Telematics and Informatics xxx (xxxx) xxx

Contents lists available at ScienceDirect

Telematics and Informatics

journal homepage: www.elsevier.com/locate/tele

Identification of key cyberbullies: A text mining and social network analysis approach☆

Yoon-Jin Choi, Byeong-Jin Jeon, Hee-Woong Kim *

Graduate School of Information, Yonsei University, 50 Yonsei-Ro, Seodaemun-Gu, Seoul 03722, Korea

ARTICLE INFO

Keywords:
Cyberbullying
Cyberbully
Losada ratio
Cyberbullying index
Text mining
Social network analysis
Centrality

ABSTRACT

Cyberbullying is a major problem in society, and the damage it causes is becoming increasingly significant. Previous studies on cyberbullying focused on detecting and classifying malicious comments. However, our study focuses on a substantive alternative to block malicious comments via identifying key offenders through the application of methods of text mining and social network analysis (SNA). Thus, we propose a practical method of identifying social network users who make high rates of insulting comments and analyzing their resultant influence on the community. We select the Korean online community of Daum Agora to validate our proposed method. We collect over 650,000 posts and comments via web crawling. By applying a text mining method, we calculate the Losada ratio, a ratio of positive-to-negative comments. We then propose a cyberbullying index and calculate it based on text mining. By applying the SNA method, we analyze relationships among users so as to ascertain the influence that the core users have on the community. We validate the proposed method of identifying key cyberbullies through a real-world application and evaluations. The proposed method has implications for managing online communities and reducing cyberbullying.

1. Introduction

With the growth of information technology and the popularization of smartphones, criminal activities are quickly moving into cyberspace. Cyberbullying is an example that has caused serious damage all over the world (Camcho et al., 2018; Jang et al., 2016; Lee and Kim, 2015; Martinez et al., 2019; Peter and Petermann, 2018). Cyberbullying is a part of everyday life on social media (Lenhart et al., 2011). It has been related to several emotional, psychological, and physical problems (Hinduja and Patchin, 2007; Ybarra et al., 2006). The Cyberbullying Research Center surveyed a nationally representative sample of 5,000 middle and high school students ranging in age from 12 to 17 in the United States. The results showed that 37% of the students had experienced cyberbullying. The cyberbullying statistics issued in 2018 by the Korea Communications Commission and the National Information Society Agency revealed that 43.1% of adults had experienced cyberbullying (Choi et al., 2018). Cyberbullying can result in tragedy. In the United States, a 12-year-old girl committed suicide after being targeted by cyberbullies (The Guardian, 2013). Cyberbullying is becoming an increasingly serious problem in modern society.

Cyberbullying takes many forms (e.g., flaming, harassment, denigration, and impersonation) (Peter and Petermann, 2018; Willard, 2007). Among them, malicious comments designed to offend are the most typical. The damage caused by cyberbullying with malicious

☆ This work was supported by the Yonsei University Research Grant of 2020.

* Corresponding author.
E-mail addresses: yoonjin_choi@yonsei.ac.kr (Y.-J. Choi), fingeredman@yonsei.ac.kr (B.-J. Jeon), kimhw@yonsei.ac.kr (H.-W. Kim).

https://doi.org/10.1016/j.tele.2020.101504
Received 26 July 2019; Received in revised form 30 June 2020; Accepted 9 September 2020
Available online 19 September 2020
0736-5853/© 2020 Elsevier Ltd. All rights reserved.

comments is only worsening. In line with this increasing pattern, the 1% rule or principle of 90-9-1 (Van Mierlo, 2014) has been proposed: In cyberspace, 1% of users produce content, 9% deliver the content, and the remaining 90% just read the content. Although those persons who bully others via malicious comments constitute a small percentage of users, they have great influence on others and on the online space in terms of the spiral of silence theory (Noelle, 1977): Although many people with few opinions are silent, the opinions (i.e., cyberbullying with malicious comments) of a few active users can be taken for granted by many people. The broken windows theory (Wilson and Kelling 1982) further explains that if a car window is broken and left on the streets, it is perceived as a message that society's law and order are not maintained, leading to a greater crime. If cyberbullying with malicious comments from such a small number of key cyberbullies is neglected, it could expand to worse cyber violations later, despite the small number of users who started it.

Therefore, identifying key cyberbullies early is critical to minimizing or defeating cyberbullying. Identification of key cyberbullies is an important and urgent issue, but most global portal managers remain spectators. Google, Baidu, Bing, and MSN have eliminated comment space, or never allowed it. Yahoo has historically made comment histories unconditionally open to anyone who commented before. In the research area, little attention has been paid to identifying the worst offenders. Most research has focused on defining, realizing, and classifying cyberbullying problems with the antecedents (Law et al., 2012; Lee and Kim, 2015; Mishna et al., 2009; Slonje and Smith, 2008; Willard, 2007; Zhang and Leidner, 2018) or on detecting cyberbullying comments (Di Capua et al., 2016; Dadvar et al., 2012; Huang et al., 2014; Nahar et al., 2014; Reynolds et al., 2011; Sanchez and Kumar, 2011). Only a few researchers have focused on identifying key cyberbullies as a way to defeat them.

Our motivation for this study lies in the gap in the research on cyberbullying noted above and in the need to confront the critical problem cyberbullying poses. This study aims to develop a practical method of identifying key cyberbullies with high rates of insulting comments and analyze the influence they have in and on the community. To achieve our research objective, we apply the methods of text mining and social network analysis (SNA). Text mining allows us to analyze unstructured text documents to extract meaningful information (i.e., malicious comments) (Prollochs and Feuerriegel, 2019; Siering et al., 2018). The SNA method allows us to examine the relationship structure between entities (i.e., Cyberbullies and other online members) (Dhang-Pham et al., 2017; Karimi and Matous, 2018; Serrat, 2017). In pursuing our goal, we propose a cyberbullying index and the Losada ratio (Fredrickson and Losada, 2005). Both the Cyberbullying Index and Losada ratio are estimated based on the text mining results. We also apply the SNA method to analyze the relationships among users so as to ascertain the influence that the core users have on the community. The combination of the Losada ratio, cyberbullying index, and SNA permits identification of the key cyberbullies, i.e., those cyberbullies who are most influential.

We validate our proposed identification method through a real-world application to Daum Agora, the most representative and popular online discussion community in South Korea. There, 20% of all Internet users have been victims of malicious comments, and the corresponding damage has become a serious problem (Lee and Kim, 2015). We use a web crawler to collect data from over 650,000 posts and comments. To verify the performance of the key cyberbully classifier, we randomly selected 3,200 users, and four coders coded whether these users were core cyberbullies according to the criteria. We used nine test scenarios to verify the impact of each feature in the proposed model. Overall, this study contributes to the literature on cyberbullying by proposing and validating a method for identifying key cyberbullies.

2. Conceptual background

2.1. Previous research on cyberbullying and its detection

Cyberbullying is a form of violent expression conducted in cyberspace (Korea Communications Commission, 2018). It is an

Table 1
Previous Research on Cyberbullying and Its Detection.

Research	Representation	Objective	Textual Feat.	Social Feat.	User Feat.	Sentiment Feat.
Reynolds et al. (2011)	Number & density of "bad" words	Cyberbullying detection	1			
Dinakar et al., 2011	TF-IDF	Cyberbullying detection	1		1	
Dadvar et al. (2012)	TF-IDF	Cyberbullying detection	1		1	
Huang et al. (2014)		Cyberbullying detection	1	1		
Chavan and Shylaja (2015)	TF-IDF	Cyberbullying detection	1			
Van Hee et al. (2015)	Binary Bag-of-Words	Cyberbullying detection	1			1
Singh et al. (2016)	Probabilistic	Cyberbullying detection	1	1		
Al-Garadi et al. (2016)	TF-IDF	Cyberbullying detection	1	1	1	
Sugandhi et al., 2016	TF-IDF	Cyberbullying detection				1
Hosseinmardi et al. (2016)	Image + text (TF-IDF)	Cyberbullying detection	1			
Rosa et al. (2018b)	Word Embeddings	Cyberbullying detection				
Rosa et al. (2018a)	Fuzzy Fingerprints	Cyberbullying detection				
This study	**CBI + Losada ratio + SNA**	**Key Cyberbully detection**	**1 (CBI)**	**1 (SNA)**		**1 (Losada ratio)**

aggressive and intentional act that is carried out by a group or a single user on an ongoing basis, using electronic media, primarily against those who cannot easily defend themselves (Slonje and Smith, 2008). Previous studies focused on the definition and classification of cyberbullying (Law et al., 2012; Mishna et al., 2009; Peter and Petermann, 2018; Slonje and Smith, 2008; Willard, 2007). In addition, there have been comparative analyses with respect to traditional violence (Brown et al., 2014) and the identification of motives for cyber violence (Lee and Kim, 2015).

Practical means of detecting cyberbullying have been researched mainly in the field of computer science, as shown in Table 1. The goal of this research into detection is largely to improve the accuracy of classification by improving a model and by expanding the available analyzable features (e.g., user features, social features). Studies to improve analytical techniques are most often based on machine learning, focusing on supervised learning. However, it has also been extended to semi-supervised and unsupervised learning to overcome the limitations encountered. Supervised learning can be used to classify constructs such as cyberbullying by training a model based on labeled input data. The label distinguishes specific comments indicative of cyberbullying from those that are not. In our studies, human coders generated the label data (Dadvar et al., 2012; Reynolds et al., 2011; Sanchez and Kumar, 2011). However, in the real world, it is difficult to secure live data with labels. It is also difficult to capture all the various and varying contexts of cyberbullying afterward, even if a concerted effort goes into creating the labels. Therefore, semi-supervised learning techniques (Nahar et al., 2014) have been studied and extended to unsupervised learning (Di Capua et al., 2016) based on streaming data without labels. Nevertheless, in most cases, extant data are considered sufficient.

Studies have been undertaken to improve the performance of cyberbullying detection by adding context data. Sanchez and Kumar (2011) applied sentiment analysis to Twitter data; Dadvar et al. (2012) used user gender information; and Huang et al. (2014) improved the performance of the model by adding social data via SNA. Because most cyberbullying detection research uses supervised learning that is based on text mining, this additional analysis improved the accuracy of detecting cyberbullying. Researchers in the field of computer science have become interested, too, in improving the performance of detection methods. The main features used to detect cyberbullying include textual features, social features, user features, and sentiment features. Recently, word embedding has been attempted (Rosa et al., 2018b). The engineering of these various features has improved the automatic detection of cyberbullying. Improvement in model performances have been observed in studies that reflected the form of communication between users by applying SNA as a social feature (Singh et al., 2016; Huang et al., 2014). These detection models analyze comments to determine whether a communication constitutes cyberbullying. However, this approach requires additional steps to detect actual key cyberbullies via their comments. In summary, there has been little research into identifying cyberbullies. Our research goes beyond the previous studies and focuses on methods of identifying key cyberbullies as a way to make the Internet a safer and more pleasant place in cyberspace.

