E-Cigarette Surveillance With Social Media Data: Social Bots, Emerging Topics, and Trends

Jon-Patrick Allem¹, MA, PhD; Emilio Ferrara², PhD; Sree Priyanka Uppu³, MS; Tess Boley Cruz¹, PhD; Jennifer B Unger¹, PhD

¹Keck School of Medicine, University of Southern California, Los Angeles, CA, United States
²Information Sciences Institute, University of Southern California, Los Angeles, CA, United States
³Department of Computer Science, University of Southern California, Los Angeles, CA, United States

Corresponding Author:
Jon-Patrick Allem, MA, PhD
Keck School of Medicine
University of Southern California
2001 N. Soto Street
Los Angeles, CA,
United States
Phone: 1 858 603 0812
Email: allem@usc.edu

Abstract

Background: As e-cigarette use rapidly increases in popularity, data from online social systems (Twitter, Instagram, Google Web Search) can be used to capture and describe the social and environmental context in which individuals use, perceive, and are marketed this tobacco product. Social media data may serve as a massive focus group where people organically discuss e-cigarettes unprimed by a researcher, without instrument bias, captured in near real time and at low costs.

Objective: This study documents e-cigarette–related discussions on Twitter, describing themes of conversations and locations where Twitter users often discuss e-cigarettes, to identify priority areas for e-cigarette education campaigns. Additionally, this study demonstrates the importance of distinguishing between social bots and human users when attempting to understand public health–related behaviors and attitudes.

Methods: E-cigarette–related posts on Twitter (N=6,185,153) were collected from December 24, 2016, to April 21, 2017. Techniques drawn from network science were used to determine discussions of e-cigarettes by describing which hashtags co-occur (concept clusters) in a Twitter network. Posts and metadata were used to describe where geographically e-cigarette–related discussions in the United States occurred. Machine learning models were used to distinguish between Twitter posts reflecting attitudes and behaviors of genuine human users from those of social bots. Odds ratios were computed from 2x2 contingency tables to detect if hashtags varied by source (social bot vs human user) using the Fisher exact test to determine statistical significance.

Results: Clusters found in the corpus of hashtags from human users included behaviors (eg, #vaping), vaping identity (eg, #vapelife), and vaping community (eg, #vapenation). Additional clusters included products (eg, #eliquids), dual tobacco use (eg, #hookah), and polysubstance use (eg, #marijuana). Clusters found in the corpus of hashtags from social bots included health (eg, #health), smoking cessation (eg, #quitsmoking), and new products (eg, #ismog). Social bots were significantly more likely to post hashtags that referenced smoking cessation and new products compared to human users. The volume of tweets was highest in the Mid-Atlantic (eg, Pennsylvania, New Jersey, Maryland, and New York), followed by the West Coast and Southwest (eg, California, Arizona and Nevada).

Conclusions: Social media data may be used to complement and extend the surveillance of health behaviors including tobacco product use. Public health researchers could harness these data and methods to identify new products or devices. Furthermore, findings from this study demonstrate the importance of distinguishing between Twitter posts from social bots and humans when attempting to understand attitudes and behaviors. Social bots may be used to perpetuate the idea that e-cigarettes are helpful in cessation and to promote new products as they enter the marketplace.

(JMIR Public Health Surveill 2017;3(4):e98) doi: 10.2196/publichealth.8641
**KEYWORDS**
electronic cigarettes; vaping; Twitter; social media; social bots; electronic nicotine delivery system; infoveillance

**Introduction**

Electronic cigarettes (e-cigarettes) have climbed in popularity in the United States and elsewhere [1-6]. As e-cigarette use (vaping) rapidly becomes more prevalent, data from online social systems (eg, Google Web Search, Instagram, Twitter, YouTube) can be used to capture and describe the social and environmental context in which individuals use, perceive, and are marketed this tobacco product [7]. These data may serve as a massive focus group allowing for people to organically discuss e-cigarettes unprimed by a researcher, without instrument bias, captured in near real time and at low costs [8].

