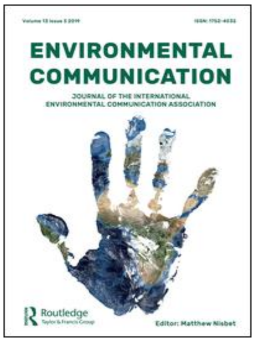

Environmental Communication

ISSN: 1752-4032 (Print) 1752-4040 (Online) Journal homepage: https://www.tandfonline.com/loi/renc20

The Uproar Over Gene-Edited Babies: A Semantic Network Analysis of CRISPR on Twitter

Christopher Calabrese, Jieyu Ding, Benjamin Millam & George A. Barnett

To cite this article: Christopher Calabrese, Jieyu Ding, Benjamin Millam & George A. Barnett (2019): The Uproar Over Gene-Edited Babies: A Semantic Network Analysis of CRISPR on Twitter, Environmental Communication, DOI: 10.1080/17524032.2019.1699135

To link to this article: https://doi.org/10.1080/17524032.2019.1699135

ENVIRONMENTAL COMMUNICATION
https://doi.org/10.1080/17524032.2019.1699135

Routledge
Taylor & Francis Group

RESEARCH ARTICLE

The Uproar Over Gene-Edited Babies: A Semantic Network Analysis of CRISPR on Twitter

Christopher Calabrese, Jieyu Ding, Benjamin Millam and George A. Barnett

Department of Communication, University of California, Davis, Davis, CA, USA

ABSTRACT

CRISPR-Cas9, a recent gene editing technology, has potential to greatly impact medicine, agriculture, and the environment; therefore, it is increasingly important to understand the public's perceptions surrounding this technology. This study examined the main themes of CRISPR during a nine-month period on Twitter through semantic network analysis. Three additional networks (before, initial, and after-event stages) were created to understand the effects of the "CRISPR babies" news event. Four themes emerged in the overall network: (1) research/applications of CRISPR, (2) the "CRISPR babies" event, (3) agricultural regulations for biotechnology, and (4) advancements in muscular dystrophy research. The "CRISPR babies" news story dominated the initial and after-event stage network clusters, indicating major events may impact how users discuss socio-scientific issues in online media. The sentiment surrounding CRISPR was mixed, demonstrating concerns over this technology. This study reveals the potential of social media as a tool to understand public discussions around new technologies.

ARTICLE HISTORY
Received 14 March 2019
Accepted 22 November 2019

KEYWORDS
CRISPR; gene editing; Twitter; thematic analysis; semantic network analysis

Introduction

Gene editing technologies, such as CRISPR-Cas9, can precisely and accurately cut and replace sequences of DNA in an organism's genome. Such a revolutionary biotechnology and its applications may likely impact agriculture and medicine, as well as our environment. Gene editing technologies have potential to feed our growing global population while minimizing impacts on the environment, such as producing crops that can grow under harsh environmental conditions and in a rapidly changing climate. Applications of gene editing in agriculture have been used to develop disease-resistant rice (Li, Liu, Spalding, Weeks, & Yang, 2012) and wheat (Wang et al., 2014). In addition to disease-resistance crops, gene editing technologies have the potential for improving human health and disease treatment. For example, recent studies have examined CRISPR-Cas9 as a possible tool against Duchenne muscular dystrophy (Bengtsson et al., 2017; Min et al., 2019).

However, applications of gene editing technologies face considerable controversy with regard to potential unintended consequences. As research develops on gene editing to reduce human and agricultural pests (Kyrou et al., 2018; Neve, 2018), there lies several potential risks that may lead to negative effects on the environment. For example, the use of gene editing to reduce a weed population may be intended to improve agricultural production, but may also have negative consequences on the overall ecosystem; changes in one population may ultimately result in an overpopulation of another species, or even the extinction of a species that is dependent on another (Saplakoglu, 2017). Further, the application of gene editing on humans has created several ethical and moral

CONTACT Christopher Calabrese cjcalabrese@ucdavis.edu Department of Communication, University of California, One Shields Avenue, Davis, CA 95616, USA

concerns. Recent research using CRISPR to edit the DNA of human embryos has sparked concerns over its safety and the potential impact it may have on the human genome itself (Marchione, 2018; Regalado, 2018b).

Despite concerns over the application of these technologies, gene editing research continues to grow and advance. With both potential benefits and risks of gene editing, it is important to assess what people's perceptions are of these technologies, especially the newer, more advanced methods like CRISPR-Cas9. News and developments on gene editing may greatly influence people's perceptions, while traditional and online media may reinforce or change them. In an age where online media dominates, people may develop their opinions surrounding new technologies through their use of social media, rather than through one-way consumption of traditional media. Therefore, it is important to understand people's perceptions of CRISPR as expressed on social media, such as microblogging sites like Twitter.

This study examines the perceptions of CRISPR online through the application of semantic network analysis of Twitter messages. The term, CRISPR, or clustered regularly interspaced short palindromic repeats, is examined instead of other more scientifically accurate terminology (e.g. CRISPR-Cas9) to capture the expressed language to better understand the perceptions of a more general audience. Semantic network analysis examines the words, concepts, and themes of CRISPR on Twitter. In addition, a case study was conducted on the effects of the breaking news story involving the creation of "CRISPR babies" by a Chinese scientist, He Jiankui. The main words and themes of CRISPR were examined through semantic networks depicting the before, initial, and after stages of the effects of this news story. Sentiment analyses were also conducted for both the overall network and the three event networks to capture the positive and negative emotions within tweets discussing CRISPR. Findings will provide researchers with a snapshot of how CRISPR is represented in an online media platform, and how it may influence public perceptions.

Literature review

Public views on gene editing

Public views of gene editing differ depending on how the technology is used; however, most surveys mainly assess its potential applications on humans. People are generally in favor of the technologies when used for treating or preventing a disease, but are not when it is used for enhancements (Blendon, Gorski, & Benson, 2016; Gaskell et al., 2017; Scheufele et al., 2017). These findings fall in line with a recent public opinion survey on using gene editing for babies, where Americans are more accepting of its application when used for medical purposes, such as for a disease or condition. However, people are strongly opposed to gene editing for enhancing the intelligence of babies (Funk & Hefferon, 2018).

While it is important to assess gene editing and its application on human health, it is also important to understand how the public views the technology in relation to other applications, such as its potential agricultural and environmental impacts. For example, when examining the public's views on gene editing for plant and animal wildlife conservation, over 80% of respondents saw gene editing as at least somewhat risky for both humans and nature (Kohl, Brossard, Scheufele, & Xenos, 2019). Therefore, to develop technology that aligns with scientific advancement, public needs, and bioethics, it is critical to understand the online discourse of new gene editing technologies, specifically more prominent ones like CRISPR-Cas9. Further understanding of the public's views related to different applications of CRISPR-Cas9 is needed for development and dissemination.

Science communication and online media

Much of the public's perceptions of new science and technology may be influenced by online media. Americans use the Internet as a main information source for science and technology-related news

(Brossard, 2013; Brossard & Scheufele, 2013; Mitchell, Funk, & Gottfried, 2017). Moreover, roughly a quarter of social media users follow science pages or accounts, thereby bringing more science information into their social media feeds (Mitchell et al., 2017). People who seek news online are more likely to form positive attitudes toward science and gain science knowledge (Dudo et al., 2011). Online science sources help with increasing knowledge about scientific issues and narrow the gaps of knowledge brought by traditional media outlets between lower and higher educated groups (Cacciatore, Scheufele, & Corley, 2014).

With the popularity of the Internet, and the linkage between online science news consumption and favorable attitudes toward science, we argue that it is important to study the contents of CRISPR-Cas9 online. In response to Brossard's (2013) call to understand the nature of online science information, we examine the discourse surrounding CRISPR on Twitter as a way to capture its audience's perceptions and sentiments.

