Open Forum Infectious Diseases

BRIEF REPORT

What Twitter Can Tell Us About #IDWeek2020

Richard J. Medford[1,2] **and Sameh N. Saleh**[2,3]

[1]Division of Infectious Diseases, University of Texas Southwestern Medical Center, Dallas, Texas, USA, [2]Clinical Informatics Center, University of Texas Southwestern Medical Center, Dallas, Texas, USA, and [3]Department of Internal Medicine, University of Texas Southwestern Medical Center, Dallas, Texas, USA

We used topic modeling, subjectivity analysis, and social graph theory to analyze 11 944 tweets relating to IDWeek 2020. Twitter is a rich medium that can successfully disseminate knowledge and allow users to engage in social networks during a medical conference, despite a virtual format.

Keywords. IDWeek 2020; machine learning; natural language processing; social network analysis; social media; Twitter; topic modeling.

The social media microblogging platform Twitter has been used previously to study the benefits and engagement of users at medical conferences [1–3]. Tweets and hashtags allow users to engage in discourse around certain topics and find common areas of interest. Twitter can be used to disseminate important scientific knowledge [4, 5] and reach a wide audience, especially those who are unable to attend in person or are present but at a different session [3]. Researchers have performed analyses to identify what constitutes a successful tweet and have examined factors associated with likelihood of a retweet (ie, presence of media or a hyperlink), a marker of impact and dissemination [2, 3]. Furthermore, in 2019, novel medical education tools within Infectious Diseases such as ID Journal Club (@IDJClub) have been implemented on the Twitter platform [6]. With the transition of IDWeek 2020 (October 21 to October 25, 2020) to a completely virtual event due to the coronavirus disease 2019 (COVID-19) pandemic, we sought to evaluate engagement of users, social networks, and topics discussed on Twitter to understand the characteristics of this new conference forum. We hypothesize that these data could be valuable to the IDWeek planning committee and may aid Twitter users, new or old, in determining accounts or users of interest.

METHODS

From October 21 to 25, 2020, we extracted English language tweets using the keyword "IDWeek 2020" leveraging Twitter's Application Programming Interface (API) via RStudio (version 1.3.1093). We collected 89 variables associated with each tweet, of which 10 are used in this analysis. Using a previously described method for tweet analysis, we processed and transformed the data into a usable format and explored basic characteristics [4].

Patient Consent Statement

Institutional review board approval was not required for this study, as we used only publicly available data. There were no patients in this study and therefore no factors necessitating patient consent.

To understand the content of tweets, we applied several natural language processing techniques. To perform topic modeling (gensim Python [version 3.6.1] package), we used an unsupervised machine learning model (Latent Dirichlet Allocation) that automatically generates topics based on grouping of similar observations or tweets. We iteratively trained the model to identify the ideal number of topics and document the top 20 weighted terms that contribute to each topic. We then manually labeled each topic based on the input of both authors. Tweets can contain more than 1 topic, but there is always a single predominant topic.

Social network analysis uses graph theory to analyze the social structures within a community by connecting nodes (ie, people) with edges (ie, connections or relationships between nodes). These are visually depicted as a circle and a line connecting 2 circles, respectively [7]. Based on tweet volume and followers, we display the social network of the top 2 "influencers" as indicated through the Symplur website at the time of accession [8], visualizing nodes as Twitter users and edges as the interactions with another user, through likes, retweets, or follows, graphed with the Fruchterman-Reingold algorithm via Gephi software (version 0.9.2). Darker or thicker lines are weighted and represent increased interactions between users.

To further explore social networks and impact, we first use a circular diagram to highlight only the retweet network among Twitter accounts. Each line abutting the periphery of the circle (nodes) represents a unique user, and lines connecting users (edges) signify a retweet. Similarly, darker or thicker lines represent increased numbers of retweets. We set the minimum number of retweet mentions at 75 and label the ensuing top 18

Received 2 November 2020; editorial decision 10 December 2020; accepted 11 December 2020.
Correspondence: R. J. Medford, MD FRCP(C), 5323 Harry Hines Blvd, Dallas, TX 75390 (richard.medford@utsouthwestern.edu).

