

UC Office of the President

Research Grants Program Office (RGPO) Funded Publications

Title

Cannabis Surveillance With Twitter Data: Emerging Topics and Social Bots.

Permalink

<https://escholarship.org/uc/item/971541m7>

Journal

American Journal of Public Health, 110(3)

ISSN

0090-0036

Authors

Allem, Jon-Patrick

Escobedo, Patricia

Dharmapuri, Likhith

Publication Date

2020-03-01

DOI

10.2105/ajph.2019.305461

Peer reviewed

Cannabis surveillance with Twitter data: emerging topics and social bots

Jon-Patrick Allem, Ph.D., M.A.,¹ Patricia Escobedo, M.A.,¹ Likhith Dharpuri,
M.S.²

¹ Keck School of Medicine, University of Southern California, Los Angeles, CA, USA

² Department of Computer Science, University of Southern California, Los Angeles, CA, USA

Abstract

Introduction: Cultural trends in the United States, the cannabis consumer marketplace, and state cannabis policies are rapidly changing. Our goal was to use publicly accessible data from people who post to Twitter to rapidly capture and describe the public's recent experiences with cannabis.

Methods: Twitter posts containing cannabis-related terms were obtained from May 1, 2018 to December 31, 2018. Methods were used to distinguish between posts from social bots and non-bots. Text classifiers were used to identify topics in posts (n = 60,861). **Results:** Prevalent topics of posts included *Using Cannabis* with mentions of cannabis initiation, and *Health and Medical* with posts suggesting that cannabis could help with sleep, pain, anxiety, depression, trauma, and post-traumatic stress disorder. *Polysubstance* Use was a common topic with mentions of cocaine, heroin, ecstasy, LSD, meth, mushrooms, and Xanax along with cannabis. Social bots

32 commonly discussed the health benefits of cannabis. **Conclusions:** Findings
33 suggest that high potency cannabis products, unsubstantiated health claims
34 about cannabis products, and the co-use of cannabis with legal and illicit
35 substances warrant considerations by public health researchers in the future.

36

37 The legalization of cannabis through state-level medical cannabis and
38 adult (21 years or older) recreational use laws has lowered prices, and
39 increased the availability of cannabis products in the United States (U.S.).¹
40 Currently, a total of 34 states have legalized medical cannabis, and ten
41 states have legalized cannabis for adult recreational use.² The 2017 National
42 Survey on Drug Use and Health found that past month cannabis use was
43 highest among adults aged 18-25 (22.1%), compared to adolescents (6.5%),
44 and adults ages 26 or older (7.9%).³ Acute health outcomes associated with
45 cannabis use include impaired short-term memory, impaired attention,
46 impaired coordination, and sleep problems,³ while repeated cannabis use is
47 associated with potential for cannabis dependence, increased risk of other
48 drug and alcohol use disorders, and increased risk of schizophrenia, among
49 individuals with genetic vulnerability.³ There is also substantial evidence
50 between cannabis use and increased risk of motor vehicle crashes.⁴ Around
51 one in ten cannabis users will become addicted, however for individuals who
52 begin using cannabis as adolescents, one in six will become dependent.⁴
53 Several longitudinal studies have suggested that heavy cannabis use during

54 adolescence may lead to lower cognitive functioning and IQ during
55 adulthood.^{5,6}

56 Publicly accessible data from people who post to social media
57 platforms, like Twitter, can be used to rapidly capture and describe the
58 context of cannabis use.^{7,8} Twitter is used by 22% of U.S. adults (24% of
59 men, 21% of women, 21% of whites, 24% of African Americans 25% of
60 Hispanics) with 42% of users on the platform daily.⁹ Twitter is also used by
61 32% of adolescents (13 to 17 years) in the U.S.¹⁰ Previous analyses of
62 cannabis-related posts to Twitter, drawn from brief time periods and
63 relatively small amounts of data, have provided the initial information on
64 what the public organically discusses, including the desire to use cannabis,
65 mentions of health benefits, legalization efforts, and frequency of use.¹¹
66 Krauss and colleagues analyzed a sample of posts to Twitter from one month
67 in 2014, and found that tweets commonly mentioned cannabis and alcohol
68 co-use.¹² However, cultural trends in the U.S., the cannabis consumer
69 marketplace, and state cannabis policies are rapidly changing. The context
70 and experiences associated with cannabis use rapidly change as well,
71 making it important to provide recent information on cannabis. The goal of
72 this study was to identify and describe cannabis-related topics of
73 conversation on Twitter to inform the public health community.