Thematic analysis

Studies have examined the emerging themes when examining the effects of environmental protests (Doğu, 2017) and nuclear risk communication (Binder, 2012) on Twitter. These studies recognize that social media are important platforms to consider, since they reveal information related to political environmental activism and online discourse surrounding risk (Binder, 2012; Doğu, 2017). Understanding the main themes of these phenomena on social media provide a new understanding for science communication researchers to better engage with the public and address specific issues. Because of the nature of social media, there is a need to examine how new technologies, such as CRISPR, are portrayed on this platform. Social media not only provide knowledge and information, but also trigger public discussions; therefore, examining how social media, such as Twitter, represent CRISPR would help scientists better understand the audience of the platform's concerns regarding this emerging technology.

Semantic network analysis

To understand the nuances of information on CRISPR in today's media environment, we explore the posts about this topic on Twitter using semantic network analysis. Semantic network analysis describes the relationships between related concepts through the analysis of word co-occurrence. Deriving from the cognitive paradigm (D'Angelo, 2002) and the linguistic theory of frame semantics (Fillmore, 1982), semantic network analysis can highlight the most salient information in a body of text through developing networks that are representative of meaning. This type of content analysis has illustrated how certain socio-scientific issues are presented online, such as on Google search results involving the HPV vaccine (Ruiz & Barnett, 2015), online representations of genetically modified foods (Jiang, Anderton, Ronald, & Barnett, 2018), and Twitter posts related to measles outbreaks (Tang, Bie, & Zhi, 2018) and nanotechnology (Veltri, 2012). Recently, semantic network analyses were conducted to examine the main themes of gene editing within the most prominent sources viewed when seeking scientific information online (Calabrese, Anderton, & Barnett, 2019). Building on this research, it is important to understand the perceptions of gene editing technologies within online platforms.

The microblogging site, Twitter, provides the ideal content to understand what people are expressing about CRISPR online. Rather than having to decipher discussions surrounding CRISPR from specific Facebook groups or subreddits, Twitter as a platform allows users to post messages to their public profile. This allows researchers to more easily pinpoint exactly when and where messages surrounding a topic are posted through the search function. Other platforms may reveal search results with an unrepresentative sample of all discussions surrounding CRISPR. The availability of the data, through Twitter's Premium API, allows researchers to gather a full, complete database of tweets mentioning CRISPR. Based on these reasons, Twitter was chosen as the online platform for analyses. Thus, the following research questions are proposed:

RQ1: What are the most central and frequently mentioned concepts in Twitter's overall "CRISPR" network?

RQ2: What are the distinct themes that are portrayed in Twitter's overall "CRISPR" network?

RQ3: What is the sentiment of Twitter's overall "CRISPR" network?

A case study

On 25 November 2018, a breaking news story sparked discussion regarding the application of gene editing technologies on human embryos. Reports indicated that a Chinese researcher, He Jiankui, attempted to use CRISPR to create HIV-resistant human babies by inactivating the CCR5 gene (Marchione, 2018; Regalado, 2018b). This episode stirred heated discussion on the ethics surrounding the use of gene editing technologies for human enhancement. Because of this event, a debate and upsurge of tweets on gene editing emerged on Twitter. We believe that the content of these tweets surrounding this event are important to study because they reflect people's perceptions and reactions towards gene editing, which may potentially influence the applications and policies of this technology. Thus, we propose the following two research questions:

RQ4: How do the themes of "CRISPR" change before, during, and after the announcement that He Jiankui, attempted to use CRISPR to create HIV-resistant human babies?

RQ5: What is the sentiment of "CRISPR" before, during, and after this major event?

Methods

This study conducts semantic network analyses to determine the concepts and themes of "CRISPR" on Twitter. Semantic network analysis reveals the position and importance of words in relation to other words within a network based on word frequency and centrality measures (Freeman, 1978; Wasserman & Faust, 1994). This analysis can also detect concept associations through identifying word clusters within the network (Blondel, Guillaume, Lambiotte, & Lefebvre, 2008). First, to examine the overall picture of how Twitter users discuss "CRISPR," a semantic network was created based on tweets posted from 1 April to 31 December 2018. The main concepts and themes were determined from the resulting network. Second, to understand the effects of a major event, three additional semantic networks were created based on the stages of the announcement that He attempted to create HIV-resistant human babies. Tweets posted one week before the event (18–24 November) were identified as the before stage, while tweets posted within a week starting on the date of the breaking news story were identified as in the initial stage (25 November–1 December), and tweets posted a week after (2–8 December) were identified as in the after stage. The three semantic networks were compared, and the words, themes, and concepts were also determined from the resulting networks.

Data collection and analysis

Data were collected from Twitter's Premium Search Tweets API, from the 30-day and full-archive endpoints, using Python and Twitter's *searchtweets* library. The Premium Search Tweets API allows users to collect a full archive of data based on the specific search terms used when calling for the data. A 5-week preliminary search was conducted to determine the most salient keywords (CRISPR, CRISPR-Cas9, gene editing, and genome editing) on Twitter. The keyword "CRISPR" resulted with the highest number of tweets overall ($n = 47{,}481$) and the highest average number of tweets per week ($n = 9{,}497$). Other terms included tweets that were not relevant to the study's aim. While the use of one search term may serve as a potential limitation of the study, the goal was to specifically identify the common themes of the emerging technology, CRISPR-Cas9. Thus, only "CRISPR" was utilized for this study's analyses.

Based on the initial data collection and costs, the entire full-archive of tweets within a nine-month period were collected. Twitter's API was queried with "crispr lang:en," to match tweet content containing the keyword "CRISPR" and to restrict results to English-language tweets. Both Python and R (Version 3.4.4) were used to manage and clean the Twitter data. Retweets were not included in this analysis to remove the redundancy of information as well as the potential influence of bots, which would both deter from the main goal of the study to understand Twitter users' perceptions of CRISPR.[1] From Python, tweet data were saved as text files in the Twitter API JSON format, then imported into R using the *jsonlite* (Ooms, 2014) package. Tweet text data were cleaned in R using the *tm* (Feinerer, Hornik, & Meyer, 2008) and *qdap* (Rinker, 2019) packages; these packages remove URLs, convert to lowercase, expand contractions, remove punctuation, and strip whitespace. The tweets were saved into text files for further processing.

Preprocessing procedures were conducted through ConText (Diesner, 2014) to further edit the text files for analysis. This included removing syntactically functional words (e.g. the, that, they), and included word stemming (e.g. claim, claiming). Then, the frequencies of the resulting words were calculated and ranked based on their frequency of occurrence. Words that occurred with frequencies above the mean were included in the analysis.

Next, semantic matrices were generated from the processed texts based on word co-occurrence in ConText (Diesner, 2014). Miller (1956) argued that people can only process "seven plus or minus two" meaningful bits of information at a time; however, more recent studies have found that this number may range from around three to five, depending on several other factors surrounding one's memory capacity (Cowan, 2001, 2016). Thus, links were created for words that occurred within five words of one another within each tweet. The frequencies of word co-occurrence were then calculated and ranked. In the overall network, the mean frequency of co-occurrence was 86. For the event networks, the mean frequency of co-occurrence was 3 (before stage), 29 (initial stage), and 12 (after stage). Links with frequencies above the average were included in the analysis.

Semantic networks were created using the network visualization software, Gephi (Bastian, Heymann, & Jacomy, 2009). The top 75 words by frequency were included in the full nine-month network visualization, and the top 50 words by frequency were included in each event network visualization. After the data were imported, the network visualizations were adjusted using the ForceAtlas2 layout (Jacomy, Venturini, Heymann, & Bastian, 2014) to examine the spatialization between words. The size of the word label indicated how frequent the word occurred. The thickness of each link represented the weight or number of co-occurrence between two words. The more closely related the words were, the shorter the link distance.

Modularity analyses and network measures were conducted using Gephi (Bastian et al., 2009). *Modularity analysis* is a type of community (group) detection method that reveals the different clusters within a network (Blondel et al., 2008). The number of communities are indicated by the colors of the word labels. For example, if there are two different colors present in the network visualization, there are two main clusters within that network. Network measures, such as the network density, degree, and eigenvector centrality, were also calculated. These indicators provide a depiction of how central and how connected words are within the network. *Network density* refers to the number of connections divided by the total number of potential connections in the network. Network density refers to how intertwined the word concepts are, indicating how complex discussions are surrounding an issue. *Degree* refers to the number of links connecting each word. *Eigenvector centrality* indicates a word's relative influence or how central it is in the network. A high degree or a high eigenvector centrality indicates how central certain words are being discussed, revealing how salient certain keywords or topics are among users.