2.2. *Key cyberbullies with spiral of silence theory and broken window theory*

Noelle-Neumann's (1977) spiral of silence theory is one of the most influential recent theories of public opinion formation (Kennamer 1990: 393). This theory regards public opinion as a social control (Moy et al. 2001). The spiral of silence theory proposes that individuals who perceive their opinions as a majority will try to voice them. However, those who think their opinion is in the minority will tend to keep silent or go along with the majority (Liu and Fahmy 2011). This phenomenon is more severe in an online community environment than in general public opinion. The 1% rule or the principle of 90-9-1 (Van Mierlo, 2014) reinforces the spiral of silence theory.

Criminal justice scholars have proposed the "broken window" theory (Wilson and Kelling 1982). According to this theory, crime can be reduced by repairing broken windows, removing graffiti, and keeping the streets clean. An unrepaired window influences others to break another window by conveying that norms do not need to be followed. This theory is controversial because of the lack of empirical support (Skogan and Frydl, 2004; Skogan, 1992; Kelling and Coles, 1997; Kelling and Sousa, 2001; Blumstein et al., 2006), but Rudy Giuliani, former mayor of New York City, based his law enforcement policy on it. He instructed the police to strictly enforce the law on minor crimes (e.g., spitting, jaywalking, etc.). This presumably signaled that crime would no longer be tolerated and subsequent dramatic drops in the city's crime rate were attributed to this policy of law enforcement (Corriss 2010).

Therefore, identifying key cyberbullies is crucial to defeating cyberbullying from the perspective of the "broken window" theory, the spiral of silence theory, and the principle of 90-9-1. Neglect of malicious comments in cyberspace may lead to more and worse cyber violence later, even though it was initially perpetrated by a minority of online users. Early management is important because it can have a significant negative impact on the formation of cybercrime behavior. Just as proposed by the 90-9-1 rule, a very few key cyberbullies can have a significant impact.

3. Cyberbully identification approach

Among the various types of cyberbullying, we select as our research focus the identification of the authors of malicious comments. By focusing on the originators of malicious comments, we conceptualize cyberbullies as those persons who post serious malicious comments. Malicious comments rank high in severity and impact in cyberspace. Users who post them are cyberbullies who are likely to adversely affect cyberspace, according to the spiral of silence (Noelle-Neumann, 1977) and the broken window theories (Wilson and Kelling 1982).

We characterized key cyberbullies according to two attributes: Those who (1) post a high number of malicious comments and (2) have high impact on the cyber community. In defining them, it is important to consider not only their postings of malicious comments, but also their impact on others in cyberspace. Influential cyberbullies can silence other users, according to the spiral of silence theory,

and lower the psychological barriers to writing malicious comments, as the broken window theory contends, so that others are induced to write malicious comments. Cyberbullying by even a small number of key cyberbullies distorts the cyberspace environment and damages many people.

This study uses text mining and SNA methodology in identifying cyberbullies. Based on the results of text mining, we apply the Losada ratio (Fredrickson and Losada, 2005) and the cyberbullying index to estimate the seriousness of the online postings by each online member. Based on the results of SNA, we further estimate various centrality indices (degree centrality, betweenness centrality, closeness centrality, and PageRank) between online users to find the most influential cyberbullies. The most effective model in identifying key cyberbullies was proposed with selected variables (Losada ratio, CBI, SNA degree centrality), after evaluating the performance of 9 scenarios reflecting various features. The combination of the cyberbullying index and the results of the degree centrality enables us to identify key cyberbullies, and the Losada ratio results provide supplementary information.

3.1. Losada ratio

The Losada ratio, also known as the positivity ratio, is the sum of positivity divided by the sum of its negativity in a system (Fredrickson and Losada, 2005). With the Losada ratio (Equation (1)), when the ratio of positive-to-negative (P/N) emotions is 2.9 or higher, the likelihood increases that a user or community will flourish (Fredrickson and Losada, 2005). The proportion of P/N language can influence the success of an individual or group, suggesting that having more positive than pessimistic tendencies has a positive result. The Losada ratio judges the possibility of prosperity based on the P/N ratio and has been used in various fields, such as behavioral science, psychology, and marketing. Previous research revealed that a high Losada ratio is associated with a better team performance (Losada and Heaphy, 2004), better mental health (Diehl et al., 2011), and higher employee creativity (Rego et al., 2012). The ratio of 2.9:1 has been debated because it lacks theoretical and empirical justification (Brown et al., 2013; Nickerson, 2014). Despite the continuing debate over the appropriateness of the precise values, the concept of a proportional correlation with health and prosperity is widely accepted (Losada and Heaphy, 2004; Diehl et al., 2011; Regoet al., 2012).

$$\boldsymbol{LosadaRatio} = \frac{\boldsymbol{NumberofPositiveComments}}{\boldsymbol{NumberofNegativeComments}} \tag{1}$$

3.2. Cyberbullying index

Willard (2007) classified cyberbullying as flaming, harassing, denigration, and impersonation. Malicious comments are the most typical tactic. The damage caused by cyberbullying, especially by malicious comments, is worsening (Lee and Kim, 2015; Jang et al., 2016). To study this most prevalent and prominent problem, we focus on malicious comments in developing our cyberbullying index. Its development first requires confronting the difficult problem of determining what constitutes a malicious comment. Because the criteria are highly subjective and the meaning changes according to context, most of the preceding effort at classification has used data coded by humans to train machine learning to recognize these comments (Reynolds et al., 2011; Nahar et al., 2014; Di Capua et al., 2016; Sanchez and Kumar, 2011; Dadvar et al., 2012; Huang et al., 2014). However, in our effort to identify key cyberbullies, we define malicious comments as those that contain insulting words. Thus, we develop the cyberbullying index based on the ratio of insulting words that appear in the comments. Furthermore, we develop cyberbullying indices for use at both the user and community levels.

The index represents the percentage of malicious comments among all those a user makes. The user-level index is a numerical representation of the degree of cyberbullying a user perpetrates. The higher the index, the more frequently a user's comment contains insulting words. Equation (2) shows how the user-level index of cyberbullying is derived:$\boldsymbol{l}$ is the number of comments made by a user, and $\boldsymbol{S}$ is a function, the value of which varies according to whether a user's $_i$th comment contains insulting words.

The community-level version of the index represents the percentage of malicious comments among all comments in the community. It can also be read as the frequency of cyberbullying across the community. The formula is shown in Equation (3) in which $\boldsymbol{L}$ is the number of all comments created in the community, and $\boldsymbol{S}$ is a function whose value varies depending on whether the $_i$th article contains insulting words.

$$\boldsymbol{f(user)} = \frac{1}{\boldsymbol{l}} \sum_{i=1}^{\boldsymbol{l}} \boldsymbol{S(i)} = \begin{cases} 1, \boldsymbol{if slang belongs to the i - th comment} \\ 0, \boldsymbol{if slang not belongs to the i - th comment} \end{cases} \tag{2}$$

$$\boldsymbol{f(community)} = \frac{1}{\boldsymbol{L}} \sum_{i=1}^{\boldsymbol{L}} \boldsymbol{S(i)} = \begin{cases} 1, \boldsymbol{if slang belongs to the i - th comment} \\ 0, \boldsymbol{if slang not belongs to the i - th comment} \end{cases} \tag{3}$$

3.3. Social network analysis

SNA is an analytical technique that focuses on the structure of relationships on the assumption they are important. The topological position of users is a main concern in the tracing of the flow of information and human interaction and has been applied to various fields (Karimi and Matous, 2018; Serrat, 2017). Centrality indices are the most common measurements of the importance of users in a network (Aleahmad et al., 2016). A single entity (e.g., A person or a thing) is defined as a node and the relationship between two nodes is defined as an edge. Analysis of the characteristics of relationships proceeds through the calculation of centrality, cohesion, and the relationship between nodes and edges. With SNA, we can determine the most influential users (Freire et al., 2017).

We calculate various centrality indices—degree centrality, betweenness centrality, closeness centrality as well as PageRank—to measure the influence these users inflicted on the community. Degree centrality is the total number of links connected to a given node, indicating a higher degree for users who communicate with more people such as those on Daum Agora. Users attributed to a high degree of centrality are more influential (Yadav et al., 2018). Degree centrality consists of in-degree as an ingoing connection and out-degree as outgoing connection. In-degree centrality has been reported to be a good indicator of people's popularity in social networks (Golub and Van, 1996). Out-degree centrality is an indicator of the number of persons who one person extends to. It is also used as a key indicator in detecting cyberbullying (Huang et al., 2014). Closeness centrality measures how close a node is to all the other nodes in terms of average length of the shortest path between the node and all other nodes in a graph (Bavelas, 1950). A user with high closeness centrality is located at the shortest distance from other users; he or she can be said to be the fastest user to spread the message to other users (Beauchamp, 1965). Betweenness centrality measures the number of times a node acts as a bridge along the shortest path between two other nodes (Freemann, 1977). Users with high betweenness centrality are likely to play an intermediate role in message propagation (Kiss & Bichler, 2008). PageRank was developed by Brin and Page (1998), the founders of Google. It is transmitted from the source page to the link target, and its value depends on the PageRank of the source page. So, a link from a page with a high PageRank contributes more than a link from a page with a low PageRank. It is a method of weighting a document having a hyperlink structure such as the World Wide Web according to its relative importance. This algorithm can be applied to any batch of references and to references linked together.

3.4. Proposed identification method

Fig. 1 shows the overall procedure for identifying key cyberbullies. Based on the use of our method of text mining, we calculate the Losada ratio through a sentiment analysis of comments and identify the degree of cyberbullying by developing and calculating the cyberbullying index by compiling a dictionary of insulting words. Losada ratios and cyberbullying indices are calculated at the community and user levels, respectively. Based on the application of SNA, we then analyze centrality to identify key users with strong, active relationships with other users in the community. Users with a low Losada ratio, a high cyberbullying index, and a high centrality is judged key cyberbullies because the percentage of negative comments is high, the percentage of insulting words are high, and communication within the community is strong.

Table 2 summarizes the concepts, application methods, and constraints of each method. Each has its own usage; however, there are constraints when using them alone. Using all three together makes it practical to efficiently detect key cyberbullies. However, the Losada ratio has a constraint when a specific user has zero positive comments: The numerator becomes zero in the P/N, making calculation impossible.

For users who comment infrequently, the cyberbullying index becomes 1 for all comments, including insulting words, when there are only one or two comments and all contain insulting words. This type of user is the most dangerous kind of cyberbullying suspect in the community. Unfortunately, the cyberbullying index cannot reflect the influence of such a user. Thus, various SNA centrality indices were used and evaluated. A user with high centrality is a highly influential user who affects many people through his or her comments. If a user has a high cyberbullying index, a low Losada ratio, and a high centrality simultaneously, he or she is identified as a key cyberbully.