74 **Methods**

75 Twitter (<https://twitter.com/>) posts containing the cannabis-related
76 terms: “blunt,” “bong,” “budder,” “cannabis,” “cbd,” “ganja,” “hash,”

77 “hemp,” “indica,” “kush,” “marijuana,” “marihuana,” “reefer,” “sativa,”
78 “thc,” and “weed,” were obtained from May 1, 2018 to December 31, 2018.
79 These terms were informed by prior research and topic matter experts.¹¹⁻¹⁴
80 There was a total of $n = 53,177,048$ posts containing these terms during this
81 time. Similar to prior research,^{15,16} we removed all retweets ($n = 34,095,967$),
82 and sampled out a subset of the remaining tweets ($n = 19,081,081$) to
83 conduct analyses. We accomplished this by grouping tweets by the week in
84 which they were posted (this allowed us to maintain temporal characteristics
85 of the tweets when sampling). We then sampled proportionally from each
86 cannabis-related term by week.¹⁷ From this, we sampled out 102,701 tweets
87 from 75,751 unique accounts.

88 Next, we filtered out non-English tweets and tweets that contained key
89 terms but did not refer to cannabis – as in the case of Emily Blunt (actress),
90 James Blunt (singer), and hash brown (food), among others. Then, we filtered
91 out social bots, or automated Twitter accounts, designed to produce content
92 and engage with legitimate human accounts on Twitter.¹⁸ Social bots may
93 bias the data limiting our ability to reliably describe the public’s recent
94 experience with cannabis.¹⁹ In order to distinguish between non-bots and
95 social bots, Botometer was used.²⁰ This program analyzes the characteristics
96 of a Twitter account and gives it a score based on how likely the account is
97 to be a social bot. This method of social bot detection is considered state of
98 the art, and has been employed in previous studies focused on social bots
99 and public health.²¹⁻²³ Through these procedures, we arrived at $n = 60,861$

100 tweets from 47,760 non-bots and $n = 8,874$ tweets from social bots. All
101 analyses relied on public, anonymized data, adhered to the terms and
102 conditions, terms of use, and privacy policies of Twitter, and were performed
103 under Institutional Review Board approval from the authors' university. To
104 protect privacy, no tweets were reported verbatim in this report.

105 To prepare tweets for data analysis we performed a number of
106 transformations, including 1) *Basic normalization* which encompasses lower
107 casing all tweets, removing extra spaces, punctuation, and special
108 characters such as brackets. 2) *Stop word removal*. Words such as 'a', 'the',
109 etc. are heavily represented in the English language, adding to the syntax,
110 but rarely adding to the meaning of a sentence. As such, we remove these
111 words. 3) *Normalizing Twitter account mentions*. On Twitter, @account_name
112 is used to tag accounts, and pages in a post. The name of each account
113 tagged has little importance to our study, but we wanted to maintain
114 information on the number of accounts tagged. Therefore, all
115 @account_name occurrences in the tweets were replaced by @person - a
116 common token for all accounts. 4) *Lemmatization*. Words such as 'walked',
117 and 'walk' can be conflated in our analysis, so we broke down words into
118 their basic form by removing inflections, and variants. 5) *Non-printable
119 character removal*. Unicode characters in tweets are often used for
120 emoticons, or as symbols from other languages. Since we are interested in
121 tweets in English, we can remove these symbols without much loss in the
122 meaning of the sentence. 6) *Removal of hashtags, and URLs*. Hashtags are

123 useful for filtering out tweets which relate to a group (e.g., #weed), but are
124 not necessary to analyze when dividing a group into further topics. As such,
125 we discarded hashtags. URLs embedded in the tweet were usually links to
126 images (which on Twitter are shown as embedded images), and external
127 links to other websites. Since we cannot obtain much information about a
128 website from its URL, we discarded it.