Open Forum Infectious Diseases®2021
© The Author(s) 2020. Published by Oxford University Press on behalf of Infectious Diseases Society of America. This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs licence (http://creativecommons.org/licenses/by-nc-nd/4.0/), which permits non-commercial reproduction and distribution of the work, in any medium, provided the original work is not altered or transformed in any way, and that the work is properly cited. For commercial re-use, please contact journals.permissions@oup.com
DOI: 10.1093/ofid/ofaa621

users based on the number of retweet mentions, with label font size being proportional to number of nodes.

Second, we identify the top 20 users based on number of followers and perform a subjectivity analysis on the aggregate of their respective IDWeek tweets through the use of a recurrent neural network [9]. We provide a median subjectivity score on a scale of 0 (completely objective) to 1 (completely subjective). Typically, an objective tweet conveys factual information, whereas a subjective tweet relays an opinion or belief.

RESULTS

Tweet Characteristics

A total of 11 944 tweets from 3603 unique users were collected during our study period. Of these, 6092 were retweets. Users were more likely to use Twitter with an iPhone (5001), followed by the web app (4230) and an Android device (1711). The number of tweets per hour peaked on October 21 at 652 tweets, coinciding with the 24-hour COVID-19 Chasing the Sun event. In anticipation of IDWeek, a small subset of tweets (699) were collected before October 21. Similarly, 849 tweets were extracted after the completion of the event.

Topic Modeling

After training the model, we identified 8 topics as the ideal topic number and labeled the following themes in order of frequency: (1) the COVID-19 pandemic and the path forward, (2) Sharing IDWeek 2020 presentations and content, (3) antimicrobial stewardship, (4) health policy and Infectious Diseases, (5) SARS-CoV-2 transmission, (6) highlighted speakers and the COVID-19 pandemic, (7) clinical discussions: HIV and *Staphylococcus* treatment, and (8) the COVID-19 pandemic response in Europe and the United States (Table 1). Discussion of the COVID-19 pandemic and the path forward emerged as the most prevalent topic (27.0% of tweets), followed by sharing IDWeek 2020 presentations and content (18.5%) and

Table 1. The 20 Terms (in Order of Weighting) Associated With the 8 Abstract Topics (Labeled by Authors), Tweet Frequency, and a Representative Tweet

Possible Topic Label	Tweets per Topic	Words Contributing to Topic Model (in Descending Order of Weighting)	Representative Tweet
The COVID-19 pandemic and the path forward	2710 (27.0)	good, effective, idsainfo, great, justice, fauci, people, thank, start, need, normal, forward, carlosdelrio, move, new, chat, evidence, see, world, vaccine	"COVID-19...has revealed issues with health justice, social justice, climate justice.... Our children deserve a better world.... We need to start moving forward, not to a new normal...[but] towards a more sustainable and better normal in the future @DrMikeRyan #IDWeek2020 *[media]*"