Once the data of the cyberbully suspects were extracted, we synthetically and systematically calculated the degree centrality around the cyberbully suspects and identified those deemed to be key cyberbullies.

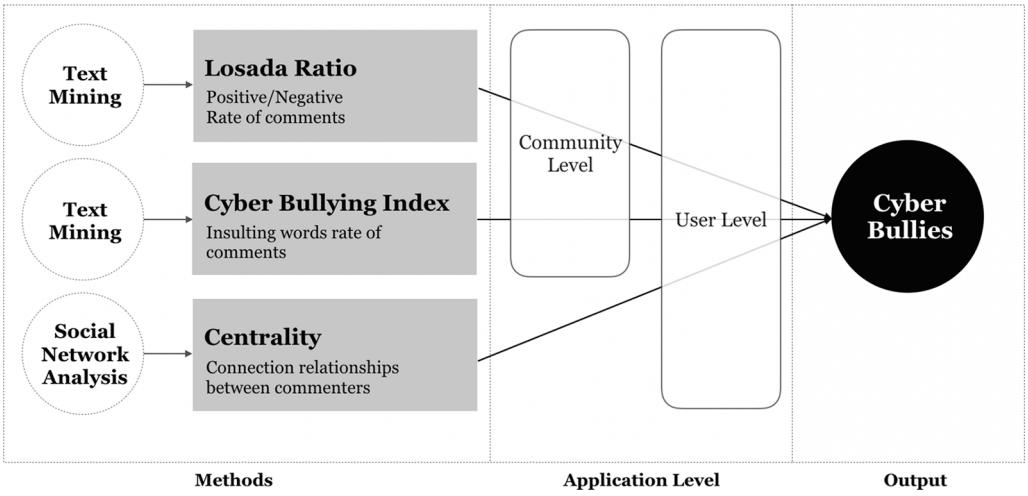

Fig. 1. Cyberbully Identification Approach.

ARTICLE IN PRESS

Table 2
Characteristics of the Proposed Elements.

Element	Concept	Usage	Constraint
Losada ratio	– P/N rate of comments (dictionary-based sentiment analysis) – The higher the positive rate (i.e., 2.9 or above) the more likely individuals or groups are to flourish (Fredrickson and Losada, 2005) – $\boldsymbol{LosadaRatio} = \frac{\boldsymbol{NumberofPositiveComments}}{\boldsymbol{NumberofNegativeComments}}$	Problematic user and community detection	– Losada ratio cannot be calculated if there are no negative comments
Cyber-bullying Index	– Rate of insulting words in comments (dictionary of insulting words) – User/Community-level ratio of malicious comments among all comments $f(\boldsymbol{user}) = \frac{1}{l}\sum_{i=1}^{l} \boldsymbol{S(i)} f(\boldsymbol{community}) = \frac{1}{L}\sum_{i=1}^{L} \boldsymbol{S(i)}$	Problematic user and community detection	– Even if the frequency of commenting is low, when all the written comments contain insulting words, that user is judged to be the most dangerous user – When the cyberbullying index is 1, it cannot reflect the degree of influence in the community.
SNA	– Degree centrality indicates the total number of users a specific user has communicated with Closeness centrality indicates how close a node is to all the other nodes – Betweenness centrality indicates the number of times a node acts as a bridge along the shortest path between two other nodes – PageRank indicates how to weight documents with hyperlink structure according to their relative importance	Identify the key users in the community. (Judged to be influential)	– Cyberbully identification should be considered with Losada ratio and cyberbullying index

4. Application of the proposed method

To validate the proposed method, we applied it to a real case. Fig. 2 shows the procedure in applying the proposed method. We selected Daum Agora in Korea for our trial. Daum Agora is a large-scale online discussion community based on the concept of the agora in ancient Greece. In Daum Agora, online users from diverse perspectives gather to exchange ideas and collaborate. It is no surprise that malicious comments are a serious problem in the Daum Agora (Yoo, 2010) because it is especially characterized by active political debate in which online users not only share their opinions but also file an online petition (Yoo, 2010). Daum Agora has 11 discussion communities. We selected seven (Politics, Economy, Real Estate, Stock/Fund, Society, Education, and Culture/Entertainment) in which the topics of discussion are clear and discussion is active. Some discussions are strongly offensive and manifest as cyberbullying.

We used a web crawler to collect data and compiled a dictionary of Korean insulting words. In compiling it, we focus especially on incorporating the latest insulting words from current slang. For preprocessing, we conducted morphological analysis (part-of-speech tagging), classified malicious comments using the said dictionary, and conducted sentiment analysis to divide sentiments into "positive," "negative," and "neutral" categories. During the analysis stage, the cyberbullying index was calculated; the Losada ratio was used; and the SNA was conducted. We extracted the most identifiable features of key cyberbullies among textual features (Cyberbullying Index), sentimental features (Losada ratio), and social features (centrality indices) derived during the analysis phase. Multiplying the variables derived from the analytical results of this evaluation, we created a cyberbullying score that could be used to identify key cyberbullies.

4.1. Data collection and preprocessing

We collected 139,849 posts and 513,763 comments, all of them written between April 6, 2015, and April 4, 2018. We collected the same number of postings for each community and extracted 31,650 users. We used a crawling technique to collect data. To ensure the privacy of users' data, we followed every protocol of the ethical guidelines outlined by the Association of Internet Researchers (AoIR) (Franzke et al., 2020). To ensure users' anonymity, we excluded all personal identifying information except the users' ID, which are pseudonyms unconnected to an actual person. The descriptive statistics for morphological analysis of the posts and comments are shown in Table 3. When comparing the numbers of comments in each community, the Economy and Real Estate communities had the largest number of comments in comparison with the number of posts. The Stock/Fund and Education communities had fewer comments when compared with others.

We conducted additional preprocessing to calculate the Losada ratio. We conducted a sentiment analysis using KoALA (Jeon et al, 2019). Based on the sentiment scoring results, we classified the comments as "positive," "negative," and "neutral." A score greater than 0 meant the comment was classified as positive, whereas one less than 0 meant that the comment was classified as negative. A score of 0 meant that the comment was neutral. Regardless of the scoring, comments that included insulting words were classified as negative.

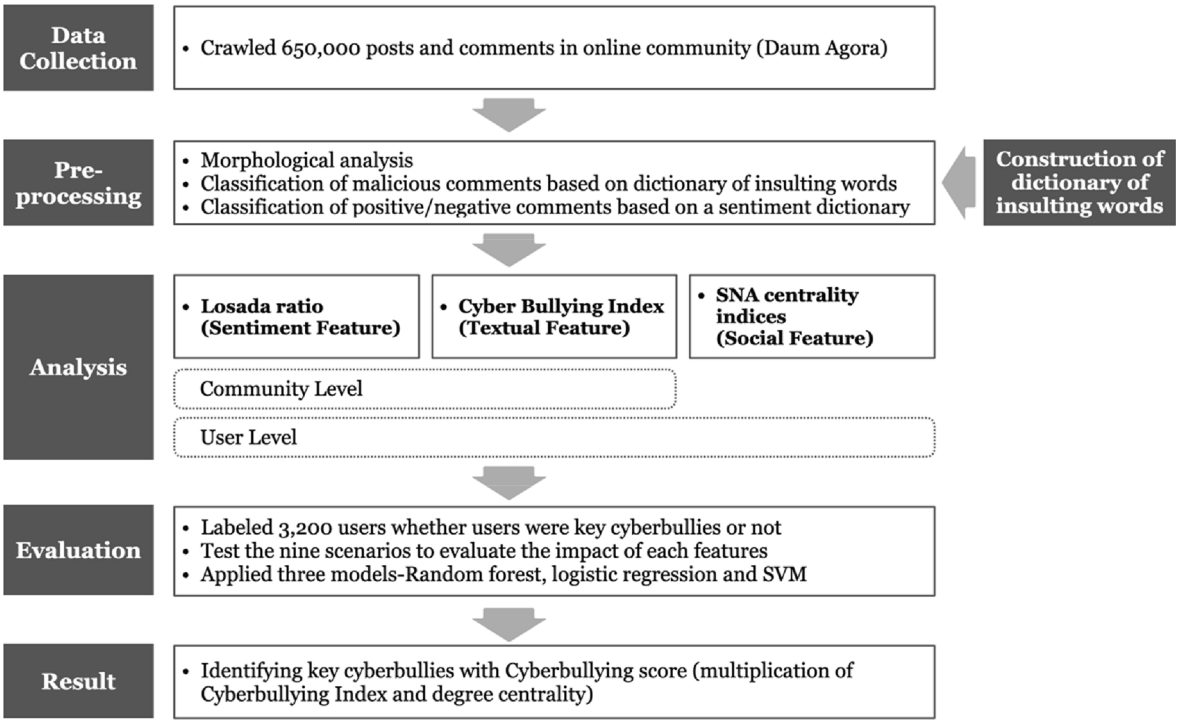

Fig. 2. Application Procedure.

Table 3
Descriptive Statistics of Daum Agora Data.

Community	Type	Count	Total Token	Noun Token
Politics	Posts	19,980	3,030,829	1,305,912
	Comments	63,098	789,379	411,466
Economy	Posts	20,000	2,538,244	1,188,445
	Comments	142,061	2,156,025	1,251,614
Real estate	Posts	19,983	9,187,575	3,013,662
	Comments	115,761	2,187,976	1,264,305
Stock/Fund	Posts	19,980	13,093,060	5,634,366
	Comments	9,531	216,251	115,771
Society	Posts	19,980	10,417,372	5,644,253
	Comments	34,081	801,976	441,068
Education	Posts	19,980	5,756,127	2,168,716
	Comments	59,868	709,711	304,650
Culture/ Entertainment	Posts	19,946	7,281,688	3,720,209
	Comments	89,363	1,471,356	515,918
Total	Posts	139,849	51,304,895	22,675,563
	Comments	513,763	8,332,674	4,304,792
	Sum	653,612	59,637,569	26,980,355

We also conducted malicious comment classification as additional preprocessing for calculation of the cyberbullying index. We classified the comments based on our dictionary of insulting words. If a comment included words found in it, it was classified as malicious.

For SNA, the nodes were Daum Agora users, and the edges were relationships between IDs (i.e., users), such as commented replies to posts. Commenting on posts was classified as a directed network. SNA centrality indices such as degree centrality, closeness centrality, betweenness centrality, and PageRanks were evaluated, and the weight of the network was not considered. We made a one-mode network that considers individual users equally in order to facilitate the main purpose of this study which is to identify key cyberbullies.