129 In order to find topics within our tweets, we generated one-grams and
130 bi-grams from each tweet. For example, the sentence “A quick brown fox
131 jumps” contains the one-grams “a,” “quick,” “brown,” “fox,” and “jumps.” It
132 contains the bi-grams “a quick,” “quick brown,” “brown fox,” and “fox
133 jumps.” By generating frequency counts of the most common one-grams and
134 bi-grams, we obtained an initial idea for the common topics discussed. From
135 this assessment, we arrived at consensus on ten commonly occurring topics
136 including, *Person Tagging* (e.g., @person), *Using Cannabis* (mentions of
137 smoking cannabis, passing a blunt), *Health and Medical* (mentions of the use
138 of cannabis to relieve health problems like anxiety and pain), *Legality*
139 (mentions of cannabis laws and legalizing cannabis), *Buy/Sell* (mentions of
140 the purchase and delivery of cannabis), *Processed Product Usage* (mentions
141 of consuming edibles, wax, dab), *CBD and Hemp Use* (mentions of using CBD
142 oil, hemp oil, CBD infused products), *Appeal or Abuse Liability* (mentions of
143 needing, wanting, or craving cannabis), *Polysubstance use* (mention of other
144 substances including alcohol, painkillers, psychedelics), and *Cannabis*
145 *Industry* (mentions of cannabis stocks, markets, and related industries).

146 Although not prominent topics, but consistent with our prior research,¹⁵ we
147 looked for words and phrases that suggested *Underage Use* (mentions of
148 cannabis use at school), and *Impairment* (mentions of cannabis use at the
149 workplace or driving under the influence).

150 Each tweet was classified to one or more topics based on the presence
151 of at least one topic-related pattern. A pattern could be a one-gram, a bi-
152 gram, or any group of words that must occur in the normalized tweets in a
153 given order. We accomplished this by using a rule-based classification script
154 written in Python where each tweet was examined for the presence of a
155 specified set of patterns representing a topic. Since there was topic overlap,
156 we report the percentage of overlap between each topic by utilizing a
157 confusion matrix as a visualization tool. Each cell in the matrix represents
158 the intersection of two topics. The value of the cell represents the
159 percentage of the total corpus which belongs to both topics. For example, a
160 hypothetical post such as “Hey @person share your edibles” would be
161 classified under *Person Tagging* and *Processed Product Usage*. The number
162 of posts containing both would be found at the intersection of the matrix for
163 these 2 topics.

164 **Results**

165 The total coverage of the 12 topics constituted 58.14% of all tweets in
166 the corpus from non-bots (**Figure 1**). The remaining 41.86% of tweets were
167 too varied to be classified into a single topic with meaningful coverage (e.g.,
168 coverage of each subsequent topic would be less than 1% of total tweets).

169 The most prevalent topic in this corpus was *Person Tagging* at 33.60%
170 followed by *Using Cannabis* at 11.89%. Among *Using Cannabis*, 2.61% of
171 posts were indicative of cannabis initiation including phrases such as “first
172 time.” *Health and Medical* was the next most prevalent topic at 5.61%.
173 Among *Health and Medical*, cannabis was suggested to help with sleep, pain,
174 anxiety, depression, trauma, and post-traumatic stress disorder (PTSD),
175 among others. *Legality* was the next most prevalent topic at 5.50%,
176 followed by *Buy/Sell* at 5.01%, *Processed Product Usage* at 3.71%, *CBD and*
177 *Hemp Use* at 2.73%, and *Appeal or Abuse Liability* at 2.65%. *Polysubstance*
178 *Use* was a common topic at 2.47%. Among *Polysubstance Use*, beer, wine,
179 vodka, tequila, cocaine, heroin, ecstasy, LSD, meth, mushrooms and Xanax
180 were mentioned along with cannabis. *Cannabis Industry* comprised 1.33% of
181 posts. *Impairment* and *Underage Use* were uncommon topics at 0.47% and
182 0.35%, respectively.

183 The total coverage of the same 12 topics constituted 55.60% of all
184 tweets in the corpus from social bots. Comparing the two corpuses, some
185 topics have similar prevalence while other topics stand out with large
186 differences. For example, the largest difference in prevalence in topics
187 between corpuses was found in *Person Tagging* (non-bots at 33.60% versus
188 social bots at 11.09%), followed by *Using Cannabis* (non-bots at 11.89%
189 versus social bots 3.88%) and *Health and Medical* (non-bots at 5.61% versus
190 social bots at 10.13%) (**Figure 2**).