Sharing IDWeek 2020 presentations and content	1857 (18.5)	check, presentation, idtwitter, virtual, patient, abstract, present, ever, see, great, twitter, want, work, test, new, link, oral, pediatric, thread, viral	"Do you know What's Hot in ID & HIV? Be sure to view this excellent session at @IDWeek2020 moderated by @Armstrws and with talks by @DrJLi @BensonCb32 @DrCindySears @PaulPauwaert"
Antimicrobial stewardship	1290 (12.8)	use, antibiotic, stewardship, treatment, idsainfo, patient, work, team, infections, patients, infection, present, check, antimicrobial, shea_epi, disease, help, opat, idtwitter, study	"Thanks to our team @NM_IDSteward & @MWUCCP, led by @cecruce & @njrhodes_rx, we found that broad-spectrum Gram-neg abx use in patients with CAP was sig associated (odds 1:2) with development of healthcare-associated CDI within 84 days of admit *[link]* #IDWeek2020 *[media] [link]*"
Health policy and Infectious Diseases	1278 (12.7)	health, policy, need, study, time, lecture, community, show, hiv, day, impact, infection, among, influenza, see, chasingthesun, adaadimora, public, antibiotic, give	"Dr. Mark Rupp from @UNMC_ID: Speaking to the politicization of public health; what can hospital epidemiologists (& ID clinicians) do? 1) Stay true to science!! Things will eventually autocorrect 2) Speak up in support of public health #IDWeek2020 @IDWeek2020 @SHEA_Epi"
SARS-CoV-2 transmission	978 (9.7)	transmission, chasingthesun, airborne, droplet, idsainfo, great, long, important, contact, overview, weber, happen, possible, direct, range, uncommon, live, infection, care, mask	"Such a great overview of the 'airborne' vs. droplet transmission of #COVID19 by Dr. Weber. Long-range airborne transmission doesn't happen, transmission >6 ft possible but uncommon, droplet and direct contact most important. @IDSAInfo #IDWeek2020 #chasingthesun *[media]*"
Highlighted speakers and the COVID-19 pandemic	917 (9.2)	health, pandemic, fauci, drtomfrieden, carlosdelrio, way, vaccine, year, even, infectious_disease, watch, shea_epi, public, kick, protect, forget, chasingthesun, idtwitter, spread, idsainfo	"Anthony Fauci's address at #IDWeek2020 to kick off a 24-hour session on #COVID19 is preceded by words of support from George W. Bush, NIH Director Francis Collins, @DrTedros and others. 'Tony has become the voice of truth on the COVID-19 pandemic,' Collins says"
Clinical discussions: HIV and *Staphylococcus* treatment	643 (6.4)	disease, hiv, present, severe, oral, study, staphylococcus, learn, virus, treatment, risk, age, abstract, say, one, hominis, live, epidermidis, idnerdhumor, lugdenensis	"Fascinating #IDWeek2020 talk on S. aureus bacteremia lessons over past 20 years I loved the preclinical research on DNMT3A (my inner AML nerd was screaming) A/C genotype ⬆ promoter methylation ⬇ serum IL-10 ⏭ may improve MRSA bacteremia resolution Also, 👋 life lessons 👂 *[media]*"
The COVID-19 pandemic: Europe and the USA	379 (3.8)	usa, pandemic, health, fauci, wear, avoid, hit, much, crowds, hard, stay, country, europe, carlosdelrio, mask, disparity, rate, wash, since, major	"Dr. Fauci on #COVID19: 1-Worst pandemic since 1918 2-USA hit very hard 3-USA did not shut down as much countries in Europe & thus did not have major ⬇ in baseline rate 4-Wear a mask! Avoid crowds, esp indoors. Wash hands. Stay 6 feet apart #IDWeek2020 @IDWeek2020"