4.2. Compilation of dictionary of Korean insulting words

We define "insulting words" as terms used to worsen a listener's (reader's) mood or to be offensive. Because of the nature of the language, insulting words in Korea are highly variable and thus difficult to collect. Dictionaries that purportedly have such compilations are difficult to apply in our study because they do not include the latest insulting words or such online-derived words and phrases. Therefore, we have compiled our own dictionary of insulting words to use to classify malicious comments. We collected banned search keywords from other representative services (Naver's blog, South Korea's No. 1 portal, café, P2P services) and extracted words frequently used in conjunction with comments to give low ratings to movies. To complete the selection of insulting words for inclusion in the dictionary, we used a method used in open coding in qualitative studies (Corbin and Strauss, 2008). That is, three researchers conducted a peer review of the compilation of insulting words. Two coders then decided whether each of the words in the compilation warranted inclusion in the dictionary of insulting words. The inter-rater agreement scores averaged 0.92, and Cohen's Kappa scores averaged 0.91 (Cohen et al., 2000). Inter-rater disagreements were reconciled through discussions with a separate coder. The resulting list of 907 insulting words was thus established through a consensus of the judges. Table 4 shows a list of some insulting words that appear frequently in Daum Agora data.

4.3. Data analysis

We conducted three sequential analyses, all using the preprocessed data. First, the Losada ratio and Cyberbullying Index were calculated at the community level. Through this, it was possible to observe, at a group level, which communities were problematic and noteworthy. The Losada ratio at the community level calculated the P/N ratio according to the sentiment analysis of all comments in the community. Fredrickson and Losada (2005) found that if the P/N ratio was 2.9:1 (or higher), the group was likelier to thrive. It is possible to identify the problematic communities by comparing their P/N scores. The community-level cyberbullying index calculates the P/N ratio of insulting words for each community. Comparisons with other communities can determine which community is

Table 4
Insulting Words That Appeared Frequently in Daum Agora Data.

Ranking	Insulting Word	Frequency	Ranking	Insulting Word	Frequency	Ranking	Insulting Word	Frequency
1	존나	338,906	6	새끼	81,884	11	시바	27,703
2	시발	210,709	7	졸라	52,370	12	좆	26,876
3	미친	152,847	8	개웃겨	32,664	13	병신	21,370
4	씨발	96,255	9	새끼들	30,362	14	좃	19,035
5	ㅅㅂ	87,161	10	지랄	29,712	15	또라이	15,597

ARTICLE IN PRESS

Table 5
Community Level of Losada Ratios and Cyberbullying Indices.

Community	Total Number of Comments	Number of Positive Comments	Number of Negative Comments	Number of Neutral Comments	Number of Malicious Comments	Losada Ratio	Cyber-bullying Index
Politics	41,268	11,084	14,930	15,254	5,277	0.74 : 1	0.13
Economy	119,150	37,891	36,942	44,317	12,909	1.03 : 1	0.11
Real estate	86,042	30,435	28,160	27,447	9,410	1.08 : 1	0.11
Stock/Fund	8,148	2,915	2,754	2,479	757	1.06 : 1	0.09
Society	16,092	5,880	5,937	4,275	1,830	0.99 : 1	0.11
Education	48,172	17,636	12,813	17,723	4,095	1.38 : 1	0.09
Culture/ Entertainment	15,463	5,616	4,332	5,515	1,503	1.30 : 1	0.10
Total	334,335	111,457	105,868	117,010	35,781	1.05 : 1	0.11

problematic based on the percentage of insulting words used.

At the individual level, the Losada ratio and the Cyberbullying Index were calculated to determine users with high potential problems. Next, various centrality indices were calculated to reflect the user's influence in the community. The Losada ratio and the P/N language ratio of an individual was calculated, suggesting that having more positive than pessimistic tendencies had a positive result. The user-level cyberbullying index calculates the ratio of insulting comments in an individual's comment. It is used to detect problematic users with a high rate of insulting comments. Users with a low Losada ratio, a high cyberbullying index, and a high centrality can be considered key cyberbullies because the rates of negativity and insulting words are high and the user has a significant impact on the community.

5. Application results and evaluation

5.1. Community Level-Losada ratio and cyberbullying index

We conducted two analyses to determine the health of a community and the severity of the cyberbullying in it. Table 5 shows the analytical results for each community. Comments are in the order of Economy, Real Estate, and Education. The Politics community had the worst Losada ratio (0.74:1) and cyberbullying index (0.13). Although this community was only fourth in number of articles, its rate of negative and malicious comments was the highest among all the communities.

Daum Agora had almost a 1:1 Losada ratio in almost all communities. Given that the criterion for a healthy community is 2.9:1, Daum Agora has a markedly high percentage of negative comments. In the case of Politics, the ratio of 0.74:1 is significantly higher than in other communities.

The cyberbullying indices of the seven communities fell between 0.09 and 0.13. The Politics community scored highest at 0.13, meaning that 13% of the comments in the community were malicious. Politics had the highest cyberbullying index because of the frequent slander. Excluding Politics, almost all communities had a cyberbullying index of approximately 0.1, which means one out of every 10 comments in Daum Agora consists of malicious remarks containing insulting words.

Daum Agora is an unhealthy community with a Losada ratio of 1:1, meaning that negative comments account for half of all content. The cyberbullying index at the community level for each of the communities is also around 0.1, meaning that the percentage of malicious comments are quite high. The Politics community's low Losada ratio and high cyberbullying index suggest that special intervention may be needed. We can measure the health status of Daum Agora's seven communities, but finding the key cyberbullies and taking appropriate action can make the community healthier. We need to identify key cyberbullies to defeat the status quo. In the next section, we discuss changing the unit of analysis from the community level to the individual level.

5.2. User Level-Losada ratio and cyberbullying index

Based on the comments collected from Daum Agora, we extracted a total of 6,476 users who had exchanged comments with 31,650 users. We calculated the Losada ratio and the cyberbullying index for each user. By analyzing the cyberbullying index and the Losada ratio at the user level, it was possible to measure the degree of violence and health of each user.

Three hundred eighty-one users had cyberbullying indices of 1.0. In most cases, users posted only one or two comments, but all included insulting words. Therefore, to measure the influence of higher levels of activity in the community, we zeroed in (by filtering) on users who had commented more than 23 times (average frequency of all users was 22.9).

Table 6
User Level of Losada Ratio and Cyberbullying Index.

Rank[1)]	ID	Total Number of Comments	Number of Positive Comments	Number of Negative Comments	Number of Neutral Comments	Losada Ratio	Cyber-Bullying Index [2)]
1	User 1	40	0	40	39	–	0.98
2	User 2	30	0	30	28	–	0.93
3	User 3	52	0	52	48	–	0.92
4	User 4	34	5	29	29	0.17:1	0.85
5	User 5	165	0	165	134	–	0.81
6	User 6	26	1	23	20	0.04:1	0.77
7	User 7	38	6	29	25	0.02:1	0.66
8	User 8	35	5	27	3	0.19:1	0.66
…	…	…	…	…	…	…	…
11	**User 9**	**313**	**45**	**240**	**28**	**0.19:1**	**0.64**
12	User 10	42	3	31	8	0.10:1	0.60
13	**User 11**	**218**	**20**	**180**	**18**	**0.11:1**	**0.58**
…	…	…	…	…	…	…	…
19	**User 12**	**1,337**	**191**	**1,070**	**613**	**0.17:1**	**0.46**

[1)] *Top 19 users were selected based on results from the Cyberbullying Index*
[2)] *Mean = 0.12, Standard deviation = 0.14 for all users*

Table 6 shows the results of the Losada ratios and the cyberbullying indices of the top users. The Losada ratio ranged from 0.02 to 0.24, showing an unhealthy status. It is far lower than 2.9, which is the criterion for thriving. The higher the cyberbullying index, the higher the percentage of users whose comments included insulting words. The score tends to increase sharply in the ascending order of rankings. Although we filtered on 23 or more comments, the users with cyberbullying indices of 0.9 or higher were ranked among the top three. At the user level, these users were considered very violent. The top three users had not made a single positive comment. Thus, they cannot be assigned a Losada ratio.

Although the Losada ratio and cyberbullying index can be used to identify suspected individual cyberbullies, they cannot be used together or alone to identify key cyberbullies because they cannot reflect a user's influence on the community. To assess an individual user's influence, we applied degree centrality.

5.3. User Level-Social network analysis

The User 1 scored highest on the cyberbullying index (see Table 6). Of his 40 comments, 39 were malicious, and his cyberbullying index was 0.98. In the case of User 12, her cyberbullying index was 0.46, which meant she ranked 19th. Her cyberbullying index was lower than that of User 1, but in terms of commenting frequency, User 12 wrote 1,337 comments, meaning that she was much more influential in the community than User 1, who wrote only 40 comments.

We conducted SNA to overcome the constraints of the Losada ratio and the cyberbullying index and to detect key cyberbullies. We analyzed various centrality indices to see the characteristics of the relationships between the nodes and the edges and calculated degree centrality around various nodes. Degree centrality, closeness centrality, betweenness centrality and PageRanks were calculated in order to identify the most effective model. Considering relationships among users, rather than simple frequencies, centrality is an effective barometer to evaluate the influence of users in the community. In particular, filtering was performed based on degree centrality to reflect the user's influence in the community, which simple frequency cannot take into account. Through this filtering, the quality of communication was reflected in the analysis.

Degree centrality is the sum of in-degree centrality and out-degree centrality. Therefore, when the $\frac{In-degreecentrality}{Degreecentrality}$ is close to 1, the user rarely leaves a comment on another person's post, and only receives a comment from other users. Conversely, users with $\frac{Out-degreecentrality}{Degreecentrality}$ close to 1 often leave comments in other people's posts, but rarely receive comments from other users. Therefore, in