Lastly, sentiment analysis was conducted using ConText (Diesner, 2014). ConText uses a predefined lexicon map developed to classify the polarity of words. The sentiment classification was previously tested for reliability and upheld a high annotator agreement (Wilson, Wiebe, & Hoffmann, 2005). The edited tweets were entered into the software, where the words were tagged as either positive, negative, or neutral in sentiment. It is important to note that the sentiment analysis methods used examine the polarity of individual words, rather than the full context of the sentences. For example, the sentiment analysis detects *baby* as a neutral word; however, depending on the context, baby may have a positive (e.g. happy baby) or negative (e.g. crying baby) tone. The use of individual words to understand sentiment has value, and has been utilized in several studies (Calabrese et al., 2019; Jiang et al., 2018; Ruiz & Barnett, 2015); furthermore, by conducting the semantic networks, researchers can examine the sentiment of words by their context.

Results

Overall network

Figure 1 depicts the number of tweets posted by date during the nine-month time period. There was one substantial peak revolving around the number of tweets posted. This peaked near the end of November, which was around the time when the “CRISPR babies” news story occurred. Other smaller peaks occurred in mid-June, likely from reports that CRISPR may lead to cancer, and in mid-July, likely from reports that CRISPR may make unwanted DNA deletions.

Figure 2 shows the semantic network of the overall nine-month period of tweets. The average degree was 17.9, and the network density was .243, meaning almost a quarter of the words in the tweets co-occurred with each other. Table 1 displays the 25 most central terms along with their frequencies and degree.

The semantic network of all the tweets posted in the nine-month period can be seen in Figure 2. Four clusters were identified in this network. The largest cluster shared 73.3% of the network and was labeled in purple. This cluster represents the research and applications of gene editing, and its most central words include, *CRISPR*, *gene*, *editing*, *technology*, and *human*. The second largest cluster, in orange, accounted for 14.7% of the network. The theme for this cluster encompasses the “CRISPR babies” news story, and its most central words include, *scientist*, *baby*, *make*, *China*, and *create*. The third largest cluster, in green constituted 8.0% of the network, concerns agricultural regulations for biotechnology. This clusters’ most central words were *GMO*, *biotech*, *crop*, *FDA*, *animal*. Lastly, the smallest cluster accounted for 4.0% of the network. The theme of this cluster involves gene editing research with muscular dystrophy, with words, such as *dog*, *muscular*, *dystrophy*.

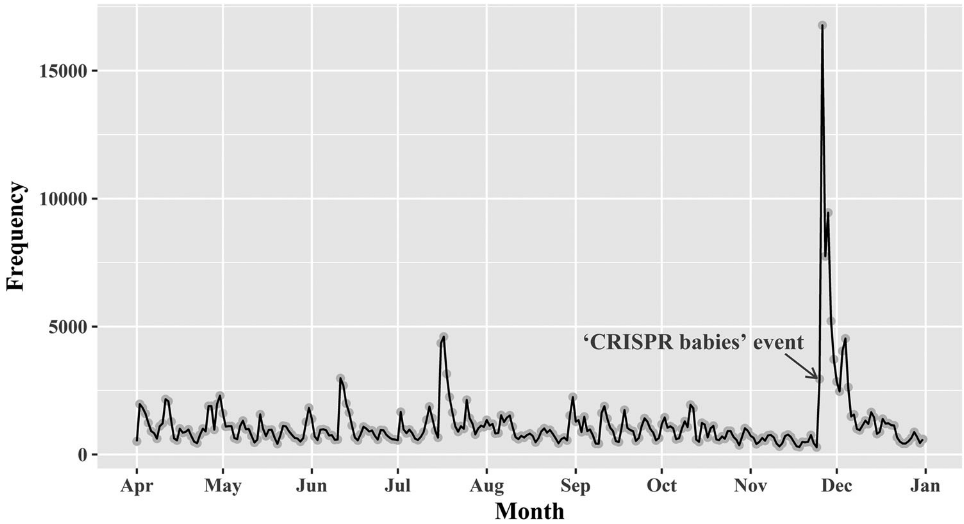

Figure 1. Line graph illustrating the number of tweets posted on “CRISPR” per month from 1 April 2018 to 31 December 2018.

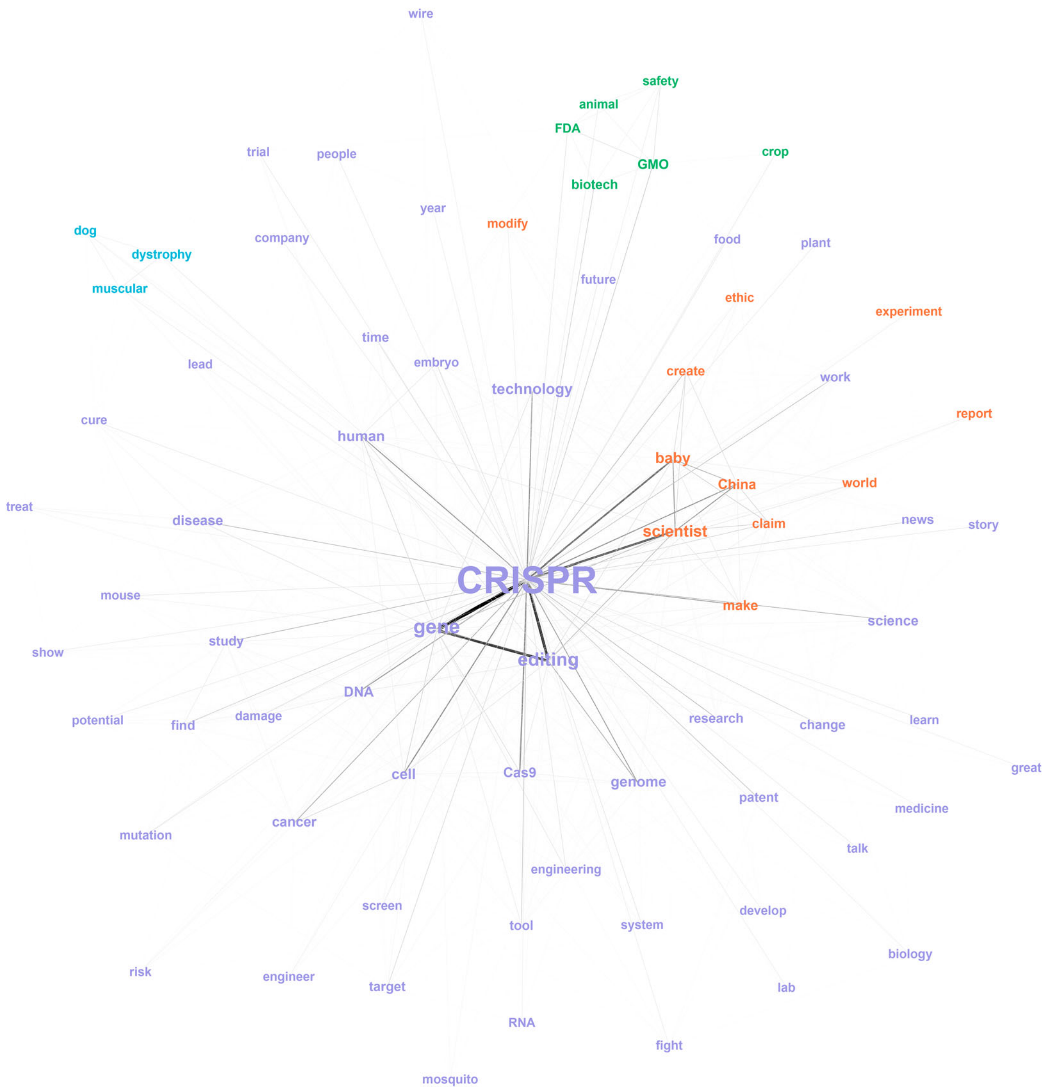

Figure 2. Semantic network representing "CRISPR" on Twitter from 1 April 2018 to 31 December 2018.