Abbreviations: COVID-19, coronavirus disease 2019; SARS-CoV-2, severe acute respiratory syndrome coronavirus 2.

antimicrobial stewardship (12.8%). The COVID-19 pandemic response in Europe and the United States was the least common topic discussed, with 3.8% of tweets.

Social Network Analysis

The top 18 users based on retweet network were (alphabetical order): chrltmrshl, CarlosdelRio7, DocWoc71, DrJRMarcelin, Dr_Mike_Stevens, ErinMcCreary, FranciscoMarty_, GermHunterMD, IDSAInfo, IDWeek2020, JGPharmD, PIDSociety, SAIRABT, SaraKellerMD1, SIDPharm, SPInfectologia, SteveRustad1, and Vaccinologist. As visualized in Figure 1A, thousands of darker weighted magenta lines (edges) can be seen connecting from a peripheral node to 1 of these top 18 labeled nodes, indicating increased retweet mentions. Figure 1B demonstrates the interactions of 2 of these user accounts (CarlosdelRio7 and GermHunterMD)

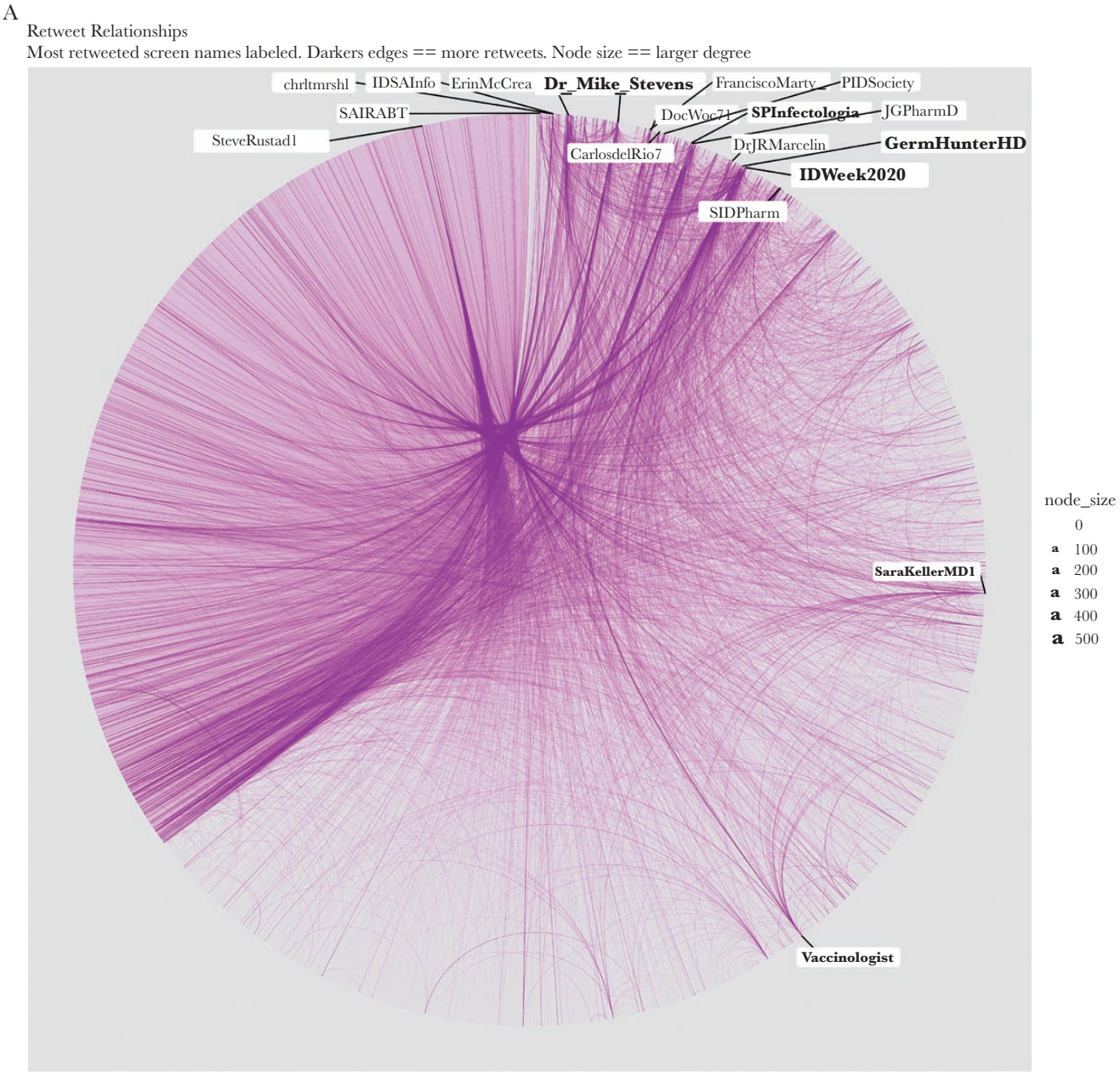

Figure 1. A, Retweet network: Top 18 usernames depicted by number of retweet mentions. Larger font size of the label corresponds to an increase in nodes (number of times that username was mentioned in a retweet). B, User network diagram of 2 active Infectious Diseases Twitter users (@CarlosdelRio7, @GermHunterMD). Nodes (circles in gray) are labeled with the user account handle, and edges (pink lines connecting circles) represent interaction between those users. Darker lines represent increased interaction, and depicted are 5 large clusters of activity. Note, 3 other nodes (@alvie_bar, @nynursesunited, @universalmaski2) emerge as active participants of the network. C, Subjectivity analysis of the top 20 users with a minimum of 5 IDWeek 2020 tweets based on follower network. A subjectivity score of 0 represents complete objectivity, while a score of 1 represents complete subjectivity. Whiskers represent interquartile range, and diamonds represent outlier values.