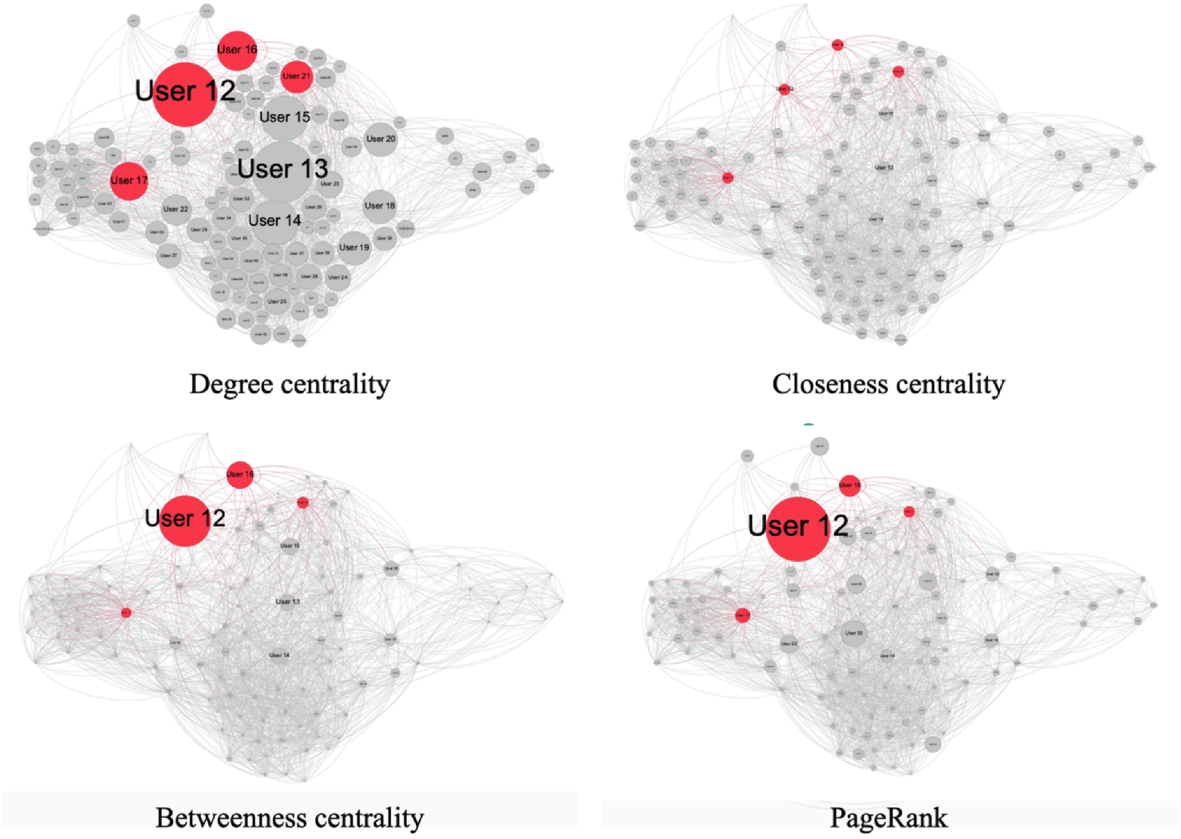

Fig. 3. Centrality Indices and the Comment Network on Daum Agora *Note: Key Cyberbullies were marked as red node.* (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

order to limit the analysis to people who send and receive more comments, users with $\frac{In-degreecentrality}{Degreecentrality}$ and $\frac{Out-degreecentrality}{Degreecentrality}$ centrality exceeding 90% were removed from the analysis.

Fig. 3 shows a comment network on Daum Agora based on 4 different centrality indices. Well known key cyberbullies were marked as red nodes. Plotting these network diagrams visualized the explanatory power of each centrality index for identifying key cyberbullies. The centrality index that most accurately reflects the influence in the community will be tested in subsequent experiments. Table 7 shows the centrality indices and cyberbullying indices of the top 10 users. Table 6 does not consider a user's influence in the community, whereas Table 7 shows the influence of each user in the community. This enables true identification of key cyberbullies by considering a user's influence in the community and his or her proportion of P/N comments (Losada ratio).

5.4. Evaluation

We randomly extracted the data of 3,200 users for validation. Four graduate students labeled whether users were cyberbullies. The coders used four criteria to label these users. Three cyberbullying criteria were used: repetition, aggression and intentionality (Rosa et al., 2019). With it, community influence was also used as a criterion because identifying key cyberbullies was the focus of this study.

These criteria were considered comprehensively in concluding whether a user could be designated as a key cyberbully. We used Cohen's kappa scores to assess the "intercoder reliability" on all transcripts. Every calculation created an overall kappa score in which true agreement was assessed, and the resulting score revealed the kappa score once overlapping coding was considered. The kappa scores in the study ranged from 0.72 to 0.94, all of which exceeded the suggested range of 0.7 (Lombard et al., 2002). The strength of the agreement level is "substantial" (Landis and Koch, 1977). This established the intercoder reliability of this study.

A consensus on evaluating cyberbullying classifiers has not been established. We followed the criteria used in the evaluation of numerous studies. Many studies have reported cyberbullying class (CB) performance using precision, recall, and F1 scores. (e.g., Dadvar et al., 2012; Huang et al., 2014; Nahar et al., 2014; Ptaszynski et al., 2016; Reynolds et al., 2011; Rosa et al., 2018, 2018b; Van Hee et al., 2015). Among these classifiers, F1 scores are the harmonic mean of precision and recall and are a criterion for comprehensive evaluation. In this study, we focused on the F1 score to evaluate the performance of key cyberbully classifications.

In this study, three features proposed for key cyberbully detection represent the features used in the existing research into detection of cyberbullying. First, the CBI is a textual feature that is similar to Bayzick et al. (2011)'s use of the number and density of "bad" words in determining what words to include in a dictionary. Second, the Losada ratio corresponds to the sentiment feature. Sentiment features have been used in existing cyberbullying detection studies (Dinakar et al., 2011; Van Hee et al., 2015; Sugandhi et al., 2016). Third, the concept of centrality of the SNA has been used as a social feature.

To evaluate the impact of each feature, we designed nine scenarios. These scenarios are detailed below:

- Baseline1: Malicious
- Baseline2: Frequency
- Scenario A: CBI
- Scenario B: CBI + Sentiment (positive/ neutral/ negative)
- Scenario C: CBI + Losada ratio
- Scenario D: CBI + Losada ratio + Degree centrality
- Scenario E: CBI + Losada ratio + Out-degree centrality
- Scenario F: CBI + Losada ratio + Degree centrality + Out-degree centrality
- Scenario G: CBI + Losada ratio + PageRanks
- Scenario H: CBI + Losada ratio + Closeness centrality
- Scenario I: CBI + Losada ratio + Betweenness centrality

Table 7
Degree Centrality and Cyberbullying Index.

Ranking [3]	User ID	Cyberbullying Index (CI)	Degree Centrality (DC) [4]	BetweennessCentrality [5]	Closeness Centrality [6]	PageRanks [7]
1	User 12	0.46	612	0.3410	0.0084	0.3410
2	User 13	0.03	581	0.3817	0.0005	0.3817
3	User 14	0.03	439	0.3739	0.0016	0.3739
4	User 15	0.04	429	0.3808	0.0005	0.3808
5	User 16	0.13	378	0.3418	0.0028	0.3418
6	User 17	0.14	361	0.3245	0.0019	0.3245
7	User 18	0.04	324	0.3357	0.0018	0.3357
8	User 19	0.04	321	0.3502	0.0006	0.3502
9	User 20	0.03	320	0.3468	0.0018	0.3468
10	User 21	0.24	307	0.3416	0.0014	0.3416

[3] *Top 10 users were selected based on degree centrality results*
[4] *Mean = 8.19, Standard deviation = 24.35 for all users*
[5] *Mean = 8230.58, Standard deviation = 66247.84 for all users*
[6] *Mean = 0.25, Standard deviation = 0.32 for all users*
[7] *Mean = 0.0001, Standard deviation = 0.0002 for all users*

To evaluate the performance of each feature combination, we applied three models—Random forest, logistic regression, and support vector machines (Rosa et al., 2019)—that are the models used most in cyberbullying detection. We evaluated precision, recall, and F1 scores as well as accuracy to ensure proper performance reviews (Powers, 2011). We also compared the performance of our approach with previous research by applying to each scenario the three machine-learning models used in the existing cyberbullying detection studies. Overall, the performance evaluation results were similar to or slightly superior to those of Rosa et al. (2019) who systematically reviewed 22 automatic cyberbullying detection studies. Of course, because the purposes of classification are different for detection of cyberbullying and detection of key cyberbullies, it can be said that the methodology presented in this study is meaningful even when simple comparisons of performance are difficult.

Performance evaluations (see Table 8) show that Scenario D based on degree centrality has the best overall performance. Other centralities such as betweenness, closeness, and PageRank showed little effect on performance improvement, but out-degree centrality and degree centrality improved the performance of the classifier. We can see that Scenario D has better performance with fewer variables than F with both centralities, so it is more efficient variable utilization.

On discussion boards such as Daum Agora, unilateral comments (high out-degree) are not influential. If a user unilaterally comments a lot but other users do not respond, the influence of that user is not high. High in-degree centrality alone makes it difficult to measure a user's impact. A lot of the comments may be because of their high impact, but they may also be controversial. If only the in-degree centrality is high, and if there are few actions for commenting on the posts of other users, it is not enough to be a high-impact key user. A user with a high degree centrality frequently responds to posts of other users (high out-degree) and receives a lot of comments from other users (high in-degree). Degree centrality is the simplest and the best centrality measure in communities like Daum Agora.

Betweenness centrality, closeness centrality, and PageRank measures were not significant in Daum Agora, which consists of posts and comments. If the connection relationship has, like Twitter, the same hierarchy for each node, the shortest path or intermediate location may be significant, but not in post comment form such as in Daum Agora.

5.5. *Identifying key Cyberbullies—Multiplication of cyberbullying index and degree centrality*

We classified users as key cyberbullies if they had high degree centrality and their cyberbullying indices exceeded Daum Agora's overall average of 0.11. The Losada ratios of all users with cyberbullying indices exceeding 0.11 were very low. Thus, we did not consider the Losada ratio separately.

We multiplied the cyberbullying index and degree centrality for each user. We call the multiplication score the *cyberbullying score*: cyberbullying $score_i$ = cyberbullying $index_i$ * degree $centrality_i$ for the $_i$th user. The cyberbullying score can be used as a criterion for the ranking of key cyberbullies with malicious comments in the targeted online space.

Table 9 shows the top 15 users with the highest cyberbullying scores. User12 is the most problematic key cyberbully. This user appears to be very influential in Daum Agora with a high cyberbullying index (0.46) and degree centrality (612), meaning that 46% of all this user's comments in 612 interactions with other users are malicious. The top 14 users whose multiplication scores exceed the threshold (45) also can be regarded as key cyberbullies.

For validation, we further multiplied degree centrality and the Cyberbullying Index of the top 19 users in Table 6. Sixteen of these have cyberbullying scores lower than 45; User 12, User 11, and User 9 are the exceptions. We can confirm that User 11 and User 9 are suspects, and User 12 is clearly the worst key cyberbully in Daum Agora. We further multiplied degree centrality and the Cyberbullying Index for the top 10 users in Table 7. Among these 10 users, only four (User 12, User 21, User 17, and User 16) have cyberbullying scores that exceed 45. Based on these results, we perceive that a cyberbullying score as the multiplication of degree centrality and the Cyberbullying Index is an effective identification criterion.

By plotting the cyberbullying scores (i.e., DC*CI in Table 9) on a graph (See Fig. 4), we perceived there are four tiers among the top 15 users. Similar to a scree plot test (Hair et al., 1998), the steep slope could be used as criteria for clear separations between tier groups. There is a steep slope between groups and a gentler slope in each group. We could thus identify key cyberbullies (i.e., retained

Table 8
User Level of Losada Ratio and Cyberbullying Index.