Sentiment analysis detected and tagged 205,284 words, equally balanced between positive and negative sentiment. Roughly 40.1% words had a positive tone; these words included *potential*, *great*, *ethical*, *improve*, and *drive*. About 39.6% of the words were negative in tone; among the words with negative sentiment, the most frequent included *cancer*, *disease*, *damage*, *trial*, and *virus*. The remaining 20.3% were neutral in tone, which included words such as *baby*, *nature*, *think*, *big*, and *idea*.

Major event networks

The number of original tweets posted per each one-week interval are as follows: 1,129 tweets (before stage), 17,960 tweets (initial stage), and 7,195 tweets (after event stage). To confirm Twitter users' attention to this event, the number of all tweets including retweets were assessed; there were 3,241 tweets (before stage), 48,708 tweets (initial stage), and 17,725 tweets (after event stage).

The semantic networks of each event stage can be seen in Figure 3(A–C). Table 2 reveals a summary of network measures for each of the three time periods. The largest difference can be

Table 1. Summary output of the semantic network analysis of CRISPR on Twitter.

#	Word	Frequency	Degree	Eigenvector centrality
1	CRISPR	120,003	74	1.000
2	gene	38,807	73	0.995
3	scientist	19,532	53	0.846
4	editing	30,372	54	0.845
5	human	10,087	50	0.798
6	technology	15,156	42	0.723
7	cell	11,352	36	0.652
8	DNA	9205	35	0.646
9	genome	10,934	34	0.620
10	disease	6773	33	0.617
11	baby	15,010	34	0.598
12	cancer	6857	30	0.584
13	make	7417	27	0.564
14	Cas9	8658	29	0.559
15	tool	4799	25	0.529
16	science	8917	26	0.520
17	China	9464	26	0.502
18	study	5810	24	0.491
19	research	5741	21	0.488
20	work	5179	23	0.481
21	create	4848	21	0.459
22	find	4199	22	0.459
23	target	3922	17	0.395
24	mutation	3068	17	0.382
25	claim	4011	15	0.371

seen between the first two networks. The average degree jumps from 11.29 to 20.71, and the network density increases from .257 to .471, indicating that the words of the initial stage network are more connected and more clustered together than the before stage network. Table 3 displays the top 25 most central terms in each event network along with their frequency and degree.

There were three clusters identified in the before stage network (Figure 3A). The largest cluster, in purple, accounted for 73.3% of the network. This theme involves general applications of CRISPR, and its most central words are *CRISPR*, *gene*, *editing*, *genome*, and *science*. The second cluster, in orange, accounted for 13.3% of the network. This theme revolves around CRISPR research on primary human T-cells, and its most central words are *cell*, *human*, *function*, *immune*, and *primary*. In green, the third cluster also accounted for 13.3% of the network. This theme revolves around the SLICE tool, a single guide RNA (sgRNA) lentiviral infection with Cas9 protein electroporation (Shifrut et al., 2018), and the network's most central words are *tool*, *system*, *slice*, *expand*, and *target*.

In the initial stage, two main clusters emerged (Figure 3B). Accounting for 71.1% of the network, the cluster labeled in pink refers to broader effects of the "CRISPR babies" news story, and its most central words are *CRISPR*, *human*, *gene*, *twin*, and *China*. The green cluster accounts for 28.9% of the network and refers to main details of the news story. Its most central words are *baby*, *scientist*, *claim*, *Chinese*, and *edit*.

Four clusters emerged from the after-event stage (Figure 3C). The largest cluster accounted for 58.7% of the network. Labeled in pink, this theme involves further developments in "CRISPR babies" news story. The most central words include *baby*, *CRISPR*, *scientist*, *twin*, and *edit*. The second largest cluster is labeled in green and accounts for 26.1% of the network. This theme refers to negative concerns over the "CRISPR babies" scandal, and the most central words were *create*, *ethical*, *infant*, *misstep*, and *worse*. The third largest cluster (10.9% of the network) is labeled in orange, and involves human applications of gene editing. The most central words include *gene, editing, human, embryo,* and *genome*. Lastly, the fourth cluster accounts for 4.4% of the network, and refers to genetic modification, with the words genetics and modify.

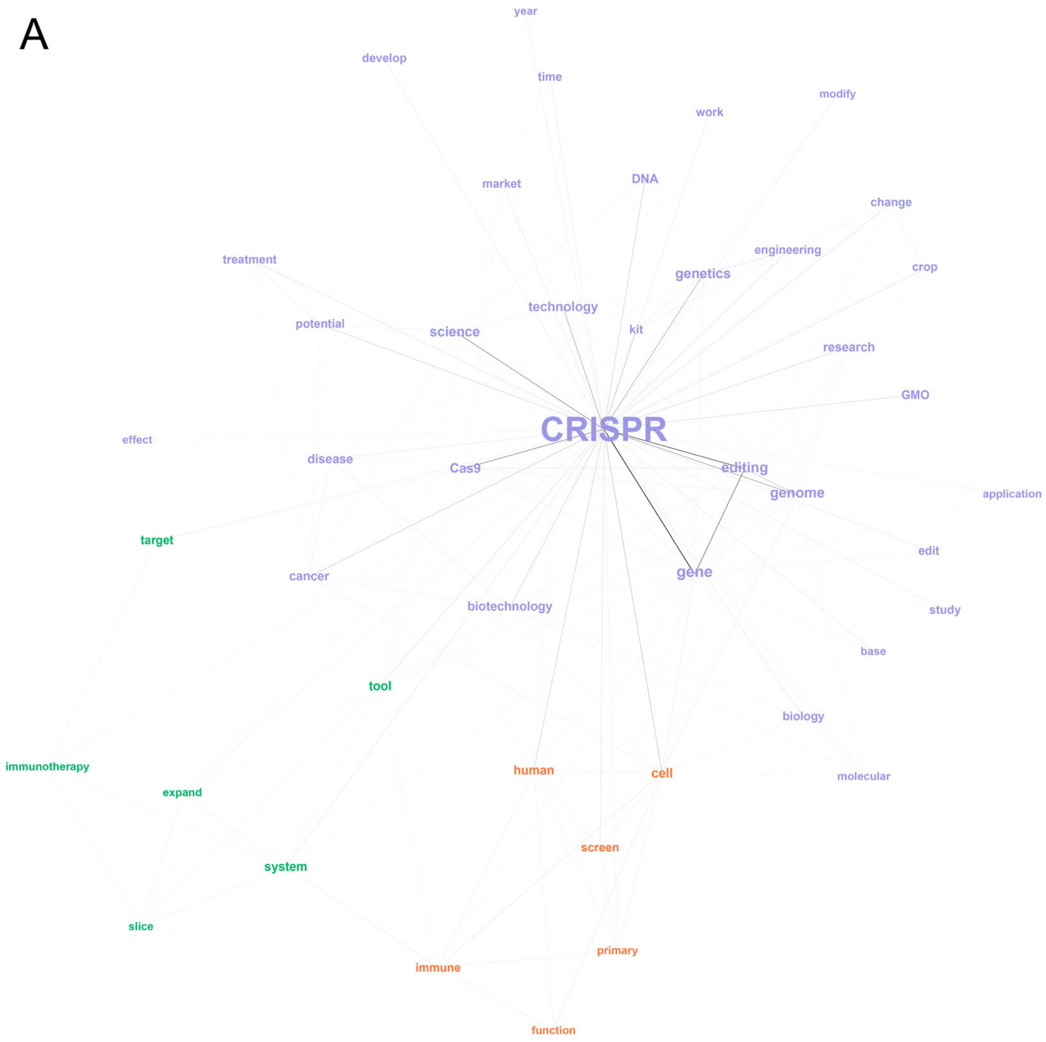

Figure 3. ABC. Semantic networks representing "CRISPR" on Twitter surrounding the gene-edited babies news story: (A) the before stage, (B) the initial stage, and (C) the after-event stage.

Sentiment analysis detected and tagged 1,735 words for the before stage. The proportion of positive words (46.3%) was higher than negative words (35.0%), while the remaining 18.7% were neutral in tone. In the initial stage, sentiment analysis tagged the polarity of 33,228 words. Most words were neutral in tone (38.0%), while there were slightly more negative words (33.0%) than positive (28.9%). For the after-event stage, about 14,873 words were detected. The plurality of words were negative in sentiment (42.2%), while there were more neutral (31.0%) than positive words (26.7%). These results are summarized in Figure 4.