Internet searches (Google Web Search) for e-cigarette–related terms increased by 450% from 2010 to 2014 in the United States with search volume for e-cigarettes greater in coastal areas (California and New York) in 2010 before becoming more uniformly searched across the contiguous United States in 2014 [9]. Searches for terms indicative of purchasing e-cigarettes have outpaced searches indicative of interest in health concerns or smoking cessation [9]. In a study analyzing e-cigarette–related posts on Instagram, Chu and colleagues [10] reported that images often showed “cloud chasing” (ie, large clouds of aerosol being blown) and “hand checks” (ie, e-cigarette device paired with specific e-juice bottle all held in one hand), suggesting these are appealing characteristics of this emerging tobacco product.

Twitter has been used in tobacco control research with studies showing how tobacco education campaigns can be informed by monitoring tweets [11,12] and which e-cigarette–related messages are likely to spread on Twitter [13], among other studies [14-21]. Ayers and colleagues [22] recently analyzed a sample of e-cigarette–related tweets and reported that social image was the most identified reason for e-cigarette use in 2015. Other identified reasons for e-cigarette use included quitting combustible cigarettes and use indoors [22].

In this study, we demonstrate the feasibility of a Twitter-based infoveillance [7] methodology to document and describe e-cigarette–related conversations on Twitter. We used social network analyses to identify discussions of e-cigarettes by describing which hashtags co-occur in a massive Twitter network. Twitter users use hashtags (ie, terms prepended by the hash mark #) to indicate the context, emotions, or subject matter related to a post. Hashtags serve as a marker for the content of posts that allows users to search for and see posts of other Twitter users even if they do not follow them. Multiple hashtags can be adopted in a single post. When 2 hashtags co-occur in the same post, one can infer that they are related. Building the network of co-occurrence of hashtags (ie, hashtag network) will illustrate concept clusters giving us insights to e-cigarette–related discussions by individuals in their own words. This clustering allows us to see underlying dimensions of meaning that might not otherwise be possible in complex data.

We also used posts and metadata from Twitter to describe where geographically e-cigarette–related discussions in the United States occur to identify priority areas for e-cigarette education campaigns. Additionally, this study builds on earlier work [23,24] and demonstrates the importance of removing social bots (ie, computer algorithms designed to automatically produce content and engage with legitimate human accounts on Twitter) from Twitter data when attempting to understand public health–related behaviors and attitudes. Taken together, findings from this study should inform tobacco control and demonstrate the utility in using Twitter data in enhancing surveillance of health behaviors in general and e-cigarette use.

**Methods**

Data were obtained by means of Python scripts that continuously polled Twitter’s streaming application programming interface. This service provides a sample stream of data based on key terms and hashtag searches. Tweets were collected between December 24, 2016, and April 21, 2017. The key terms used to collect the tweets included e-cigarette, vaping, etc (see Multimedia Appendix 1 for complete list). The key terms could have appeared in the post or in an accompanying hashtag (ie, vaping or #vaping). The university’s institutional review board approved all procedures.

The terms used to collect tweets during the study period resulted in an initial corpus of 6,185,153 tweets. However, Twitter has quickly become subject to third-party manipulation where computer algorithms designed to automatically produce content and engage with legitimate human accounts on Twitter (social bots) are created to influence discussions and promote specific ideas or products [25]. Social bots are meant to appear as genuine human users operating Twitter accounts; their profiles are often complete with metadata (name, location, pithy quote) and a photo/image. Social bots on average generate more tweets than the average human user. Therefore, social bots are producing more content on a topic. Social bots make indiscriminate references to an array of content while at the same time perpetuating select conversations, giving the appearance that a specific topic is more prominent than it is offline. Their adoption has been documented in a variety of domains, including political astroturfing [26], stock market manipulation [27], spread of misinformation [28], promotional content [29], and in sentiment classification [24].