191 **Discussion**

192 This study is one of the largest Twitter studies to date focused on
193 cannabis-related conversations, describing over 60,000 unique posts from
194 over 40,000 unique accounts. We identified a number of important, novel
195 topics of conversation ranging from cannabis initiation to health claims about
196 cannabis' ability to relieve an array of ailments, including depression and
197 trauma. Posts discussed edibles, hemp, legalization, buying products, and
198 cannabis' appeal or abuse liability, among other topics. We found that
199 Twitter users often discussed polysubstance use, with beer, wine, vodka,
200 tequila, cocaine, heroin, ecstasy, LSD, meth, mushrooms, and Xanax
201 mentioned along with cannabis. This is also the first study to date to
202 distinguish cannabis-related topics of conversations by social bots and non-
203 bots on Twitter. When we compared posts from non-bots to post from social
204 bots, we found that some topics comprised similar proportions, while other
205 topics stood out with differences. For example, posts indicating that cannabis
206 could allay health concerns represented a larger proportion of posts by social
207 bots compared to non-bots. Unsubstantiated health claims perpetuated by
208 social bots may have offline consequences, such as leaving Twitter users
209 with the impression that cannabis use can allay problems that have not been
210 scientifically supported.

211 In line with previous research,^{15,16} *Person Tagging* was a predominant
212 theme in the current study of cannabis-related posts to Twitter. The act of
213 person tagging is indicative of a distinct communicative practice where
214 Twitter users communicate their attitudes and experiences with cannabis.

215 Posts classified under *Person Tagging* consistently used @Person to involve
216 others in conversations about cannabis. These online messages may impact
217 cannabis use. For example, Cabrera-Nguyen and colleagues found that
218 current cannabis use was significantly associated with higher levels of
219 exposure to pro-cannabis content on Twitter among young adults.²⁴ Roditis
220 and colleagues demonstrated that adolescents who reported seeing
221 messages about the benefits of cannabis use on social media were more
222 likely to report cannabis use than adolescents reporting not seeing such
223 messages.²⁵ The current study's findings should be important to the public
224 health community, as repeated exposure to pro-cannabis messaging, and
225 cannabis use by others, can influence the social norms of those exposed to
226 the content and lead to imitation of the behaviors.²⁶

227 *Using Cannabis*, including initiation of use, was a prevalent topic in the
228 current study. Cannabis-related posts to Twitter can also be leveraged for
229 intervention efforts to curb initiation.²⁷ Interventions could be designed to
230 engage with Twitter users posting about their first-time experience with
231 cannabis. Such interventions could inform participants experimenting with
232 cannabis about the health consequences of use in hopes to prevent
233 dependence. *Processed Product Usage* and *CBD and Hemp Use* were also
234 common topics in the current study, and similar to prior research. For
235 example, analysis of edible-related posts to Twitter demonstrated that
236 cannabis edibles were generally positively perceived among Twitter users
237 despite some posts suggesting that edibles were unreliable (e.g., variability

238 in effect intensity and duration).⁷ Cavazos-Rehg and colleagues analyzed a
239 sample of edible-related post to Twitter from one month in 2015 and found
240 that most posts normalized or encouraged edibles use and described the
241 intense or long-lasting effects following use.⁸ A content analysis of tweets
242 about high-potency cannabis demonstrated that posts often mentioned the
243 physiological and psychological effects from use, and that the most common
244 physiologic effects were passing out, and respiratory effects, such as
245 coughing.²⁸ Despite positive perceptions of cannabis concentrates, the
246 amount of tetrahydrocannabinol (THC), the cannabinoid responsible for
247 intoxication, found in some cannabis concentrates, and extracts, can range
248 from 50 to over 80%,²⁹ while high potency cannabis plants have THC levels
249 around 20-25%.^{4,30} While not lethal, overdoses are common when using
250 cannabis concentrates and extracts,³¹ and high amounts of THC may produce
251 acute psychotic symptoms (e.g., hallucinations, delusions, and anxiety) in
252 some users.³² Interventions could be designed to monitor Twitter in hopes of
253 identifying high risk cannabis users. Messages could be designed to inform
254 Twitter users about the risks of high potency cannabis products in hopes of
255 reducing overdoses and dependence.