Discussion

This study examined how Twitter presents information about CRISPR online through identifying the words, themes, and sentiment of all tweets, excluding retweets, related to the technology. Findings from this study reveal that most discussions on Twitter involve recent research and applications of gene editing. However, it is noteworthy that the presentations and emotions expressed by Twitter users differed before, during, and after an important event.

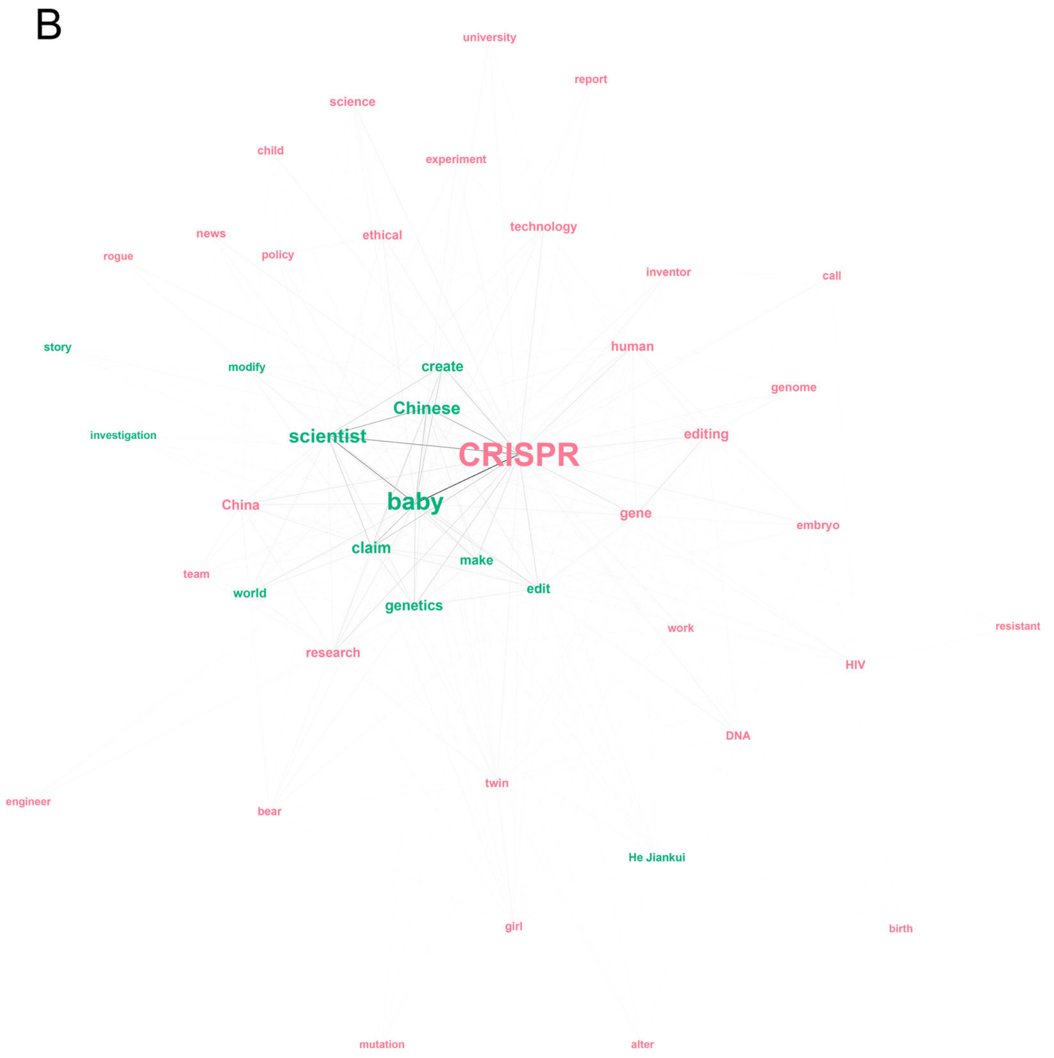

Figure 3 Continued

Overall network

From the overall network, the largest cluster referred to applications of gene editing. Most of the words referred to human applications, such as *medicine*, *cancer*, *cure*, and *human*. However, other applications were mentioned in the cluster, including agriculture (e.g. *food* and *plant*). The word *mosquito* may refer to recent research on gene drives to eradicate mosquitos carrying malaria (Kyrou et al., 2018), though there may be unanticipated ecological or environmental consequences (Saey, 2018). Overall, this cluster encompasses the research and several different applications of gene editing, indicating that the majority of online discourse on Twitter relates to the function of the technology, indicating that Twitter serves as a platform to share scientific knowledge.

Interestingly, the “CRISPR babies” theme took up a substantial amount of the overall nine-month network of tweets, despite only occurring within the last month. This highlights the significance of how certain extreme events may be in influencing the major themes and issues regarding new technologies. It also reveals that Twitter users may have great concerns on how these technologies should be used. These findings are similar to the results of public opinion surveys, where users express

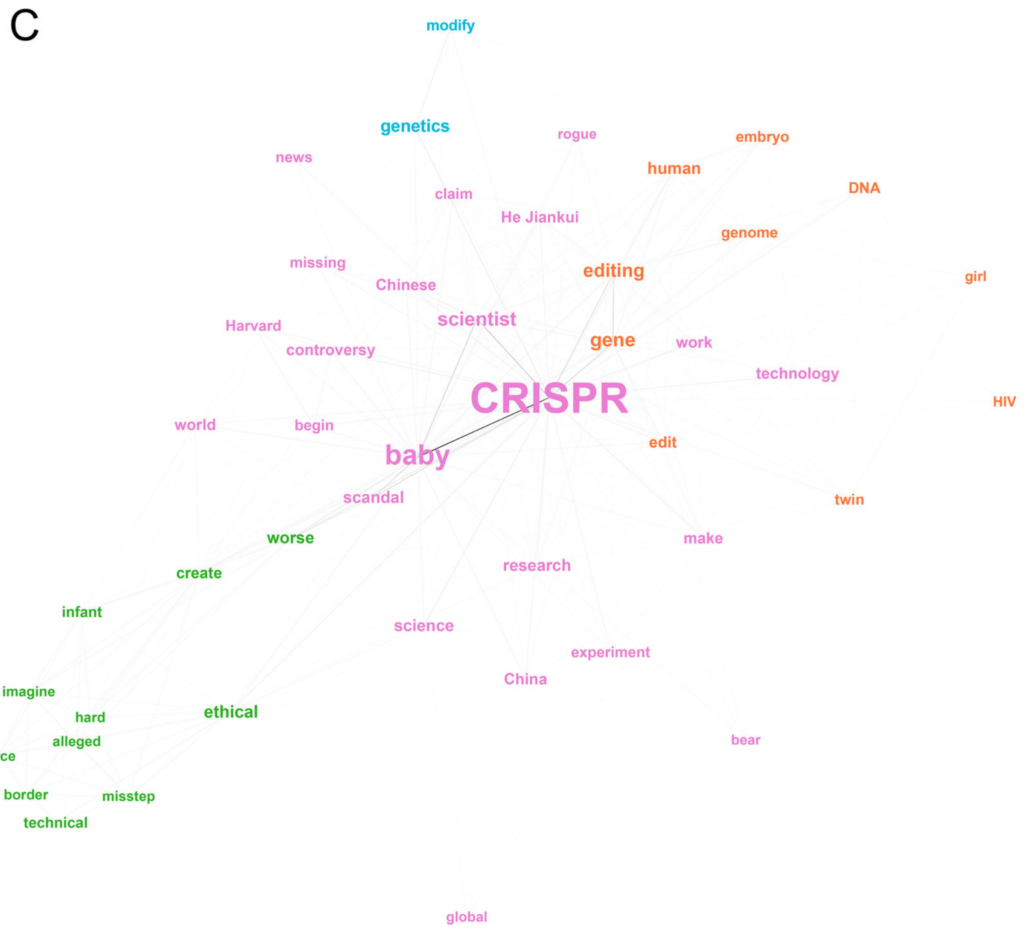

Figure 3 *Continued*

hesitation and general disagreement with using gene editing for embryo or germline enhancement (Blendon et al., 2016; Funk & Hefferon, 2018; Gaskell et al., 2017; Scheufele et al., 2017).