In order to distinguish between human users and social bots, certain criteria such as information diffusion patterns (based on retweets or mentions), friend features (for example, ratio of followers to followees), content (frequency of nouns/verbs/adverbs in a tweet), and sentiment features (emotion scores) are used. The BotOrNot algorithm combines these features to obtain a single score between 0 and 1 that indicates if a Twitter account is a social bot or not [28,30]. Evaluations of the BotOrNot program have shown that an account is most likely to be a bot if the account score is ≥0.6 [24,27,29]. The method used for bot detection has a detection accuracy above...
Table 1. Most common hashtags in each respective cluster from the bot-free corpus.

| Cluster | Hashtags                                                                 |
|---------|--------------------------------------------------------------------------|
| 1 (pink) | vaping, ecigs, vapelife, vapeporn, weed, buzz, vaporizer, vapenation, eliquid, cannabis, vape, vapes, bigtobacco, ejuice, smokeshop |
| 2 (green)| eliquids, vaper, vapelife, smoke, instavape, vapecommunity, ecig, vapors, atomizer, vapelclub, vapetagram, vapesociety   |
| 3 (orange)| smokers, nowsmoking, cigaretters, tobacco, week, marijuana, cigars, whisky, scotch, smoker, cigarettes, hookah, addiction, blu |

*Colors correspond to the figure found in Multimedia Appendix 2.*
Table 2. Most common hashtags in each respective cluster from the bot corpus.

| Cluster  | Hashtags                                                                 |
|----------|-------------------------------------------------------------------------|
| 1 (orange) | Cigars, cigar, blu, tobacco, cigarette, smoke, smoking, photography, galanecigars, lifelove |
| 2 (gray)  | vapes, vape, vaping, vapor, ecig, ecigs, vapefam, vapelifge, vapor, smok, vaporstorm, eliquids, vapepen, vapefamily, vapeshop, vapecomunity, vaporn, vapers, vaporizer, ecigaretters |
| 3 (blue)  | esmoke, esmoking, online, beast, mod, cheap, cigpet, starterskit, esmoker, mobile, ismog, modbox |
| 4 (green) | marijuana, smoking, health, weed, tobacco, cannabis, cbd, thc, cool, bongs, tobacco, quality, cheap, vapes, ejuice, quismaoking |

*Colors correspond to the figure found in Multimedia Appendix 3.*
## Table 3. Associations between hashtags and data source (bots vs humans coded) with an odds ratio $> 1$ indicating greater likelihood from a bot.

| Hashtags   | Odds ratio | P value |
|------------|------------|---------|
| addiction  | 0.62       | .006    |
| atomizer   | 0.53       | <.001   |
| beast      | 2.72       | <.001   |
| bigtobacco | 0.08       | <.001   |
| blu        | 1.33       | <.001   |
| bongs      | 1.79       | <.001   |
| tobacco    | 1.05       | .002    |
| buzz       | 0.66       | <.001   |
| cannabis   | 1.81       | <.001   |
| cheap      | 1.81       | <.001   |
| cigar      | 0.66       | <.001   |
| cigarette  | 1.65       | <.001   |
| cigarettes | 0.55       | <.001   |
| cigars     | 0.54       | <.001   |
| cigpet     | 2.73       | <.001   |
| cool       | 1.58       | .03     |
| ecig       | 0.54       | <.001   |
| ecigs      | 1.52       | <.001   |
| ejuice     | 0.97       | .23     |
| eliquid    | 0.73       | <.001   |
| eliquids   | 1.06       | .40     |
| esmoke     | 2.88       | <.001   |
| esmoking   | 2.87       | <.001   |
| esmoker    | 2.89       | <.001   |
| health     | 1.00       | .92     |
| hookah     | 0.29       | <.001   |
| instavape  | 0.80       | .03     |
| ismog      | 2.89       | <.001   |
| marijuana  | 1.25       | <.001   |
| mobile     | 1.68       | <.001   |
| mod        | 2.47       | <.001   |
| modbox     | 2.41       | <.001   |
| nowsmoking | 0.70       | .003    |
| online     | 2.78       | <.001   |
| photography| 0.96       | .87     |
| quality    | 1.80       | <.001   |
| quitsmoking| 2.27       | <.001   |
| scotch     | 0.02       | <.001   |
| smoke      | 1.00       | .85     |
| smoker     | 2.61       | <.001   |
| smokers    | 1.66       | <.001   |
| smokeshop  | 1.33       | .07     |
The cluster analysis from the corpus of hashtags from social bots contained 4 clusters with 137 hashtags or nodes and 1600 edges (Multimedia Appendix 3). Cluster 1 (orange) contained hashtags indicative of behaviors and dual tobacco use (Table 2). Cluster 2 (gray) contained hashtags indicative of behaviors and vaping identity and vaping community. Cluster 3 (blue) contained hashtags indicative of products (eg, #starterskit, #modbox), including brand new products (eg, #ismog, a new smart device with touch technology on a vaping box, #cigpet, a new high wattage tank or “super tank”). Cluster 4 (green) contained hashtags indicative of smoking cessation (eg, #quitsmoking), interest in health (eg, #health), and polysubstance use.