256 Analyses of cannabis-related posts to Twitter from one month in 2014
257 identified a number of themes including a desire to use cannabis, mentions
258 of health benefits, legalization efforts, and frequency of use.¹¹ *Appeal or*
259 *Abuse Liability, Health and Medical, and Legality* were common topics in the
260 current study suggesting continuity of cannabis-related discussions on

261 Twitter over the past few years. Since 2014, nine states have voted to
262 legalize medical cannabis,² and six states have voted to legalize recreational
263 adult use of cannabis.² Among adults living in a state with legalized
264 recreational cannabis, more than half (54.8%) reported seeing cannabis
265 advertising during the past month, regardless of individual cannabis use.³³ As
266 states continue to legalize access to cannabis products, cannabis-related
267 discussions about product appeal, health claims and legalization may
268 continue to be popular topics on Twitter. Local and state public health
269 agencies can use Twitter to disseminate evidence-based information about
270 cannabis use.

271 Krauss and colleagues analyzed a sample of posts to Twitter from one
272 month in 2014, and found that tweets commonly mentioned polysubstance
273 use (cannabis and alcohol).¹² Our study corroborates the findings from this
274 initial study, and extends them by demonstrating Twitter users report
275 cannabis use along with other substances, including heroin, ecstasy, LSD,
276 cocaine and prescription drugs. There is moderate evidence that cannabis
277 use is likely to increase the risk of developing a substance abuse disorder for
278 other substances, including alcohol, tobacco, and other illicit drugs, with
279 some studies indicating that the risk of developing substance use disorders
280 is higher among younger cannabis users.⁴

281 In contrast to prior Twitter studies focused on substance (nicotine)
282 use,¹⁵ the current study found few posts indicative of underage use of
283 cannabis. Thompson and colleagues assessed cannabis-related content

284 posted to Twitter by adolescents, finding that a majority of the tweets
285 reflected a positive attitude toward cannabis, and 42.9% indicated personal
286 use.³⁴ Adolescents' posts also suggested that their parents were supportive
287 of their cannabis use.³⁴

288 This study found that posts from social bots comprised smaller
289 proportions of *Person Tagging*, but almost double the proportion of posts
290 pertaining to *Health and Medical* compared to non-bots. Social bots have
291 previously been found to spread unsubstantiated health claims on Twitter.
292 For example, in 2017, Allem and colleagues found that social bots were more
293 than two times as likely to make claims about the effectiveness of electronic
294 cigarettes in smoking cessation compared to non-bots.²³ In 2018, Martinez
295 and colleagues found similar results, suggesting that the majority of the
296 Twitter data they collected pertaining to e-cigarettes originated from social
297 bots, and often touted the use of e-cigarettes in cessation.²² Most recently,
298 Broniatowski and colleagues reported that social bots were responsible for
299 disseminating antivaccine messages in the U.S.²¹ Taken all together,
300 unsubstantiated health claims perpetuated by social bots may have offline
301 consequences, such as leaving Twitter users with the impression that
302 cannabis use can allay problems that have not been scientifically supported.
303 The findings from the current study may serve as an early warning. Tech
304 companies, like Twitter, have shown concern over misinformation appearing
305 on their platforms. Cannabis may be an emerging area for misinformation on

306 Twitter, requiring the company to regulate content internally, or face
307 external regulations from state or federal agencies.

308 *Limitations*

309 This study focused on posts to Twitter, and findings may not extend to
310 other social media platforms. The posts in this study were collected from an
311 eight-month period and may not extend to other time periods. Data
312 collection relied on Twitter’s Streaming API, which prevented collection of
313 posts from private accounts. Findings may not generalize to all Twitter users
314 or to the U.S. population. Not all tweets were covered by the established
315 categories, and topics of conversation were not segmented by geographic
316 location, preventing this study from understanding the impact of different
317 state cannabis policies on the public’s experience with cannabis. In some
318 instances, one-grams and bi-grams used to define topics may have multiple
319 meanings that were ignored in the current study. For example, the phrase
320 “first time” in cannabis-related posts may not always indicate initiation but
321 rather describe an established cannabis user in a novel situation or context.
322 Similarly, it is unclear that the word “school” always identifies underage use,
323 as college students or other educational professional may be adult cannabis
324 users.