Scenario	Precision			Recall			F1-score		
	RF	LR	SVM	RF	LR	SVM	RF	LR	SVM
Base 1	0.75	0.71	0.65	0.76	0.59	0.64	0.75	0.63	0.65
Base 2	0.60	0.66	0.62	0.60	0.52	0.55	0.61	0.53	0.57
A	0.66	0.49	0.49	0.62	0.50	0.50	0.64	0.50	0.49
B	0.79	0.74	0.74	0.71	0.57	0.54	0.75	0.60	0.56
C	0.69	0.49	0.49	0.68	0.50	0.50	0.69	0.50	0.50
D	**0.81**	**0.72**	**0.67**	**0.78**	**0.60**	**0.63**	**0.80**	**0.63**	**0.65**
E	0.88	0.62	0.61	0.65	0.52	0.59	0.71	0.53	0.60
F	0.75	0.74	0.63	0.61	0.59	0.57	0.65	0.62	0.59
G	0.62	0.49	0.49	0.56	0.50	0.50	0.58	0.50	0.50
H	0.75	0.49	0.49	0.71	0.50	0.50	0.72	0.49	0.50
I	0.84	0.66	0.49	0.65	0.52	0.50	0.71	0.54	0.50
All	0.88	0.78	0.49	0.65	0.60	0.50	0.71	0.64	0.50

Note: Random forest (RF), Logistic regression (LR), Support Vector machine (SVM).

Table 9
Multiplication of Degree Centrality and Cyberbullying Index.

Ranking [5)]	User ID	Degree Centrality (DC)	Cyberbullying Index (CI)	DC * CI [6)]	Tier
1	User 12[T6,T7]	612	0.46	280.60	Tier 1
2	User 22	214	0.37	79.50	Tier 2
3	**User 21**[T7]	307	0.24	72.90	
4	User 11	106	0.58	61.75	
5	User 23	230	0.24	54.12	
6	User 24	198	0.27	54.06	
7	User 17[T7]	361	0.14	49.49	Tier 3
8	User 16[T7]	378	0.13	48.54	
9	User 25	178	0.27	48.11	
10	User 26	178	0.26	46.67	
11	User 27	118	0.39	46.60	
12	User 28	212	0.22	46.38	
13	User 29	210	0.22	46.14	
14	User 9[T6]	72	0.64	45.78	
15	User 30[T6]	79	0.50	39.17	Tier 4

[5)] *Top 15 users were selected based on the cyberbullying score (DC*CI)* [6)] *Mean = 7.14, Standard deviation = 14.00 for all users*
[T6] *Listed in* Table 6 *(Top 19 users based on the Cyberbullying Index)*
[T7] *Listed in* Table 7 *(Top 10 users based on degree centrality results)*

as significant) compared to other cyberbullies by applying the slope check of scree plot test. The first tier consists of the top user (User 12), a second tier of the top five users whose scores exceed 54 (User 22, User 21, User 11, User 23, User 24), and a third tier with eight suspect users whose scores exceed 45). We chose 45 as the threshold, which divides the third and fourth tiers.

6. Discussion and implications

6.1. Discussion of the study

In this study, we proposed a method of identifying key cyberbullies through text mining and SNA techniques. We also proposed a cyberbullying index and used the Losada ratio (Fredrickson and Losada, 2005) based on the text mining approach. More than 650,000 posts and comments from Daum Agora were collected and analyzed to identify key cyberbullies.

The Losada ratio and the cyberbullying index were used to measure the health of the community and its degree of cyberbullying. The Losada ratio was low and the cyberbullying index was higher in communities such as Politics and Society that lend themselves to highly controversial discussions. Thus, the rate of positive comments was low, and the rate of comments containing insulting words was high. These communities are places where cyberbullying was manifested.

We also calculated the cyberbullying index and the Losada ratio at the user level, which had its constraints. They did not reflect, either together or alone, users' influence on the community; the Losada ratio simply calculated the P/N ratio of comments. Consequently, we used SNA, which uses centrality to track how users interact, to analyze which users played a key role in the community. Users can be considered key cyberbullies if they have a low Losada ratio, a high cyberbullying index, and high centrality.

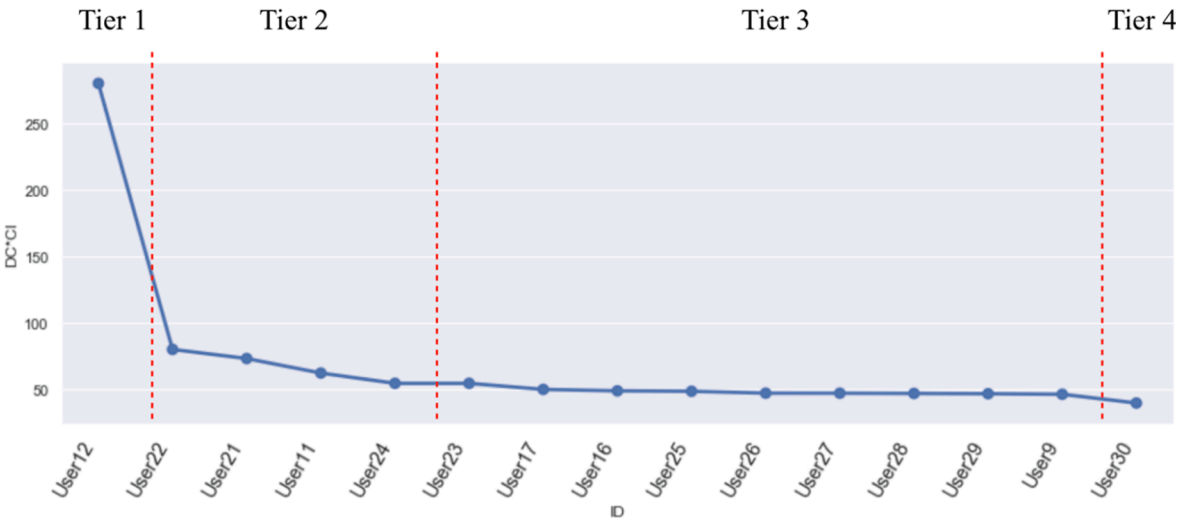

Fig. 4. Results of Cyberbullying Scores on a Scree Plot Test.

Nine scenarios were created to verify the impact of each feature. Among the various SNA centrality indices, we evaluated the centrality indices that most effectively identified key cyberbullies. Degree centrality emerged from this process as the most suitable among the various SNA concepts of centrality. For verification, we used data that labeled the key cyberbullies. The F1 score of this study's best classifier is 0.80, which indicated relatively high efficacy in identifying key cyberbullies.

A higher Losada ratio facilitates a community's or an individuals' prosperity (Fredrickson and Losada 2005). If the number of posts or comments were defined as a measure of prosperity, the number of posts would increase with a low Losada ratio. From the Daum Agora data, the Losada ratio and the number of posts have a high correlation of −41.72%. The higher the rate of negative posts, the greater the number of posts in the community. This phenomenon may be explained by how negative cues attract more attention (Kanouse and Hanson, 1987).

6.2. Ltations and future research direction

At the community level, we faced limitations in data collection. As a result, some data were concentrated in a certain period for specific communities. This may have rendered some results atypical. Future studies based on more diverse and larger amounts of data collected at many points in time may yield more dependable results.

Second, this study has a methodological limitation because of the use of a dictionary-based text-mining approach. Our dictionary incorporated as many insulting words and diverse abbreviations as we could acquire; however, there is a room for improvement. Dictionary-based text mining is limited in its capability to judge subjective malicious comments that lack cursing or insulting words. Subjective malicious comments include sarcasm such as "Well done. I'm glad it happened" or "Great job!" to someone who made a mistake. These phrases cannot be detected as malicious comments. Future research can apply deep learning methods in identifying malicious comments. For example, recurrent neural networks (RNN), i.e., long-short term memory (LSTM), can be used for text classification (Liu et al., 2016). Future studies can also consider applying deep learning methods in identifying cyberbullies.

Third, purchases of illegal IDs and revolving Internet Protocols (IPs) are used to mask behavior (Joonang Daily, 2018). Thus, malicious users can quickly create public opinion and make it appear many different people chimed in. In our study, we did not deal with this type of manipulation. It was impossible to distinguish the IP of users with multiple IDs. In future research, we should consider detecting the various IPs of individual users.

Fourth, future studies need to implement the proposed method in online communities. By developing a community of prototypes equipped with functions to prevent cyberbullying by extracting posts by key cyberbullies in real time, it will be possible to use the methods our research suggests to conduct empirical verification through applying real community mirroring and cyber violence prevention measures. This aim should further increase the performance of the classifiers to achieve high levels of performance that enable automatic detection of key cyberbullies. Future studies can also consider quantifying the actual impact of cyberbullies on the users they bully.

6.3. Implications for research and practice

Our work has several implications for research. First, this study expands the literature on big social data. Big social data is "high-volume, high-velocity, high-variety and/or highly semantic data that is generated from technology-mediated social interactions and actions in the digital realm, and which can be collected and analyzed to model social interactions and behavior (Olshannikova et al., 2017)." In the cyberbullying problem, which is a big problem in modern society, we analyzed the interaction between cyberbullies and general users and proposed a methodology for detecting key cyberbullies. This study also contributes to the literature on cyberbullying by proposing a method to identify cyberbullies that combines text mining and SNA. We concentrated on cyberbully identification and expanded the scope of cyberbullying research. Previous studies mainly focused on the detection of cyberbullying comments and the question of whether specific comments constitute cyberbullying. They were interested in improving model performance by improving analytical techniques (Di Capua et al., 2016; Nahar et al., 2014; Reynolds et al., 2011) or in expanding the input data (Dadvar et al., 2012; Huang et al., 2014; Sanchez and Kumar, 2011). No methodology has hitherto has been proposed to identify key cyberbullies. We focused on this identification problem and presented a unique method to use real data to identify them. This will serve as the basis for future research.

Second, this study contributes to the literature by demonstrating how to coordinate the use of SNA with estimation of the Losada ratio and a cyberbullying index based on the results of text mining results. We also proposed a new cyberbullying index that can be used to estimate the ratio of malicious comments among all comments at the user level and at the community level. Although the Losada ratio has been used in such fields as behavioral science, psychology, and marketing (e.g., Fredrickson and Losada, 2005), we have extended its use by demonstrating how it can be used in identifying cyberbullies. In addition to the application of the cyberbullying index and the Losada ratio, we have demonstrated how the centrality indices of SNA can be used to identify key cyberbullies. One topic of prior research on cyberbullying detection is to expand machine learning methods from supervised learning to semi supervised and unsupervised learning (Nahar et al., 2014; Di Capua et al., 2016). Such research is being conducted because it is difficult to obtain cyberbully-labeled data in real–life situations. The proposed methodology can easily identify key cyberbully suspects by integrating simple index calculations and unlabeled data.