Social bots were more likely to post hashtags that referenced smoking cessation and new e-cigarette devices compared to human users (Table 3). For example, social bots were significantly more likely to post #quitsmoking, #ismog, and #cigpet compared to human users.

The heat map representing 26,565 tweets collected from December 24, 2016, to April 21, 2017, shows that the volume of tweets is highest in the Mid-Atlantic (eg, Pennsylvania, New Jersey, Maryland, and New York) and high on the West Coast and Southwest (eg, California, Arizona and Nevada) (Figure 1).

| Hashtags     | Odds ratio | P value |
|--------------|------------|---------|
| smoking      | 1.04       | .31     |
| starterskit  | 2.88       | <.001   |
| the          | 2.43       | <.001   |
| tobacco      | 1.05       | .002    |
| vape         | 0.82       | <.001   |
| vapecommunity| 0.19       | <.001   |
| vapefam      | 0.34       | <.001   |
| vapefamily   | 1.42       | <.001   |
| vapelifes     | 0.63       | <.001   |
| vapenation   | 0.78       | <.001   |
| vapestagram  | 0.69       | <.001   |
| vapers       | 1.67       | <.001   |
| vapeshop     | 1.13       | .03     |
| vapersociety | 0.34       | <.001   |
| vapestagram  | 0.69       | <.001   |
| vapor        | 1.15       | <.001   |
| vaporizer    | 0.29       | <.001   |
| vapors       | 0.52       | .0002   |
| vaporstorm   | 2.94       | <.001   |
| weed         | 1.03       | .5694   |
| whiskey      | 0.02       | <.001   |
Discussion

Principal Findings

Data from online social systems may be used to complement and extend the surveillance of health behaviors including tobacco product use. The hashtags we studied here provide several direct insights into e-cigarette–related attitudes and behaviors with the identification of 3 clusters that represent the most cohesive posts. The cluster analysis from the corpus of hashtags from human users demonstrated the existence of a vaping identity and vaping community. Use of these hashtags may serve further internalization of, and social bonding around, vaping-related identities. These hashtags also suggest discussions of vaping may occur in an echo chamber on Twitter in which ideas and beliefs are amplified by those in the network [34], normalizing vaping.

In the cluster analysis from the corpus of hashtags from human users, we found many references to vaping-related products. These hashtags represent a way for commercial users to make their posts searchable and integrate themselves into online communities of vapers. Noncommercial users may also use these hashtags to communicate to their followers which products they recently purchased or which products they like to use together (eg, their favorite modifiable device paired with their favorite e-liquid) [10].

The third hashtag cluster found in the corpus from human users, we found many references to vaping-related products. These hashtags represent a way for commercial users to make their posts searchable and integrate themselves into online communities of vapers. Noncommercial users may also use these hashtags to communicate to their followers which products they recently purchased or which products they like to use together (eg, their favorite modifiable device paired with their favorite e-liquid) [10].