325 *Public Health Implications*

326 Use of cannabis as well as initiation, health-related claims about
327 cannabis products, and polysubstance use were common contexts
328 associated with Twitter posts about cannabis. These results suggest that

329 high potency cannabis products, unsubstantiated health claims about
330 cannabis products, and the co-use of cannabis with legal and illicit
331 substances warrant considerations by public health researchers in the future.
332 Twitter may be a platform to engage with those experimenting with cannabis
333 as well as established cannabis users to inform them of the potential for
334 cannabis dependence and additional health consequences of use. This study
335 also highlights the ability of Twitter data to help understand the public's
336 recent experiences with cannabis.

337 **Acknowledgments**

338
339 The authors have no conflicts of interest to declare.
340

341 **References**

- 342
343 1. Leung J, Chiu CY, Stjepanović D, Hall W. Has the legalisation of medical
344 and recreational cannabis use in the USA affected the prevalence of
345 cannabis use and cannabis use disorders?. *Curr Addict Rep.* 2018;5(4):403-
346 17.
347
348 2. National Conference of State Legislators. From
349 [http://www.ncsl.org/research/civil-and-criminal-justice/marijuana-](http://www.ncsl.org/research/civil-and-criminal-justice/marijuana-overview.aspx)
350 [overview.aspx](http://www.ncsl.org/research/civil-and-criminal-justice/marijuana-overview.aspx) Published May 28, 2019. Accessed June 4, 2019.
351
352 3. National Institute on Drug Abuse. Marijuana. From
353 <https://www.drugabuse.gov/drugs-abuse/marijuana>. Accessed June 4, 2019.
354
355 4. National Academies of Sciences, Engineering, and Medicine. The health
356 effects of cannabis and cannabinoids: The current state of evidence and
357 recommendations for research. National Academies Press; Published March
358 31, 2017.
359
360 5. Rubino T, Zamberletti E, Parolaro D. Adolescent exposure to cannabis as a
361 risk factor for psychiatric disorders. *J Psychopharmacol.* 2012;26(1):177-188.
362
363 6. Auer R, Vittinghoff E, Yaffe K, et al. Association between lifetime marijuana
364 use and cognitive function in middle age: the Coronary Artery Risk

- 365 Development in Young Adults (CARDIA) Study. *JAMA Intern Med.*
366 2016;176(3):352-61.
367
- 368 7. Lamy FR, Daniulaityte R, Sheth A, et al., "Those edibles hit hard":
369 Exploration of Twitter data on cannabis edibles in the US. *Drug Alcohol*
370 *Depend.* 2016;164:64-70.
371
- 372 8. Cavazos-Rehg PA, Zewdie K, Krauss MJ, Sowles SJ. "No high like a brownie
373 high": a content analysis of edible marijuana tweets. *Am J Health Promot.*
374 2018;32(4):880-6.
375
- 376 9. Perrin A, Anderson M. Share of U.S. adults using social media, including
377 Facebook, is mostly unchanged since 2018. From
378 [https://www.pewresearch.org/fact-tank/2019/04/10/share-of-u-s-adults-using-](https://www.pewresearch.org/fact-tank/2019/04/10/share-of-u-s-adults-using-social-media-including-facebook-is-mostly-unchanged-since-2018/)
379 [social-media-including-facebook-is-mostly-unchanged-since-2018/](https://www.pewresearch.org/fact-tank/2019/04/10/share-of-u-s-adults-using-social-media-including-facebook-is-mostly-unchanged-since-2018/) Published
380 April 10, 2019. Accessed June 4, 2019.
381
- 382 10. Anderson M, Jiang J. Teens, social media & technology. From
383 [https://www.pewinternet.org/2018/05/31/teens-social-media-technology-](https://www.pewinternet.org/2018/05/31/teens-social-media-technology-2018/)
384 [2018/](https://www.pewinternet.org/2018/05/31/teens-social-media-technology-2018/) Published May 31, 2018. Accessed June 4, 2019
385
- 386 11. Cavazos-Rehg PA, Krauss M, Fisher SL, Salyer P, Grucza RA, Bierut LJ.
387 Twitter chatter about marijuana. *J Adolesc Health.* 2015;56(2):139-45.
388
- 389 12. Krauss MJ, Grucza RA, Bierut LJ, Cavazos-Rehg PA. "Get drunk. Smoke
390 weed. Have fun.": A content analysis of tweets about marijuana and alcohol.
391 *Am J Health Promot.* 2017;31(3):200-8.
392
- 393 13. Montgomery L, Yockey A. Rolling and scrolling: The portrayal of
394 marijuana cigars (blunts) on YouTube. *J Subst Use.* 2018;23(4):436-40.
395
- 396 14. Meacham MC, Paul MJ, Ramo DE. Understanding emerging forms of
397 cannabis use through an online cannabis community: An analysis of relative
398 post volume and subjective highness ratings. *Drug and alcohol dependence.*
399 *Drug Alcohol Depend.* 2018;188:364-9.
400
- 401 15. Allem JP, Dharmapuri L, Unger JB, Cruz TB. Characterizing JUUL-related
402 posts on Twitter. *Drug Alcohol Depend.* 2018;190:1-5.
403
- 404 16. Allem JP, Dharmapuri L, Leventhal AM, Unger JB, Cruz TB. Hookah-related
405 posts to Twitter from 2017 to 2018: thematic analysis. *J Med Internet Res.*
406 2018;20(11):e11669.
407
- 408 17. Vitter JS. Random sampling with a reservoir. *ACM Transactions on*
409 *Mathematical Software (TOMS).* 1985;1(1):37-57.
410