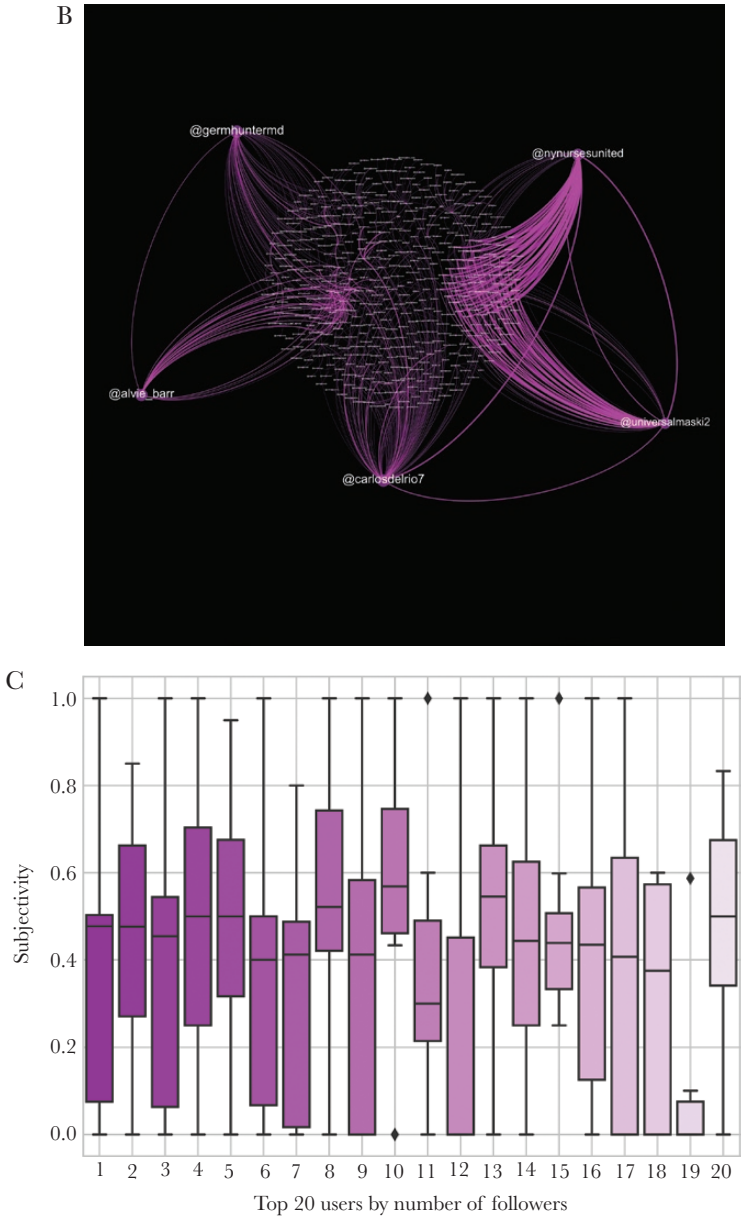

Figure 1. Continued.

with other users, representing a total of 464 nodes (users) and 1406 edges (connections). The darker and thicker pink lines converge in 5 distinct areas (pulled into the periphery) and demonstrate increased clustering and network activity among those users.

Subjectivity Analysis

The top 20 users (with a minimum of 5 IDWeek 2020 tweets) based on followers had a median (interquartile range [IQR]) subjectivity score of 0.425 (0.125–0.725) (Figure 1C). The highest median subjectivity score was 0.569 and lowest was 0. For privacy concerns, we have removed the account names.

DISCUSSION

In this study, we explore the social networks and topics discussed at the first virtual-only IDWeek on Twitter. Similar to previous studies, there was a substantial amount of tweets, retweets, and user activity related to this medical conference. When compared with IDWeek 2016 (3249 tweets from 488 users) [2], there was nearly a 4-fold increase in tweet activity and 7-fold increase in users. This large uptick may be a result of expected growth on the Twitter platform or the virtual format of the conference, which lent itself to increased internet access and impetus for virtual connection. With over half of all tweets representing retweets, there was clear evidence of amplification of original messages, facilitating sharing of ideas and thoughts.

This permits engagement in conversation and is one of the most attractive reasons for using this medium.

While social networks cannot replicate in-person networks [10], significant activity is noted around certain users, exemplified by both network analyses. Weighted lines demonstrate that even 1 individual or user account has the potential to initiate dialogue while many others participate. Interestingly, an account unrelated specifically to Infectious Diseases emerged at the center of the largest cluster (@nynursesunited) in our analysis, demonstrating the reach and impact of the social network.

Understanding the accounts with the highest impact as evidenced through retweets or number of followers may help new or existing users expand their own network, as it demonstrates how others already on the Twitter network are interacting. Through our subjectivity analysis of the top 20 users based on followers, we see a large range of subjectivity. We believe this provides unique insight into whether particularly impactful accounts are more likely to tweet factual or objective information vs a subjective opinion, and based on these data, a user can make a better informed decision to follow a particular individual or account. We do, however, recognize the potential for a few accounts to dominate conversations and stifle alternative opinions and interpretation of data and therefore caution users to be mindful of this possibility.