The third hashtag cluster found in the corpus from human users indicated dual tobacco use and polysubstance use. These co-occurring hashtags may reflect a syndrome of risky behavior among select vapers. While research is accumulating about dual e-cigarette and cigarette use [35,36], there is a dearth of research on the associations between vaping and hookah, marijuana, alcohol, and other substance use. The findings from this study should spur efforts to investigate these associations further. When the population-level impact of e-cigarettes is being debated, the co-occurrence of vaping with alcohol and other substances should also be considered.

In the corpus of hashtags from social bots, several results stood out in contrast to the results from the human user corpus. For one, a cluster of hashtags was detected that referenced smoking cessation. This suggests social bots may be used to perpetuate discussions on e-cigarettes as a cessation device. While earlier research has suggested Twitter posts about vaping referenced the use of e-cigarettes in cessation [22], it is important to distinguish between individual users and social bots when analyzing posts on Twitter [23,24,37]. Social bots may perpetuate misinformation about the efficacy of e-cigarettes in cessation, thus requiring education campaigns to serve as a vehicle to correct this misinformation.

Hashtags from social bots also represented newly introduced products to the marketplace (eg, #ismog and #cigpet) which were significantly less prevalent in the human user corpus of hashtags. This finding highlights a clear benefit of using social media data in public health surveillance. In addition to searching for known keywords and observing trends in the number of social media posts that contain those keywords, the concept cluster analysis can identify new keywords or hashtags posted on Twitter. This process can serve as an early warning system informing public health researchers about new products or new ways in which products are appealing to the public. By using social media data and keyword co-occurrence analyses we can identify new products (like ismog or cigpet), brands, marketing themes, activities, and events associated with tobacco product use as they emerge in near real time. The findings from this study complement recent research that relied on search navigation data to detect growing interest in heat-not-burn tobacco products [38]. Taken together, public health researchers could use data from online social systems to fill knowledge gaps in real time.
gaps quickly and respond more readily to the populations they serve.

Most posts were from the Mid-Atlantic and Southwest, which is compatible with earlier research relying on search navigation data [9]. The findings mark priority areas for e-cigarette education campaigns. Social media may be one way to engage with nonusers of tobacco products to inform them of the addictive properties of nicotine as well as the harms of e-cigarette use [39]. Using social media as a complementary surveillance system could allow public health researchers to identify geographic disparities in emerging tobacco product use earlier than traditional methods. While Twitter data should not be used to supersede traditional health behavior surveillance systems, social media could be used to fill information gaps quickly and can provide an important starting point to address an issue of great import to public health or policy.

Limitations

Data collection relied on Twitter’s streaming application programming interface, which prevents collecting tweets from private Twitter accounts. As a result, findings may not represent the attitudes and behaviors from individuals with private accounts. This study used hashtags to identify themes in posts on Twitter but did not specifically read and interpret each post that the hashtags accompanied. Additional valuable information could have been learned from the content of the posts that was not described herein. Approximately 1% of all users in the analytical sample provided data that allowed us to describe the geographic areas in which e-cigarette–related discussions took place in the United States. While this is a small percentage, it is compatible with earlier work [25,40] and represents 36,549 users in the United States. Additionally, we did not have the necessary demographic information (eg, age) of Twitter users to consider population density and age distributions of geographic areas.

Conclusion

The findings from this study can inform the design of public health surveillance in the future. This study demonstrated the utility in using social media data in understanding attitudes and behaviors and the importance of distinguishing between Twitter posts from social bots and humans during this process if the intent is to assess views held by real users. Findings should spur efforts to better understand the consequences of e-cigarette–related discussions on Twitter.

Acknowledgments

Research reported in this publication was supported by grant #P50CA180905 from the National Cancer Institute, National Institutes of Health, and the Center for Tobacco Products, US Food and Drug Administration, which had no role in study design, collection, analysis, and interpretation of data; writing the report; and the decision to submit the report for publication. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health or US Food and Drug Administration.

Authors' Contributions

JPA and EF conceived of the study. EF collected the data. EF, SPU, and JPA analyzed the data. JPA drafted the initial manuscript. JPA, EF, TBC, and JBU revised the manuscript for important intellectual content and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Key search terms.