The last two themes refer to specific events that may have affected Twitter users. One theme involves applications of biotechnology in general. This may refer to the U. S. Department of Food and Agriculture's (2018) plan on regulating the use of biotechnology on animals and crops for ensuring the safety of its products. Because new regulations may greatly affect the agriculture industry and scientific research, users may find this issue very important. The second theme refers to advancements in muscular dystrophy research, specifically regarding the improvement of muscle tissue in a dog model (Amoasii et al., 2018). This highlights how events and recent advancements may be influential of people's perceptions of CRISPR.

The overall network sentiment analysis indicated that most words were either positive or negative in tone, indicating that Twitter users may already have strong views and perceptions about CRISPR

Table 2. Summary of network measures for each network of the event.

	Before	Initial	After
Average Degree	11.29	20.71	15.83
Average Weighted Degree	147.6	6066.0	1430.7
Graph Density	.257	.471	.352

Table 3. Summary output of semantic network analysis before, initial, and after event stages.

	Before			Initial			After		
#	Word	Degree	Eigenvector centrality	Word	Degree	Eigenvector centrality	Word	Degree	Eigenvector centrality
1	CRISPR	43	1	CRISPR	44	1	baby	40	1
2	gene	34	0.877	baby	43	0.990	CRISPR	39	0.996
3	editing[a]	24	0.687	scientist	39	0.938	scientist	32	0.904
4	cell	23	0.680	claim	35	0.882	gene	29	0.851
5	genome	19	0.595	human	34	0.881	editing[a]	27	0.772
6	science	19	0.585	Chinese	34	0.879	twin	24	0.755
7	human	18	0.564	gene	33	0.856	create	29	0.727
8	cancer	18	0.551	twin	32	0.844	human	22	0.704
9	Cas9	17	0.522	China	33	0.840	edit	22	0.690
10	tool	15	0.506	research	32	0.840	Chinese	21	0.679
11	system	16	0.502	edit	31	0.837	make	20	0.646
12	disease	12	0.434	editing[a]	31	0.823	claim	18	0.607
13	treatment	12	0.412	genetics	30	0.808	China	19	0.588
14	biotechnology	12	0.407	girl	29	0.778	research	17	0.587
15	technology	12	0.403	create	29	0.761	DNA	17	0.586
16	DNA	11	0.401	make	25	0.709	technology	17	0.550
17	potential	10	0.376	HIV	24	0.693	embryo	15	0.527
18	slice	11	0.355	embryo	23	0.686	genetics	15	0.518
19	genetics	11	0.343	science	22	0.641	girl	15	0.516
20	function	9	0.341	world	21	0.631	world	19	0.506
21	immune	10	0.323	DNA	20	0.615	work	14	0.486
22	research	8	0.306	bear	19	0.612	He Jiankui	14	0.484
23	primary	8	0.305	He Jiankui	18	0.588	genome	12	0.443
24	change	8	0.289	technology	20	0.569	modify	13	0.440
25	edit	6	0.288	modify	16	0.495	ethical	18	0.427

[a] Editing was distinguished from edit in reference to the word as a methodology.

technologies. Despite the prominent online information sources for gene editing being neutral in tone (Calabrese et al., 2019), our findings, which use the same sentiment analysis tool, reveal that the online perceptions of these technologies may be mixed. One explanation may be that users are ambivalent to CRISPR, expressing both the positive and negative aspects of the technology.

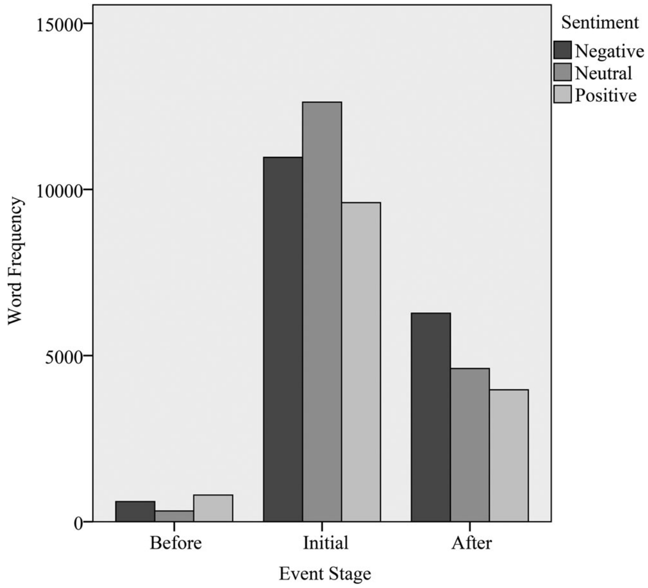

Figure 4. Bar graph illustrating the word frequency of the sentiment expressed grouped by event stage.

From the network alone, there is a mix of positive and negative words, such as *risk, great, potential, cure,* and *damage*. Users' describe their perceptions of the benefits, as well as the risks, of CRISPR. When a research article or news story is published, Twitter users may discuss its potential impact and implications. Another possible explanation may involve the polarization of the technology, where some may feel positively about the technology, while others feel negatively. Users may also fall victim to the mediatization of science (Peters, 2012), where news stories may report inaccurate or exaggerated science and health information to build hype (Dumas-Mallet, Smith, Boraud, & Gonon, 2018; Sumner et al., 2016). Such exaggerations could explain some of the positive or negative reactions toward some of the claims mentioned in news articles.

Major event networks

Based on examining the three stages of the event, there is a definite shift in themes after the news story broke. Before the story, the main themes referred to applications of gene editing in general, as well as recent advances in research. For example, two of the themes referred to a study that used a new technique called SLICE (Shifrut et al., 2018), which can improve the immune system's response to cancer (Alvarez, 2018). Once the "CRISPR babies" story broke, the main themes on Twitter shifted toward the details of the news event. This demonstrates how the impact of the news event, since applications and recent research of gene editing did not emerge as themes in this network. In the after-event stage, the clusters still primarily focused on the news story, including its aftermath and negative consequences. For example, the words *missing*, *Chinese*, and *scientist*, may refer to reports of He Jiankui going missing (Haynes, 2018). In addition, the words *Harvard* and *begin* appeared in one of the frames, possibly referring to the public's reaction toward the university's initiation of research on gene-edited sperm, though it was reported by news media earlier (Regalado, 2018a). Our findings from the after-event network indicate that the event may have influenced Twitter users' perceptions toward CRISPR. The overall shift in themes toward the "CRISPR babies" story reveals how powerful certain events can shift one's perceptions toward new technologies.

Through the sentiment analysis for each event time period, there is a gradual change toward a negative tone as time progresses. Initially, the majority of words are positive or negative, similar to the findings from the nine-month sentiment analysis. However, once the news story broke, a shift happened where neutral and negative words were more prominent. A week later, the words were mostly negative in tone. These findings indicate that as time passed, Twitter users posted more negative-valanced tweets likely due to their own evaluation and impact of the breaking news story. Rather than reporting the event, Twitter users are now voicing their opinion and evaluating the consequences of the event. Diffusion of innovations theory may explain the slow spread of information and opinions at the start of the event before the majority of Twitter users begin to express negative emotions (Rogers, 2010; Vishwanath & Barnett, 2011). The theory explains how a phenomenon is spread through a system over time, where a new idea is at first adopted slowly; the idea begins to spread faster as more and more become aware and adopt it. In this way, the slow increase in negative emotion could be due to processes explained by the adoption of a new idea. Moreover, this phenomenon can be explained by the complex, dynamic role of both online media and traditional press and broadcast on agenda setting (McCombs, 2004). Traditional media may highlight certain aspects of a topic, which then may influence how other media may pick up on the topic and how they may make that information more salient. In our highly saturated media environment, these different outlets may influence each other over time. Thus, negative emotion expressed on one type of media may then have influenced the negative emotion expressed on Twitter.