Third, this study resulted in the compilation of a dictionary of Korean insulting words and the creation of its correlated cyberbullying index. The dictionary of Korean insulting words was compiled based on the prohibited words found in other services (Naver Café, blog, and P2P services). The dictionary contains a variety of insulting words reflecting the latest terms and derivatives. These tools could support cyberbullying research in the future.

This study also has several implications for practice. First, this study presented a practical method for platform operators and content moderators to use in identifying key cyberbullies. By combining the three elements of our approach at the community and individual levels, they can identify problematic communities and key cyberbullies on their platforms. It is possible by calculating the Losada ratio and the cyberbullying index to assess a situation in terms of its cyberbullying status. The comparison between the Losada ratio and the cyberbullying index with the Daum Agora data can expose the status of their platforms.

Second, this study has another implication in predictive policing (Perry et al., 2010). Predictive policing is defined refers to the science that calculates risks in relation to a crime using models and relevant data (Rienks, 2015). Predictive policing has been cited as 'the minority report-style policing' (Newbold, 2015). For 'Minority Report-style' predictive forum policing, our model explicitly enables the prediction of future cyberbully suspects. It is possible to identify key cyberbullies and rank them as threats, according to a multiplication score (i.e., Cyberbullying score) based on the cyberbullying index and degree centrality. Such assessments and subsequent rankings can help platform operators mitigate cyberbullying before it takes a turn for the worse. In addition, the Korean National Police Agency's Cyber Bureau can proactively screen cyberbullies to prevent cybercrime based on application of the proposed method.

The proposed methodology also has the potential to go beyond identifying cyberbullies and negating cyberbullying and instead promote benevolent comments (i.e., Comments that express goodwill and/or help others) in the online space (Jang et al., 2016). Such promotion of benevolent comments has the potential to lead to the development of online social norms that themselves in turn could serve to reduce the posting of malicious comments and lead to a lessening of cyberbullying (Jang et al., 2016). To promote the posting of benevolent comments, we could consider using the proposed method to rank benevolent commenters. The multiplication score of the Losada ratio and degree centrality can be used to estimate the benevolence score in a selected online space. A higher Losada ratio implies more positive comments compared with negative comments. Those online users with higher Losada ratios and more interactions with others, (i.e., Higher degree centrality) could influence others and help develop online social norms.

7. Conclusion

Cyberbullying that uses malicious comments is a major problem in our society, and the damage it causes is becoming increasingly significant. Such comments have led to suicide as well as to psychological shock. Previous studies on cyberbullying focused more on detecting and classifying malicious comments than on identifying cyberbullies (Law et al., 2012; Mishna et al., 2009; Slonje and Smith, 2008; Willard, 2007). Going beyond this previous research, our study focused on a substantive alternative that enables the blocking of malicious comments through identification of key cyberbullies.

This study proposed a method to identify key cyberbullies and focused on the problem of malicious comments. To identify key cyberbullies with high rates of posting insulting comments and resultant significant influence on the community, we applied text mining and SNA methods. Based on the application of text mining, we proposed a new cyberbullying index and estimated the cyberbullying index and Losada ratio of each user and each community. Moreover, based on the application of SNA, we calculated centrality indices (degree centrality, closeness centrality, betweenness centrality, and PageRanks) and evaluated them based on nine scenarios to identify features that are effective identifiers of key cyberbullies. We found out that Cyberbullying Index, Losada ratio and Degree centrality work most effectively. In addition to the complementary information of the Losada ratio, the multiplication score (i.e., cyberbullying score) composed of degree centrality and the cyberbullying index permits the ranking of cyberbullies that in turn leads to identification of key cyberbullies — the worst offenders.

Our novel approach sheds light on identifying key cyberbully topics and extends the literature on cyberbullying. Identification of its worst perpetrators is essential to developing practical measures to prevent cyberbullying. The concrete method proposed in this study will help online platform operators identify key offenders as a step in creating healthy online communities and also help government agencies prevent potential cyberbullying. We hope that more effort will be devoted to this important research area and that the proposed method will serve as a useful tool for such future work.

Yoon-Jin Choi is a PhD candidate in the Graduate School of Information at Yonsei University, Seoul Korea. Her research focuses on social media marketing, electronic commerce, and cyberbullying. She presented her research work at the *International Conference on Information Systems*. Her research work is forthcoming in *Journal of Global Information Management*.

Byeong-Jin Jeon is an expert in Data Analytics. His research focuses on text mining and the applications. He presented his research work at the *International Conference on Information Systems*.

Hee-Woong Kim is a Professor in the Graduate School of Information at Yonsei University, Seoul Korea. Before joining Yonsei University, he was a faculty member in the Department of Information Systems and Analytics at the National University of Singapore (NUS). He has served on the editorial boards of the *Journal of the Association for Information Systems* and *IEEE Transactions on Engineering Management*. His research work has appeared in *Communications of the ACM. Computers in Human Behavior, Cyberpsychology Behavior and Social Networking, Decision Support Systems, European Journal of Operational Research, IEEE Transactions on Engineering Management, Information & Management, Information Systems Research, International Journal of Electronic Commerce, International Journal of Information Management, Journal of the Association for Information Systems, Journal of Management Information Systems, Journal of Retailing,* and *MIS Quarterly*.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

Aleahmad, A., Karisani, P., Rahgozar, M., Oroumchian, F., 2016. OLFinder: Finding opinion leaders in online social networks. J. Inf. Sci. 42 (5), 659–674.
Al-Garadi, M., Varathan, K.D., Ravana, S.D., 2016. Cybercrime detection in online communications: The experimental case of cyberbullying detection in the Twitter network. Comput. Hum. Behav. 63, 433–443. https://doi.org/10.1016/j.chb. 2016.05.051.
Bavelas, A., 1950. Communication patterns in task-oriented groups. J. Acoust. Soc. Am. 22 (6), 725–730.
Bayzick, J., Kontostathis, A., & Edwards, L. 2011. Detecting the presence of cyberbullying using computer software. Proceedings of the 3rd annual ACM web science conference (WebSci '11). Retrieved from http://www.websci11.org/fileadmin/websci/Posters/63_paper.pdf.
Beauchamp, M.A., 1965. An improved index of centrality. Behav. Sci. 10 (2), 161–163.
Blumstein, A., Wallman, J., & Farrington, D. (Eds.). 2006. The crime drop in America. Cambridge University Press.
Brin, S., & Page, L. 1998. The anatomy of a large-scale hypertextual web search engine.
Brown, C.F., Demaray, M.K., Secord, S.M., 2014. Cyber victimization in middle school and relations to social emotional outcomes. Comput. Hum. Behav. 35, 12–21.
Brown, N.J., Sokal, A.D., Friedman, H.L., 2013. The complex dynamics of wishful thinking: The critical positivity ratio. Am. Psychol. 1–35.
Camcho, C., Hassanein, K., Head, M., 2018. Cyberbullying impacts on victim's satisfaction with information and communication technologies: The role of perceiving cyber bullying severity. Inf. Manage. 55 (4), 494–507.
Chavan, V. S., & Shylaja, S. S. 2015. Machine learning approach for detection of cyber-aggressive comments by peers on social media network. Proceedings of the international conference on advances in computing, communications and informatics (pp. 2354–2358). ICACCI. https://doi.org/10.1109/ICACCI.2015.7275970.
Choi et al., 2018. Cyberbullying Survey, Korea Broadcasting Commission, National Information Society Agency, 2018. Retrieved from: https://kcc.go.kr/user.do?mode=view&page=A02060400&dc=60400&boardId=1030&cp=1&boardSeq=46802.
Cohen, S., Underwood, L. G., & Gottlieb, B. H. (Eds.). 2000. Social support measurement and intervention: A guide for health and social scientists. Oxford University Press, 29–52.
Corbin, J., & Strauss, A. 2008. Strategies for qualitative data analysis. Basics of Qualitative Research. Techniques and procedures for developing grounded theory, 3.
Corriss, L. 2010. Information security governance: Integrating security into the organizational culture. In Proceedings of the 2010 Workshop on Governance of Technology, Information and Policies, 35–41.
Dadvar, M., Jong, D. F., Ordelman, R., & Trieschnigg, D. 2012. Improved cyber bullying detection using gender information. In Proceedings of the Twelfth Dutch-Belgian Information Retrieval Workshop (DIR 2012). University of Ghent.
Dhang-Pham, D., Pittayachawan, S., Bruno, V., 2017. Applying network analysis to investigate interpersonal influence of information security behaviors in the workplace. Inf. Manage. 54 (5), 625–637.
Di Capua, M., Di Nardo, E., & Petrosino, A. 2016. Unsupervised cyber bullying detection in social networks. In Pattern Recognition (ICPR), 2016 23rd International Conference, 432–437.
Diehl, M., Hay, E.L., Berg, K.M., 2011. The ratio between positive and negative affect and flourishing mental health across adulthood. Aging Mental Health 15 (7), 882–893.
Dinakar, K., Reichart, R., Lieberman, H., 2011. Modeling the detection of textual cyberbullying. Social Mobile Web 11 (02), 11–17.
Franzke, A. S., Bechmann, A., Ess, C. M., & Zimmer, M. 2020. Internet Research: Ethical Guidelines 3.0. Retrives from https://aoir.org/reports/ethics3.pdf.
Fredrickson, B.L., Losada, M.F., 2005. Positive affect and the complex dynamics of human flourishing. Am. Psychol. 60 (7), 678–686.
Freemann, L.C., 1977. A set of measures of centrality based on betweenness. Sociometry 40, 35–41.
Freire, M., Antunes, F., & Costa, J. P. 2017. A semantics extraction framework for decision support in context-specific social web networks. In International Conference on Decision Support System Technology, 133-147.
Golub, G., Van, C., 1996. Matrix Computations, 3rd ed.,. Johns Hopkins University Press.
Hair, J.F., Anderson, R.E., Tatham, R.L., Black, W.C., 1998. Multivariate Data Analysis, 5th ed. Prentice-Hall International.
Hinduja, S., Patchin, J.W., 2007. Offline consequences of online victimization: School violence and delinquency. J. Sch. Violence 6 (3), 89–112.
Hosseinmardi, H., Rafiq, R. I., Han, R., Lv, Q., & Mishra, S. 2016. Prediction of cyberbullying incidents in a media-based social network. Proceedings of the 2016 IEEE/ACM international conference on advances in social networks analysis and mining (pp. 186–192). IEEE Press. https://doi.org/10.1109/asonam.2016.7752233.
Huang, Q., Singh, V. K., & Atrey, P. K. 2014. Cyber bullying detection using social and textual analysis. In Proceedings of the 3rd International Workshop on Socially-Aware Multimedia (pp. 3–6).
Jang, Y.J., Kim, H.W., Jung, Y., 2016. A mixed methods approach to the posting of benevolent comments online. Int. J. Inf. Manage. 36 (3), 414–424.
Jeon, B.J., Choi, Y.J., Kim, H.W., 2019. Application development for text mining: KoALA. Inf. Syst. Rev. 21 (1), 117–137.
Joongang Daily. 2018. Comments also make public opinion less than two hours. Joongang Daily, March 20. http://news.joins.com/article/22456090.
Kanouse, D. E., & Hanson Jr, L. R. 1987. Negativity in evaluations. In Preparation of this paper grew out of a workshop on attribution theory held at the University of California, Los Angeles, Aug 1969. Lawrence Erlbaum Associates, Inc.
Karimi, F., Matous, P., 2018. Mapping diversity and inclusion in student societies: A social network perspective. Comput. Hum. Behav. 88, 184–194.
Kelling, G.L., Coles, C.M., 1997. Fixing Broken Windows: Restoring Order and Reducing Crime in Our Communities. Simon and Schuster.
Kelling, G. L., & Sousa, W. H. 2001. Do police matter?: An analysis of the impact of new york city's police reforms. CCI Center for Civic Innovation at the Manhattan Institute.
Kennamer, J.D., 1990. Self-serving biases in perceiving the opinions of others: Implications for the spiral of silence. Commun. Res. 17 (3), 393–404.
Kiss, C., Bichler, M., 2008. Identification of influencers—measuring influence in customer networks. Decis. Support Syst. 46 (1), 233–253.
Korea Communications Commission. 2018, "Results of cyber violence survey in 2017", http://www.kcc.go.kr/user.do?boardId=1113&page=A05030000&dc=K00000001&boardSeq=45505&mode=view.
Landis, J.R., Koch, G.G., 1977. The measurement of observer agreement for categorical data. Biometrics 159–174.
Law, D.M., Shapka, J.D., Hymel, S., Olson, B.F., Waterhouse, T., 2012. The changing face of bullying: An empirical comparison between traditional and internet bullying and victimization. Comput. Hum. Behav. 28 (1), 226–232.
Lee, S.H., Kim, H.W., 2015. Why people post benevolent and malicious comments online. Commun. ACM 58 (11), 74–79.
Lenhart, A., Madden, M., Smith, A., Purcell, K., Zickuhr, K., & Rainie, L. 2011. Teens, kindness and cruelty on social network sites: How American teens navigate the new world of" Digital Citizenship". Pew Internet & American Life Project.
Liu, P., Qiu, X., & Huang, X. 2016. Recurrent neural network for text classification with multi-task learning, International Joint Conference on Artificial Intelligence (IJCAI), 2016. [Accessed at https://openreview.net/forum?id=Hk4onQfuWr].
Liu, X., Fahmy, S., 2011. Exploring the spiral of silence in the virtual world: Individuals willingness to express personal opinions in online versus offline settings. J. Media Commun. Stud. 3 (2), 45–57.
Lombard, M., Snyder-Duch, J., Bracken, C.C., 2002. Content analysis in mass communication: Assessment and reporting of intercoder reliability. Human Commun. Res. 28 (4), 587–604.
Losada, M., Heaphy, E., 2004. The role of positivity and connectivity in the performance of business teams: A nonlinear dynamics model. Am. Behav. Sci. 47 (6), 740–765.
Martinez, I., Murgui, S., Garcia, O.F., Garcia, F., 2019. Parenting in the digital era: Protective and risk parenting styles for traditional bullying and cyberbullying victimization. Comput. Hum. Behav. 90, 84–92.
Mishna, F., Saini, M., Solomon, S., 2009. Ongoing and online: Children and youth's perceptions of cyber bullying. Child. Youth Serv. Rev. 31 (12), 1222–1228.
Moy, P., Domke, D., Stamm, K., 2001. The spiral of silence and public opinion on affirmative action. J. Mass Commun. Q. 78 (1), 7–25.
Nahar, V., Al-Maskari, S., Li, X., & Pang, C. 2014. Semi-supervised learning for cyberbullying detection in social networks. In Australasian Database Conference, 160–171.
Newbold, J. 2015. 'Predictive Policing,' 'Preventative Policing'or 'Intelligence Led Policing.' What Is the Future? Warwick Business School, Coventry, UK.