[PDF File (Adobe PDF File), 12KB-Multimedia Appendix 1]

Multimedia Appendix 2

Concept cluster built from hashtags collected from a bot-free corpus.

[PNG File, 1MB-Multimedia Appendix 2]

Multimedia Appendix 3

Concept cluster built from hashtags collected from a bot corpus.

[JPG File, 592KB-Multimedia Appendix 3]

References

1. Arrazola RA, Singh T, Corey CG, Husten CG, Neff LJ, Apelberg BJ, et al. Tobacco use among middle and high school students—United States, 2011-2014. MMWR Morb Mortal Wkly Rep 2015 Apr 17;64(14):381-385 [FREE Full text] [Medline: 25879896]
2. Schoenborn CA, Gindi RM. Electronic cigarette use among adults: United States, 2014. NCHS Data Brief 2015 Oct(217):1-8. [FREE Full text] [Medline: 2655932]

3. Leventhal AM, Strong DR, Kirkpatrick MG, Unger JB, Sussman S, Riggs NR, et al. Association of electronic cigarette use with initiation of combustible tobacco product smoking in early adolescence. JAMA 2015 Aug 18;314(7):700-707. [FREE Full text] [doi: 10.1001/jama.2015.8950] [Medline: 26284721]

4. Wills TA, Knight R, Sargent JD, Gibbons FX, Pagano I, Williams RJ. Longitudinal study of e-cigarette use and onset of cigarette smoking among high school students in Hawaii. Tob Control 2017 Jan;26(1):34-39. [doi: 10.1136/tobaccocontrol-2015-052705] [Medline: 26811353]

5. Unger JB, Soto DW, Leventhal A. E-cigarette use and subsequent cigarette and marijuana use among Hispanic young adults. Drug Alcohol Depend 2016 Jun 01;163:261-264. [doi: 10.1016/j.drugalcdep.2016.04.027] [Medline: 27141841]

6. Soneji S, Barrington-Trimis JL, Wills TA, Leventhal AM, Unger JB, Gibson LA, et al. Association between initial use of e-cigarettes and subsequent cigarette smoking among adolescents and young adults: a systematic review and meta-analysis. JAMA Pediatr 2017 Jun 26. [doi: 10.1001/jamapediatrics.2017.1488] [Medline: 28654986]

7. Eysenbach G. Infodemiology and infoveillance tracking online health information and cyberbehavior for public health. Am J Prev Med 2011 May;40(5 Suppl 2):S154-S158. [doi: 10.1016/j.amepre.2011.02.006] [Medline: 21521589]

8. Ayers JW, Althouse BM, Dredze M. Could behavioral medicine lead the web data revolution? JAMA 2014 Apr 9;311(14):1399-1400. [doi: 10.1001/jama.2014.1505] [Medline: 24577162]

9. Ayers JW, Althouse BM, Allem J, Lesa EC, Dredze M, Williams RS. Revisiting the rise of electronic nicotine delivery systems using search query surveillance. Am J Prev Med 2016 Jun;50(6):e173-e181. [FREE Full text] [doi: 10.1016/j.amepre.2015.12.008] [Medline: 26876772]

10. Chu K, Allem JP, Cruz T, Unger J. Vaping on Instagram: cloud chasing, hand checks, and product placement. Tob Control 2016;26(5).

11. Allem JP, Escobedo P, Chu K, Soto DW, Cruz TB, Unger JB. Campaigns and counter campaigns: reactions on Twitter to e-cigarette education. Tob Control 2017 Mar;26(2):226-229. [doi: 10.1136/tobaccocontrol-2015-052757] [Medline: 26956467]

12. Ayers JW, Westmaas JL, Lesa EC, Benton A, Chen Y, Dredze M, et al. Leveraging big data to improve health awareness campaigns: a novel evaluation of the Great American Smokeout. JMIR Public Health Surveil 2016;2(1):e16. [FREE Full text] [doi: 10.2196/publichealth.5304] [Medline: 27227151]