Overall, this study revealed the themes and sentiment of Twitter users' posts surrounding CRISPR. While a large portion of the overall network involved the "CRISPR babies" event, three other themes, human and agricultural applications, regulations, and scientific research advancements, were also present. This suggests that there is some indication of discussions on CRISPR's impacts on society, agriculture, and the environment; it provides a starting point for developing engagement strategies with the general public around the applications of CRISPR technologies. Environmental benefits and concerns were not directly mentioned, and though our study examined the posts of Twitter users, this may be an indication that there is still a need to engage the general public with these issues.

In addition, the case study revealed that the news story may have influenced what people posted on Twitter through the differences between themes and the increase in negative emotion sentiment as time passed. Future research should examine message strategies to effectively address the concerns over negative news events. This may aid in improving the people's perception of CRISPR, or gene editing technologies in general, as well as preventing any potential negative impacts.

Limitations

This study is not without limitations. Twitter is just one of many social media platforms that people use to express their views on specific issues; this study does not claim to represent the entire online media environment or general population (Wojcik & Hughes, 2019), but does provide some insight into what people are expressing on this specific platform. Future work should examine how CRISPR is portrayed on other social media platforms, as well as through traditional media, such as newspapers or television. In addition, another potential limitation was the use of only one search term "CRISPR" in our analyses. While the use of several search terms would increase the sample size of posts, we chose "CRISPR" based only on our study's aims and to limit and remove posts irrelevant to the new, emerging gene editing technology. Future work should examine online discourse on other gene editing technologies. We examined Twitter users' perception of CRISPR during specific time intervals on Twitter. Though we depicted the differences of the main themes and the sentiment during the before, initial, and after-event stages of a major event, we cannot make any causal inferences regarding public opinion. Future work should examine how public perceptions of gene editing change over time, especially with regard to the long-term effects of events. It would also be beneficial to examine the main themes on Twitter by partitioning users by their stakeholder group, such as scientists, agricultural workers or policymakers. This would help identify and address the key concerns surrounding gene editing technologies among each community. Similar steps are currently being taken, as our research group is conducting a national survey on the general public and other stakeholder groups' knowledge, perceptions, and associations with gene editing and CRISPR. Furthermore, this research was conducted under the assumption that people are posting these messages themselves. Because retweets may provide redundant information and possible exposure to bots, which would have deterred from our main goal to analyze Twitter users' perceptions of CRISPR online, they were not included in the analyses. Lastly, although the sentiment analysis algorithm has been empirically tested and validated in previous research (Wilson et al., 2005), there is a chance that some classification of words are not accurate based on their context, as with any computational tool that measures sentiment. There is still value in understanding sentiment at the individual word level. Future research may examine sentiment through advanced computational techniques beyond the scope of this study (Danowski & Riopelle, 2018).

Conclusion

This study examined the main words and themes of CRISPR on Twitter using semantic network analysis. Our findings indicate that the presentation of information about CRISPR on Twitter, can inform scientists about the different opinions and aspects of an issue, which may aid scientists to better address issues and concerns surrounding these technologies. In addition, while our study found that the sentiment toward CRISPR was mixed, researchers should continue to monitor the presentation of technologies on social media platforms, especially around major events. This will allow researchers to gain an understanding of the concerns over new technologies, and effectively address issues through subsequent messaging strategies.

Note

1. The total number of tweets, including retweets, was 328,137 (overall network), 3,241 (before stage), 48,708 (initial stage), and 17,725 (after-event stage). The total number of tweets, excluding retweets, was 135,578 (overall network), 1,129 (before stage), 17,960 (initial stage), and 7,195 (after-event stage).

Disclosure statement

No potential conflict of interests were reported by the authors.

Funding

This work is supported by the Innovative Genomics Institute at the University of California [Innovative Genomics Institute Society Research Award].

ORCID

Christopher Calabrese http://orcid.org/0000-0001-6216-0489

References

Alvarez, J. (2018, November 15). New ‘SLICE’ tool can massively expand immune system’s cancer-fighting repertoire. *UCSF News Center*. Retrieved from https://www.ucsf.edu/news/2018/11/412356/new-slice-tool-can-massively-expand-immune-systems-cancer-fighting-repertoire

Amoasii, L., Hildyard, J. C. W., Li, H., Sanchez-Ortiz, E., Mireault, A., Caballero, D., … Olson, E. N. (2018). Gene editing restores dystrophin expression in a canine model of Duchenne muscular dystrophy. *Science*, *362*(6410), 86–91. doi:10.1126/science.aau1549

Bastian, M., Heymann, S., & Jacomy, M. (2009). *Gephi: An open source software for exploring and manipulating networks*.

Bengtsson, N. E., Hall, J. K., Odom, G. L., Phelps, M. P., Andrus, C. R., Hawkins, R. D., … Chamberlain, J. S. (2017). Muscle-specific CRISPR/Cas9 dystrophin gene editing ameliorates pathophysiology in a mouse model for Duchenne muscular dystrophy. *Nature Communications*, *8*, 14454. doi:10.1038/ncomms14454

Binder, A. R. (2012). Figuring out #fukushima: An initial look at functions and content of US twitter commentary about nuclear risk. *Environmental Communication*, *6*(2), 268–277. doi:10.1080/17524032.2012.672442

Blendon, R. J., Gorski, M. T., & Benson, J. M. (2016). The public and the gene-editing revolution. *New England Journal of Medicine*, *374*(15), 1406–1411. doi:10.1056/NEJMp1602010

Blondel, V. D., Guillaume, J.-L., Lambiotte, R., & Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, *2008*(10), P10008. doi:10.1088/1742-5468/2008/10/p10008

Brossard, D. (2013). New media landscapes and the science information consumer. *Proceedings of the National Academy of Sciences of the United States of America*, *110*, 14096–14101. doi:10.1073/pnas.1212744110

Brossard, D., & Scheufele, D. A. (2013). Science, new media, and the public. *Science*, *339*(6115), 40–41. doi:10.1126/science.1232329

Cacciatore, M. A., Scheufele, D. A., & Corley, E. A. (2014). Another (methodological) look at knowledge gaps and the internet’s potential for closing them. *Public Understanding of Science*, *23*(4), 376–394. doi:10.1177/0963662512447606

Calabrese, C., Anderton, B. N., & Barnett, G. A. (2019). Online representations of “genome editing” uncover opportunities for encouraging engagement: A semantic network analysis. *Science Communication*, *41*(2), 222–242. doi:10.1177/1075547018824709

Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences*, *24*(1), 87–114. doi:10.1017/S0140525X01003922

Cowan, N. (2016). *Working memory capacity: Classic edition*. New York: Routledge.

D’Angelo, P. (2002). News framing as a multiparadigmatic research program: A response to Entman. *Journal of Communication*, *52*(4), 870–888. doi:10.1111/j.1460-2466.2002.tb02578.x

Danowski, J., & Riopelle, K. (2018). *Sentiment in semantic networks and effects on media content and volume*. Paper presented at the North American Social Networks Conference, Washington, D.C.

Diesner, J. (2014). *ConText: Software for the integrated analysis of text data and network data*. Paper presented at the Social and Semantic Networks in Communication Research. Preconference at Conference of International Communication Association (ICA), Seattle, WA.

Doğu, B. (2017). Environment as politics: Framing the Cerattepe protest in Twitter. *Environmental Communication*, *13*(5), 617–632. doi:10.1080/17524032.2017.1406384

Dudo, A., Brossard, D., Shanahan, J., Scheufele, D. A., Morgan, M., & Signorielli, N. (2011). Science on television in the 21st century: Recent trends in portrayals and their contributions to public attitudes toward science. *Communication Research*, *38*(6), 754–777. doi:10.1177/0093650210384988

Dumas-Mallet, E., Smith, A., Boraud, T., & Gonon, F. (2018). Scientific uncertainty in the press: How newspapers describe initial biomedical findings. *Science Communication*, *40*(1), 124–141. doi:10.1177/1075547017752166

Feinerer, I., Hornik, K., & Meyer, D. (2008). *Text mining infrastructure in R* (Version 2008-03-18). Retrieved from https://www.jstatsoft.org/v025/i05

Fillmore, C. J. (1982). *Linguistics in the morning calm. Linguistics Society of Korea. Frame Semantics*. Seoul: Hanshin.