- 411 18. Allem JP, Ferrara E. Could social bots pose a threat to public health? *Am J*
412 *Public Health*. 2018;108(8):1005-1006.
413
- 414 19. Allem JP, Ferrara E. The importance of de-biasing social media data to
415 better understand e-cigarette-related attitudes and behaviors. *J Med Internet*
416 *Res*. 2016;18(8):e219.
417
- 418 20. Ferrara E, Varol O, Davis C, Menczer F, Flammini A. The rise of social
419 bots. *Communications of the ACM* 2016;59(7):96-104.
420
- 421 21. Broniatowski DA, Jamison AM, Qi S, AlKulaib L, Chen T, Benton A, Quinn
422 SC, Dredze M. Weaponized health communication: Twitter bots and Russian
423 trolls amplify the vaccine debate. *Am J Public Health*. 2018;108(10):1378-84.
424
- 425 22. Martinez LS, Hughes S, Walsh-Buhi ER, Tsou MH. "Okay, we get it. You
426 vape": An analysis of geocoded content, context, and sentiment regarding e-
427 cigarettes on Twitter. *J Health Commun*. 2018;23(6):550-62.
428
- 429 23. Allem JP, Ferrara E, Uppu SP, Cruz TB, Unger JB. E-cigarette surveillance
430 with social media data: social bots, emerging topics and trends. *JMIR Public*
431 *Health Surveill*. 2017;3(4):e98.
432
- 433 24. Cabrera-Nguyen EP, Cavazos-Rehg P, Krauss M, Bierut LJ, Moreno MA.
434 Young adults' exposure to alcohol-and marijuana-related content on Twitter.
435 *J Stud Alcohol Drugs*. 2016;77(2):349-53.
436
- 437 25. Roditis ML, Delucchi K, Chang A, Halpern-Felsher B. Perceptions of social
438 norms and exposure to pro-marijuana messages are associated with
439 adolescent marijuana use. *Prev Med*. 2016;93:171-6.
440
- 441 26. Fujimoto K, Valente TW. Social network influences on adolescent
442 substance use: disentangling structural equivalence from cohesion. *Soc Sci*
443 *Med*. 2012;74(12):1952-60.
444
- 445 27. Moreno MA, Kerr B, Lowry SJ. A longitudinal investigation of associations
446 between marijuana displays on Facebook and self-reported behaviors among
447 college students. *J Adolesc Health*. 2018;63(3):313-9.
448
- 449 28. Cavazos-Rehg PA, Sowles SJ, Krauss MJ, Agbonavbare V, Grucza R, Bierut
450 L. A content analysis of tweets about high-potency marijuana. *Drug Alcohol*
451 *Depend*. 2016;166:100-8.
452
- 453 29. National Institute on Drug Abuse. A Rise in Marijuana's THC Levels. From
454 <https://archives.drugabuse.gov/rise-in-marijuanas-thc-levels>. Accessed June
455 4, 2019.
456

457 30. ElSohly MA, Mehmedic Z, Foster S, Gon C, Chandra S, Church JC. Changes
458 in Cannabis Potency Over the Last 2 Decades (1995-2014): Analysis of
459 Current Data in the United States. *Biol Psychiatry*. 2016;79(7):613-619.
460