13. Chu K, Unger JB, Allem JP, Pattarroyo M, Soto D, Cruz TB, et al. Diffusion of messages from an electronic cigarette brand to potential users through Twitter. PLoS One 2015;10(12):e0145387. [FREE Full text] [doi: 10.1371/journal.pone.0145387] [Medline: 26684746]

14. Huang J, Kornfield R, Szczypka G, Emery SL. A cross-sectional examination of marketing of electronic cigarettes on Twitter. Tob Control 2014 Jul;23 Suppl 3:26-30. [FREE Full text] [doi: 10.1136/tobaccocontrol-2014-051551] [Medline: 24935894]

15. van der Tempel J, Noormohamed A, Schwartz R, Norman C, Malas M, Zawertailo L. Vape, quit, tweet? Electronic cigarettes and smoking cessation on Twitter. Int J Public Health 2016 Mar;61(2):249-256. [doi: 10.1007/s00038-016-0791-2] [Medline: 26841895]

16. Sanders-Jackson A, Brown CG, Prochaska JJ. Applying linguistic methods to understanding smoking-related conversations on Twitter. Tob Control 2015 Mar;24(2):136-138. [FREE Full text] [doi: 10.1136/tobaccocontrol-2013-051243] [Medline: 24227540]

17. Jo CL, Kornfield R, Kim Y, Emery S, Ribisl KM. Price-related promotions for tobacco products on Twitter. Tob Control 2016 Jul;25(4):476-479. [doi: 10.1136/tobaccocontrol-2015-052260] [Medline: 26130449]

18. Cole-Lewis H, Pugatch J, Sanders A, Varghese A, Posada S, Yun C, et al. Social listening: a content analysis of e-cigarette discussions on Twitter. J Med Internet Res 2015;17(10):e243. [FREE Full text] [doi: 10.2196/jmir.4969] [Medline: 26508089]

19. Myslín M, Zhu S, Chapman W, Conway M. Using twitter to examine smoking behavior and perceptions of emerging tobacco products. J Med Internet Res 2013;15(8):e174. [FREE Full text] [doi: 10.2196/jmir.2534] [Medline: 23989137]

20. Aphiyanaphongs Y, Lulejian A, Brown D, Bonneau R, Krebs P. Text classification for automatic detection of e-cigarette use and use for smoking cessation from Twitter: a feasibility pilot. Pac Symp Biocomput 2016;21:480-491.

21. Dai H, Hao J. Mining social media data for opinion polarities about electronic cigarettes. Tob Control 2017 Mar;26(2):175-180. [doi: 10.1136/tobaccocontrol-2015-052818] [Medline: 26980151]

22. Ayers JW, Lesa EC, Allem JP, Benton A, Dredze M, Althouse BM, et al. Why do people use electronic nicotine delivery systems (electronic cigarettes)? A content analysis of Twitter, 2012-2015. PLoS One 2017;12(3):e0170702. [FREE Full text] [doi: 10.1371/journal.pone.0170702] [Medline: 28248987]

23. Allem JP, Ferrara E. The importance of debiasing social media data to better understand e-cigarette-related attitudes and behaviors. J Med Internet Res 2016 Aug 09;18(8):e219. [FREE Full text] [doi: 10.2196/jmir.6185] [Medline: 27507563]