Despite the online format, important themes emerged from our topic modeling. Unsurprisingly, COVID-19 dominated much of the discourse on Twitter, with topics ranging from discussion surrounding airborne vs droplet transmission, emphasis on public health measures including nonpharmaceutical interventions, potential vaccine timelines and candidates, public health responses, and a path beyond the pandemic. However, other important topics were discussed, including the need to address health policy and disparities within Infectious Diseases, antibiotic stewardship, and clinical pearls. We believe the IDWeek organizers can use these data as a measure of success by understanding which topics or sessions were most commonly discussed. Similarly, if there were certain sessions they wished emphasized or highlighted through the program but these were not significantly discussed on Twitter, they could use this as an opportunity to promote these future sessions on Twitter to garner more interest.

There are limitations with our analysis. Using a single keyword to identify tweets relating to an entire conference may have introduced selection bias, and other important keywords such as "chasingthesun," which emerged as a weighted term in our topic modeling, may have been omitted. Despite this, we believe limiting our analysis to "IDWeek2020" strengthened our topic modeling as it did not introduce potential confounding tweets related to other events. Similarly, we extracted tweets only in the English language, and this study does not help inform topics or networks where the discussion was in another language. Finally, we recognize the risk of labeling bias by subjectively ascribing topics to weighted terms. To minimize this risk, 1 author performed the analysis while the other primarily labeled the topics with ultimate mutual agreement.

In summary, the migration of IDWeek 2020 to a virtual medium still allowed for educational and topical conversation on the Twitter platform. We urge users to be thorough when evaluating information on Twitter, given the sometimes superficial nature in the assessment of evidence. However, Twitter can be useful when trying to keep abreast of the rapidly changing landscape of medical literature, and we believe the information gained from our analyses can help a general Twitter user identify accounts of interest and uncover valuable information for IDWeek organizers as they plan future conferences. Despite the loss of the personal and humanistic feel to previous IDWeeks, social networks allowed individuals to engage in a network far beyond their in-person ones. Given the increasing Twitter activity at medical conferences, we expect Twitter to continue to play a significant role in engaging the Infectious Diseases community in the future.

Acknowledgments

Potential conflicts of interest. R.J.M. has received research funding from Verily (Google Life Sciences) and research grants from the Centers for Disease Control and Prevention. Both authors have submitted the ICMJE Form for Disclosure of Potential Conflicts of Interest. Conflicts that the editors consider relevant to the content of the manuscript have been disclosed.

Author contributions. Both authors contributed to study concept and design. R.J.M. contributed to data acquisition and extraction. Both authors contributed to data analysis. Both authors contributed to interpretation of data. Both authors contributed to manuscript preparation. Both authors read and approved the final manuscript.

References

1. McKendrick DR, Cumming GP, Lee AJ. Increased use of Twitter at a medical conference: a report and a review of the educational opportunities. J Med Internet Res **2012**; 14:e176.
2. Mitchell BG, Russo PL, Otter JA, et al. What makes a tweet fly? Analysis of twitter messaging at four infection control conferences. Infect Control Hosp Epidemiol **2017**; 38:1271–6.
3. Cevik M, Ong DSY, Mackenzie G. How scientists and physicians use Twitter during a medical congress. Clin Microbiol Infect **2019**; 25:1561.e7–12.
4. Medford RJ, Saleh SN, Sumarsono A, Perl TM, Lehmann CU. An "infodemic": leveraging high-volume twitter data to understand early public sentiment for the coronavirus disease 2019 outbreak. Open Forum Infect Dis **2020**; 7:XXX–XX.
5. Saleh SN, Lehmann CU, McDonald SA, Basit MA, Medford RJ. Understanding public perception of coronavirus disease 2019 (COVID-19) social distancing on Twitter. Infect Control Hosp Epidemiol. **In press**.
6. Twitter Inc. IDJClub. Available at: https://twitter.com/IDJClub. Accessed 29 October 2020.
7. Brunson JC, Laubenbacher RC. Applications of network analysis to routinely collected health care data: a systematic review. J Am Med Inform Assoc **2018**; 25:210–21.
8. Symplur. #IDWeek2020 conference hashthag. Available at: https://www.symplur.com/healthcare-hashtags/idweek2020/. Accessed 29 October 2020.
9. Loria S. TextBlob: simplified text processing. Available at: https://textblob.readthedocs.io/en/dev. Accessed 5 December 2020.
10. NEJM. HIV and ID Observations, blog. Available at: https://t.co/gFFP2iBMKz?amp=1. Accessed 29 October 2020.