461 31. Barrus DG, Capogrossi KL, Cates SC, et al. Tasty THC: Promises and
462 Challenges of Cannabis Edibles. *Methods Rep RTI Press*. 2016;
463 doi:10.3768/rtipress.2016.op.0035.1611
464

465 32. Wilkinson ST, Radhakrishnan R, D'Souza DC. Impact of cannabis use on
466 the development of psychotic disorders. *Curr Addict Rep*. 2014;1(2):115-28.
467

468 33. Fiala SC, Dilley JA, Firth CL, Maher JE. Exposure to marijuana marketing
469 after legalization of retail sales: Oregonians' experiences, 2015-2016. *Am J*
470 *Public Health*. 2018;108(1):120-7.
471

472 34. Thompson L, Rivara FP, Whitehill JM. Prevalence of marijuana-related
473 traffic on Twitter, 2012-2013: a content analysis. *Cyberpsychol Behav Soc*
474 *Netw*. 2015;18(6):311-9.
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502

503
 504
 505
 506
 507
 508
 509
 510
 511
 512
 513
 514
 515

Figure 1 Prevalence of topics from non-bot corpus of tweets

Person Tagging	20450, 33.60%												
Using Cannabis	1965, 3.23%	7235, 11.89%											
Health and Medical	1231, 2.02%	218, 0.36%	3416, 5.61%										
Legality	1434, 2.36%	274, 0.45%	279, 0.46%	3345, 5.50%									
Buy/Sell	939, 1.54%	136, 0.22%	193, 0.32%	201, 0.33%	3051, 5.01%								
Processed Product Usage	948, 1.56%	62, 0.10%	127, 0.21%	71, 0.12%	183, 0.30%	2260, 3.71%							
CBD and Hemp Use	503, 0.83%	11, 0.02%	345, 0.57%	53, 0.09%	177, 0.29%	165, 0.27%	1661, 2.73%						
Appeal or Abuse Liability	267, 0.44%	160, 0.26%	29, 0.05%	15, 0.02%	51, 0.08%	17, 0.03%	10, 0.02%	1614, 2.65%					
Polysubstance Use	618, 1.02%	241, 0.40%	42, 0.07%	79, 0.13%	60, 0.10%	43, 0.07%	18, 0.03%	20, 0.03%	1504, 2.47%				
Cannabis Industry	217, 0.36%	22, 0.04%	34, 0.06%	75, 0.12%	47, 0.08%	1, 0.00%	7, 0.01%	2, 0.00%	11, 0.02%	808, 1.33%			
Impairment	107, 0.18%	77, 0.13%	12, 0.02%	32, 0.05%	13, 0.02%	6, 0.01%	0, 0.00%	5, 0.01%	4, 0.01%	0, 0.00%	288, 0.47%		
Underage Use	87, 0.14%	68, 0.11%	2, 0.00%	6, 0.01%	17, 0.03%	7, 0.01%	0, 0.00%	3, 0.00%	25, 0.04%	0, 0.00%	2, 0.00%	215, 0.35%	
	Person Tagging	Using Cannabis	Health and Medical	Legality	Buy/Sell	Processed Product Usage	CBD and Hemp Use	Appeal or Abuse Liability	Polysubstance Use	Cannabis Industry	Impairment	Underage Use	

516
 517
 518
 519
 520
 521
 522
 523
 524
 525
 526
 527
 528
 529
 530

531
532
533
534
535
536
537
538
539
540
541
542
543

Figure 2 Comparison of prevalence of topics between social bots and non-bots

Topic	Non-Bots	Bots	Delta
Appeal or Abuse Liability	2.65%	0.59%	2.07%
Buy/Sell	5.01%	4.73%	0.28%
CBD and Hemp Use	2.73%	6.41%	-3.68%
Cannabis Industry	1.33%	3.13%	-1.81%
Health and Medical	5.61%	10.13%	-4.52%
Impairment	0.47%	0.20%	0.27%
Legality	5.50%	4.88%	0.62%
Person Tagging	33.60%	11.09%	22.51%
Polysubstance Use	2.47%	1.15%	1.32%
Processed Product Usage	3.71%	4.02%	-0.31%
Underage Use	0.35%	0.10%	0.25%
Using Cannabis	11.89%	3.88%	8.01%

544