24. Allem JP, Ramanujam J, Lerman K, Chu K, Boley Cruz T, Unger J. Identifying sentiment of hookah-related posts on Twitter. JMIR Public Health Surveil 2017 Oct 18;3(4):e74. [FREE Full text] [doi: 10.2196/publichealth.8133] [Medline: 29046267]
25. Ferrara E, Varol O, Davis C, Menczer F, Flammini A. The rise of social bots. Communications of the ACM 2016;59(7):96-104.
26. Bessi A, Ferrara E. Social bots distort the 2016 US Presidential election online discussion. First Monday 2016;21(11).
27. Ferrara E. Manipulation and abuse on social media. ACM SIGWEB Newsletter 2015:4.
28. Varol O, Davis C, Menszer F, Flammini A. Online human-bot interactions: detection, estimation, and characterization. 2017 Presented at: Proceedings of the 11th International AAAI Conference on Web and Social Media; 2017; Montreal p. 280-289.
29. Varol O, Ferrara E, Menczer F, Flammini A. Early Detection of Promoted Campaigns on Social Media. EPJ Data Science 2017;6(1).
30. Davis CA, Varol O, Ferrara E, Flammini A, Menczer F. BotorNot: a system to evaluate social bots. 2016 Presented at: 25th International Conference Companion on World Wide Web; 2016; Geneva p. 273-274.
31. Bastian M, Heymann S, Jacomy M. Gephi: an open source software for exploring and manipulating networks. ICWSM 2009:8.361.
32. Fruchterman T, Reingold E. Graph drawing by force-directed placement. Software Prac Exp 1991;21(11):1129.
33. Blondel V, Guillaume J, Lambiotte R, Lefebvre E. The Louvain method for community detection in large networks. J Stat Mech-Theory E 2011.
34. Colleoni E, Rozza A, Arvidsson A. Echo chamber or public sphere? Predicting political orientation and measuring political homophily in Twitter using big data. J Comm 2014;64:317-332.
35. Gilreath TD, Leventhal A, Barrington-Trinis JL, Unger JB, Cruz TB, Berhane K, et al. Patterns of alternative tobacco product use: emergence of hookah and e-cigarettes as preferred products amongst youth. J Adolesc Health 2016 Feb;58(2):181-185 [FREE Full text] [doi: 10.1016/j.jadohealth.2015.10.001] [Medline: 26598059]
36. Leventhal AM, Stone MD, Andrabri N, Barrington-Trinis J, Strong DR, Sussman S, et al. Association of e-cigarette vaping and progression to heavier patterns of cigarette smoking. JAMA 2016 Nov 08;316(18):1918-1920. [doi: 10.1001/jama.2016.14649] [Medline: 27825000]
37. Clark EM, Jones CA, Williams JR, Kuri AN, Norotsky MC, Danforth CM, et al. Vaporous marketing: uncovering pervasive electronic cigarette advertisements on Twitter. PLoS One 2016;11(7):e0157304 [FREE Full text] [doi: 10.1371/journal.pone.0157304] [Medline: 27410031]
38. Caputi TL, Leas E, Dredze M, Cohen JE, Ayers JW. They're heating up: Internet search query trends reveal significant public interest in heat-not-burn tobacco products. PLoS One 2017;12(10):e0185735 [FREE Full text] [doi: 10.1371/journal.pone.0185735] [Medline: 29020019]
39. Reidel B, Radicioni G, Clapp P, Ford AA, Abdelwahab S, Rebuli ME, et al. E-cigarette use causes a unique innate immune response in the lung involving increased neutrophilic activation and altered mucin secretion. Am J Respir Crit Care Med 2017 Oct 20. [doi: 10.1164/rcrm.201708-1590OC] [Medline: 29053025]
40. Dai H, Hao J. Mining social media data for opinion polarities about electronic cigarettes. Tob Control 2017 Mar;26(2):175-180. [doi: 10.1136/tobaccocontrol-2015-052818] [Medline: 26980151]

Please cite as:
Allem JP, Ferrara E, Uppu SP, Cruz TB, Unger JB
E-Cigarette Surveillance With Social Media Data: Social Bots, Emerging Topics, and Trends
JMIR Public Health Surveill 2017;3(4):e98
URL: http://publichealth.jmir.org/2017/4/e98/
do: 10.2196/publichealth.8641
PMID: 29263018

©Jon-Patrick Allem, Emilio Ferrara, Sree Priyanka Uppu, Tess Boley Cruz, Jennifer B Unger. Originally published in JMIR Public Health and Surveillance (http://publichealth.jmir.org), 20.12.2017. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Public Health and Surveillance, is properly cited. The complete bibliographic information, a link to the original publication on http://publichealth.jmir.org, as well as this copyright and license information must be included.