Freeman, L. C. (1978). Centrality in social networks conceptual clarification. *Social Networks*, *1*(3), 215–239. doi:10.1016/0378-8733(78)90021-7

Funk, C., & Hefferon, M. (2018). *Public views of gene editing for babies depend on how it would be used.* Retrieved from http://www.pewresearch.org/science/2018/07/26/public-views-of-gene-editing-for-babies-depend-on-how-it-would-be-used/

Gaskell, G., Bard, I., Allansdottir, A., da Cunha, R. V., Eduard, P., Hampel, J., … Zwart, H. (2017). Public views on gene editing and its uses. *Nature Biotechnology*, *35*, 1021–1023. doi:10.1038/nbt.3958

Haynes, S. (2018, December 3). The Whereabouts of China's Gene-Editing Doctor He Jiankui Remain Unknown. *TIME*. Retrieved from http://time.com/5469111/he-jiankui-scientist-missing-gene-edited-babies/

Jacomy, M., Venturini, T., Heymann, S., & Bastian, M. (2014). Forceatlas2, a continuous graph layout algorithm for Handy network visualization designed for the Gephi software. *PLOS ONE*, *9*(6), e98679. doi:10.1371/journal.pone.0098679

Jiang, K., Anderton, B. N., Ronald, P. C., & Barnett, G. A. (2018). Semantic network analysis reveals opposing online representations of the search term "GMO". *Global Challenges*, *2*(1), 1700082. doi:10.1002/gch2.201700082

Kohl, P. A., Brossard, D., Scheufele, D. A., & Xenos, M. A. (2019). Public views about gene editing wildlife for conservation. *Conservation Biology*, *33*(6), 1286–1295. doi:10.1111/cobi.13310

Kyrou, K., Hammond, A. M., Galizi, R., Kranjc, N., Burt, A., Beaghton, A. K., … Crisanti, A. (2018). A CRISPR–Cas9 gene drive targeting doublesex causes complete population suppression in caged Anopheles gambiae mosquitoes. *Nature Biotechnology*, *36*, 1062–1066. doi:10.1038/nbt.4245

Li, T., Liu, B., Spalding, M. H., Weeks, D. P., & Yang, B. (2012). High-efficiency TALEN-based gene editing produces disease-resistant rice. *Nature Biotechnology*, *30*, 390–392. doi:10.1038/nbt.2199

Marchione, M. (2018, November 26). Chinese researcher claims first gene-edited babies. *Associated Press*. Retrieved from https://www.apnews.com/4997bb7aa36c45449b488e19ac83e86d

McCombs, M. (2004). *Setting the agenda: The mass media and public opinion*. Cambridge, UK: Polity Press.

Miller, G. A. (1956). The magical number seven plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, *63*(2), 81–97. doi:10.1037/h0043158

Min, Y.-L., Li, H., Rodriguez-Caycedo, C., Mireault, A. A., Huang, J., Shelton, J. M., … Olson, E. N. (2019). CRISPR-Cas9 corrects Duchenne muscular dystrophy exon 44 deletion mutations in mice and human cells. *Science Advances*, *5*(3), eaav4324. doi:10.1126/sciadv.aav4324

Mitchell, A., Funk, C., & Gottfried, J. (2017). *Science news and information today*. Retrieved from https://www.courthousenews.com/wp-content/uploads/2017/09/pew-science-news.pdf

Neve, P. (2018). Gene drive systems: Do they have a place in agricultural weed management? *Pest Management Science*, *74*(12), 2671–2679. doi:10.1002/ps.5137

Ooms, J. (2014). The jsonlite package: A practical and consistent mapping between json data and r objects. *arXiv preprint arXiv:1403*.2805.

Peters, H. P. (2012). Scientific Sources and the Mass Media: Forms and Consequences of Medialization. In S. Rödder, M. Franzen, & P. Weingart (Eds.), *The Sciences' Media Connection –Public Communication and its Repercussions* (pp. 217–239). Dordrecht: Springer Netherlands.

Regalado, A. (2018a, November 29). Despite CRISPR baby controversy, Harvard University will begin gene-editing sperm. *MIT Technology Review*. Retrieved from https://www.technologyreview.com/s/612494/despite-crispr-baby-controversy-harvard-university-will-begin-gene-editing-sperm/

Regalado, A. (2018b, November 25). EXCLUSIVE: Chinese scientists are creating CRISPR babies. *MIT Technology Review*. Retrieved from https://www.technologyreview.com/s/612458/exclusive-chinese-scientists-are-creating-crispr-babies/

Rinker, T. W. (2019). *Quantitative discourse analysis package*. Buffalo, New York. Retrieved from http://github.com/trinker/qdap

Rogers, E. M. (2010). *Diffusion of innovations*. New York: Simon and Schuster.

Ruiz, J. B., & Barnett, G. A. (2015). Exploring the presentation of HPV information online: A semantic network analysis of websites. *Vaccine*, *33*(29), 3354–3359. doi:10.1016/j.vaccine.2015.05.017

Saey, T. H. (2018, September 24). In lab tests, this gene drive wiped out a population of mosquitoes. *ScienceNews*, *194*, 6.

Saplakoglu, Y. (2017, February 19). New gene drive technology could wipe out malaria, but is it safe? *Science Magazine*. Retrieved from https://www.sciencemag.org/news/2017/02/new-gene-drive-technology-could-wipe-out-malaria-it-safe

Scheufele, D. A., Xenos, M. A., Howell, E. L., Rose, K. M., Brossard, D., & Hardy, B. W. (2017). U.S. Attitudes on human genome editing. *Science*, *357*(6351), 553–554. doi:10.1126/science.aan3708

Shifrut, E., Carnevale, J., Tobin, V., Roth, T. L., Woo, J. M., Bui, C. T., … Marson, A. (2018). Genome-wide CRISPR screens in primary human T cells reveal key regulators of immune function. *Cell*, *175*(7), 1958–1971.e1915. doi:10.1016/j.cell.2018.10.024

Sumner, P., Vivian-Griffiths, S., Boivin, J., Williams, A., Bott, L., Adams, R., … Chambers, C. D. (2016). Exaggerations and caveats in press releases and health-related science news. *PLOS ONE*, *11*(12), e0168217. doi:10.1371/journal.pone.0168217

Tang, L., Bie, B., & Zhi, D. (2018). Tweeting about measles during stages of an outbreak: A semantic network approach to the framing of an emerging infectious disease. *American Journal of Infection Control*, *46*(12), 1375–1380. doi:10.1016/j.ajic.2018.05.019

U. S. Food and Drug Administration. (2018). *Plant and animal biotechnology innovation action plan*. Silver Spring, MD. Retrieved from https://www.fda.gov/downloads/Safety/Biotechnology/UCM624517.pdf

Veltri, G. A. (2012). Microblogging and nanotweets: Nanotechnology on Twitter. *Public Understanding of Science*, *22*(7), 832–849. doi:10.1177/0963662512463510

Vishwanath, A., & Barnett, G. A. (2011). *The diffusion of innovations: A communication science perspective*. New York, NY: Peter Lang.

Wang, Y., Cheng, X., Shan, Q., Zhang, Y., Liu, J., Gao, C., & Qiu, J.-L. (2014). Simultaneous editing of three homoeoalleles in hexaploid bread wheat confers heritable resistance to powdery mildew. *Nature Biotechnology*, *32*, 947–951. doi:10.1038/nbt.2969. https://www.nature.com/articles/nbt.2969#supplementary-information

Wasserman, S., & Faust, K. (1994). *Social network analysis: Methods and applications* (Vol. 8). New York: Cambridge University Press.

Wilson, T., Wiebe, J., & Hoffmann, P. (2005). *Recognizing contextual polarity in phrase-level sentiment analysis*. Paper presented at the Conference on Human Language Technology and Empirical Methods in Natural Language Processing, Vancouver, British Columbia, Canada.

Wojcik, S., & Hughes, A. (2019). *Sizing up Twitter users*. Washington, DC: Pew Internet & American Life Project.