L. Event detection over twitter social media streams. *VLDB J.* **2014**, *23*, 381–400. [[CrossRef](#)]
34. Shannon, C.E.; Weaver, W. *A Mathematical Model of Communication*; University of Illinois Press: Urbana, IL, USA, 1949; Volume 11, pp. 11–20.
35. Anthonisse, J.M. The Rush in a Directed Graph. Stichting Mathematisch Centrum. Mathematische Besliskunde, 1971 (BN 9/71). Available online: <https://www.scinapse.io/papers/1513185775> (accessed on 20 July 2023).
36. Brandes, U. On variants of shortest-path betweenness centrality and their generic computation. *Soc. Netw.* **2008**, *30*, 136–145. [[CrossRef](#)]
37. Girvan, M.; Newman, M.E. Community structure in social and biological networks. *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 7821–7826. [[CrossRef](#)]
38. Korean Ministry of Environment, Ministry of Environment-Industry, Urgent Discussions to Normalize the Supply of Diesel Exhaust Fluid for Vehicles. 2021. Available online: <https://www.me.go.kr/tablet/file/readDownloadFile.do?fileId=225978&fileSeq=1> (accessed on 20 July 2023).
39. Yoon, J. Detecting weak signals for long-term business opportunities using text mining of Web news. *Expert Syst. Appl.* **2012**, *39*, 12543–12550. [[CrossRef](#)]
40. Lee, D.; Kwon, H. Keyword analysis of the mass media’s news articles on maker education in South Korea. *Int. J. Technol. Des. Educ.* **2022**, *32*, 333–353. [[CrossRef](#)]
41. Jo, W.; Chang, D. Political consequences of COVID-19 and media framing in South Korea. *Front. Public Health* **2020**, *8*, 425. [[CrossRef](#)]
42. Blondel, V.D.; Guillaume, J.L.; Lambiotte, R.; Lefebvre, E. Fast unfolding of communities in large networks. *J. Stat. Mech. Theory Exp.* **2008**, *2008*, P10008. [[CrossRef](#)]
43. Newman, M.E.; Girvan, M. Finding and evaluating community structure in networks. *Phys. Rev. E* **2004**, *69*, 026113. [[CrossRef](#)]
44. Newman, M.E. Analysis of weighted networks. *Phys. Rev. E* **2004**, *70*, 056131. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.