Nickerson, C.A., 2014. No empirical evidence for critical positivity ratios. Am. Psychol. 69 (6), 626–629.
Noelle-Neumann, E., 1977. Turbulences in the climate of opinion: Methodological applications of the spiral of silence theory. Publ. Opin. Q. 41 (2), 143–158.
Olshannikova, E., Olsson, T., Huhtamäki, J., Kärkkäinen, H., 2017. Conceptualizing big social data. J. Big Data 4 (1), 19. https://doi.org/10.1186/s40537-017-0063-x.
Perry, W. L., McInnis, B., Price, C. C., Smith, S. C., & Hollywood, J. S. 2010. Predictive Policing, the role of Crime forecasting in Law Enforcement. RAND Corporation.
Peter, I., Petermann, F., 2018. Cyberbullying: A concept analysis of defining attributes and additional influencing factors. Comput. Hum. Behav. 86, 350–366.
Powers, D., 2011. Evaluation: From precision, recall, and F-measure to ROC, informedness, markedness, and correlation. J. Mach. Learn. Technol. 2 (1), 37–63.
Prollochs, N., & Feuerriegel, S. 2019. Business analytics for strategic management: Identifying and assessing corporate challenges via topic modeling. Information & Management, Articles in press.
Ptaszynski, M., Masui, F., Nitta, T., Hatakeyama, S., Kimura, Y., Rzepka, R., et al., 2016. Sustainable cyberbullying detection with category-maximized relevance of harmful phrases and double-filtered automatic optimization. Int. J. Child. Comput. Interact. 8, 15–30. https://doi.org/10.1016/j.ijcci.2016.07.002.
Rego, A., Sousa, F., Marques, C., Cunha, M.P.E., 2012. Optimism predicting employees' creativity: The mediating role of positive affect and the positivity ratio. Eur. J. Work Org. Psychol. 21 (2), 244–270.
Reynolds, K., Kontostathis, A., & Edwards, L. 2011. Using machine learning to detect cyber bullying. In Machine learning and applications and workshops (ICMLA), 2011 10th International Conference, 2, 241–244.
Rienks, R. 2015. Predictive Policing: Taking a chance for a safer future. Korpsmedia, PDC.
Rosa, H., Carvalho, J. P., Astudillo, R., & Batista, F. 2018. Page rank versus katz: Is the centrality algorithm choice relevant to measure user influence in Twitter? In L. Kóczy, J. Medina (Ed.). Studies in Computational Intelligence (pp. 1–9). Cham: Springer. https://doi.org/10.1007/978-3-319-74681-4_1.
Rosa, H., Carvalho, J. P., Calado, P., Martins, B., Ribeiro, R., & Coheur, L. 2018a. Using fuzzy fingerprints for cyberbullying detection in social networks. Proceedings of the IEEE International Conference on Fuzzy Systems (pp. 56–62). https://doi.org/10. 1109/FUZZ-IEEE.2018.8491557.
Rosa, H., Matos, D., Ribeiro, R., Coheur, L., & Carvalho, J. P. 2018b. A deeper look at detecting cyberbullying in social networks. Proceedings of the International Joint Conference on Neural Networks (pp. 323–330). https://doi.org/10.1109/IJCNN. 2018.8489211.
Rosa, H., Pereira, N., Ribeiro, R., Ferreira, P.C., Carvalho, J.P., Oliveira, S., Trancoso, I., 2019. Automatic cyberbullying detection: A systematic review. Comput. Hum. Behav. 93, 333–345.
Sanchez, H., & Kumar, S. 2011. Twitter bullying detection. ser. NSDI, 12, 15-15.
Serrat, O. 2017. Social network analysis. In Knowledge solutions, 39–43.
Siering, M., Deokar, A.V., Janze, C., 2018. Disentangling consumer recommendations: Explaining and predicting airline recommendations based on online reviews. Decis. Support Syst. 107, 52–63.
Singh, V. K., Huang, Q., & Atrey, P. K. 2016. Cyberbullying detection using probabilistic socio-textual information fusion. Proceedings of the international conference on advances in social networks analysis and mining (pp. 884–887). ASONAM. https://doi.org/10.1109/asonam.2016.7752342.
Skogan, W.G., 1992. Disorder and Decline: Crime and the Spiral of Decay in American Neighborhoods. Univ of California Press.
Skogan, W. G., & Frydl, K. 2004. Fairness and effectiveness in policing: The evidence.
Slonje, R., Smith, P.K., 2008. Cyberbullying: Another main type of bullying? Scand. J. Psychol. 49 (2), 147–154.
Sugandhi, R., Pande, A., Agrawal, A., & Bhagat, H. 2016. Automatic monitoring and prevention of cyberbullying. Int. J. Comput. Appl., 8, 17–19. Retrieved from https://pdfs.semanticscholar.org/eb09/ e30150f3adbe00cb3e384d45fdd7e7df70af.pdf.
The Guardian. 2013. Florida cyberbullying: Girls arrested after suicide of Rebecca Sedwick, 12. The Guardian, 2013.10.15, http://www.theguardian.com/world/2013/oct/15/florida-cyberbullying-rebeccasedwick-two-girls-arrested.
Van Hee, C., Lefever, E., Verhoeven, B., Mennes, J., Desmet, B., De Pauw, G., & Hoste, V. 2015. Detection and fine-grained classification of cyberbullying events. Proceedings of the international conference recent advances in natural language processing (pp. 672–680). RANLP. Retrieved from https://biblio.ugent.be/publication/ 6969774/file/6969839.pdf.
Van Mierlo, T., 2014. The 1% rule in four digital health social networks: an observational study. J. Med. Internet Res. 16 (2).
Willard, N. E. 2007. Cyberbullying and cyberthreats: Responding to the challenge of online social aggression, threats, and distress. Research Press.
Wilson, J.Q., Kelling, G.L., 1982. Broken windows. *i* 249 (3), 29–38.
Yadav, A., Wilder, B., Rice, E., Petering, R., Craddock, J., Yoshioka-Maxwell, A., Woo, D., 2018. Influence maximization in the field. Artif. Intell. Social Work 57.
Ybarra, M.L., Mitchell, K.J., Wolak, J., Finkelhor, D., 2006. Examining characteristics and associated distress related to Internet harassment: findings from the Second Youth Internet Safety Survey. Pediatrics 118 (4), e1169–e1177.
Yoo, H., 2010. The Candlelight Girls' Playground: Nationalism as Art of Dialogy, The 2008 Candlelight Vigil Protests in South Korea. Invisible Culture. Electron. J. Vis. Cult. 15, 40–78.
Zhang, S., Leidner, D., 2018. From improper to acceptable: How perpetrators neutralize workplace bullying behaviors in the cyberworld. Inf. Manage. 55 (7), 850–865.