Original Paper

Patterns of Twitter Behavior Among Networks of Cannabis Dispensaries in California

Nicholas C Peiper[1*], MPH, PhD; Peter M Baumgartner[2*], MS; Robert F Chew[2], MS; Yuli P Hsieh[3], PhD; Gayle S Bieler[2], MS; Georgiy V Bobashev[2], PhD; Christopher Siege[4], MS; Gary A Zarkin[1], PhD

[1]RTI International, Behavioral Health and Criminal Justice Research Division, Research Triangle Park, NC, United States

[2]RTI International, Center for Data Science, Research Triangle Park, NC, United States

[3]RTI International, Survey Research Division, Research Triangle Park, NC, United States

[4]RTI International, Research Computing Division, Research Triangle Park, NC, United States

[*]these authors contributed equally

Corresponding Author:
Nicholas C Peiper, MPH, PhD
RTI International
Behavioral Health and Criminal Justice Research Division
3040 E. Cornwallis Road
PO Box 12194
Research Triangle Park, NC, 27709
United States
Phone: 1 919 316 3314
Fax: 1 919 485 5555
Email: npeiper@rti.org

Abstract

Background: Twitter represents a social media platform through which medical cannabis dispensaries can rapidly promote and advertise a multitude of retail products. Yet, to date, no studies have systematically evaluated Twitter behavior among dispensaries and how these behaviors influence the formation of social networks.

Objectives: This study sought to characterize common cyberbehaviors and shared follower networks among dispensaries operating in two large cannabis markets in California.

Methods: From a targeted sample of 119 dispensaries in the San Francisco Bay Area and Greater Los Angeles, we collected metadata from the dispensary accounts using the Twitter API. For each city, we characterized the network structure of dispensaries based upon shared followers, then empirically derived communities with the Louvain modularity algorithm. Principal components factor analysis was employed to reduce 12 Twitter measures into a more parsimonious set of cyberbehavioral dimensions. Finally, quadratic discriminant analysis was implemented to verify the ability of the extracted dimensions to classify dispensaries into their derived communities.

Results: The modularity algorithm yielded three communities in each city with distinct network structures. The principal components factor analysis reduced the 12 cyberbehaviors into five dimensions that encompassed account age, posting frequency, referencing, hyperlinks, and user engagement among the dispensary accounts. In the quadratic discriminant analysis, the dimensions correctly classified 75% (46/61) of the communities in the San Francisco Bay Area and 71% (41/58) in Greater Los Angeles.

Conclusions: The most centralized and strongly connected dispensaries in both cities had newer accounts, higher daily activity, more frequent user engagement, and increased usage of embedded media, keywords, and hyperlinks. Measures derived from both network structure and cyberbehavioral dimensions can serve as key contextual indicators for the online surveillance of cannabis dispensaries and consumer markets over time.

KEYWORDS

Introduction

Dramatic population-based shifts in cannabis use have occurred over the past 15 years in the United States [1]. As of July 2017, 29 states and Washington DC have enacted laws that permit medical cannabis use. Much research has helped to understand individuals who use cannabis for medical purposes [2], ranging from their consumption patterns and motivations for use to service satisfaction and clinical preferences [3-6]. Similar efforts have explored how state laws differentially impact operation and enforcement of cannabis businesses, health centers, and cultivation practices [7,8]. Nevertheless, significant public debate remains about the medicinal value of cannabis given the large body of clinical and population-based studies showing increased risk of many adverse outcomes [8], especially with regard to the effects of high potency strains and concentrated products like edibles [9,10].

Notably, these debates coincide with the growing availability of medical and recreational cannabis products at dispensaries across the United States. In California, the world's largest legal market for cannabis, medical cannabis patients report that they vary their purchasing behaviors based upon product pricing and availability at dispensaries as well as the specific conditions for which they received a physician recommendation [11]. Patients also report that experiences and interactions with dispensary staff like budtenders greatly influence their purchasing behaviors, including their willingness to try new products [12]. Because dispensaries serve as the purveyors of cannabis products and strongly influence population-based consumption [11,12], many advertise their products and services through a wide variety of platforms, including social network platforms like Twitter.

While it is estimated that 1 in 2000 tweets pertain to cannabis [13], there are currently no studies that specifically focus on how dispensaries use Twitter to engage with medical cannabis patients and their larger follower base. This gap is particularly salient as content analyses of influential Twitter users show that cannabis-related tweets tend to elicit positive sentiments towards cannabis, including heavy and frequent use behaviors [13]. A recent study found that WeedTweets (@stillblazingtho), a Twitter account with over 1 million followers, posts an average of 10 tweets per day and that these tweets tend to normalize regular cannabis use, especially among youth and certain minority populations [14]. In addition, other studies have detected higher frequencies of tweets related to cannabis concentrates (eg, edibles, dabs, and oils) in states that allow medical and recreational consumption [15,16], which may be partially attributable to increasingly permissive and accepting attitudes toward cannabis.

Considering the well-documented impacts of social networks on consumer preferences and behaviors [17-19], an explicit focus on the exchange of cannabis-related information may provide valuable insights into how networks of cannabis consumers form around dispensaries on Twitter. For instance, some dispensaries regularly use Twitter to share their menus, inform their followers about new products, offer coupons and promotions, promote retail services, post industry trends and events, and mention findings from scientific studies. Other dispensaries, however, may engage in these practices less frequently or have more sporadic Twitter usage, which could influence their ability to form strong and sustained networks of followers.

More importantly, systematic investigation of dispensaries on Twitter can provide insights into how dispensaries behave on the Internet and how these behaviors influence the formation of shared follower communities. This comparative study therefore examines a set of 12 Twitter cyberbehaviors among two samples of cannabis dispensaries from the San Francisco Bay Area (SFBA) and Greater Los Angeles (GLA). For each metropolitan area, we visualize overall network structure and community formation based upon shared followers, then reduce the cyberbehaviors into a more parsimonious set of dimensions with principal components factor analysis. Finally, we utilize quadratic discriminant analysis to investigate whether the extracted dimensions of cyberbehavior significantly differentiate between dispensary communities in California.

Methods

Study Sample

We adapted aspects of targeted sampling methods to select cannabis dispensaries in SFBA and GLA. Traditionally, these methods have been used in social science and public health studies to access "hidden" populations (eg, medical cannabis patients or people who inject drugs) outside of community or medical settings [20]. Targeted sampling integrates components of street ethnography, theoretical sampling, stratified survey sampling, quota sampling, and respondent-driven sampling [21-24]. As it improves upon convenience samples through a purposive and rigorous process, a growing body of studies has used targeted sampling to recruit representative samples that are comparable to those achieved through random sampling techniques [25-28].

In the context of cannabis dispensaries in California, we included licensed, registered, and commercially zoned dispensaries from the Medical Dispensary Program in San Francisco, the Cannabis Regulatory Commission in Oakland, the Medical Cannabis Commission in Berkeley, and the Medical Marijuana ID Program in Los Angeles. We then cross-referenced the initial database with Leafly, WeedMaps, and THCFinder, three popular cannabis sites that allow users to geolocate dispensaries throughout the United States, including California. These sites include streamlined platforms with comprehensive information about social network profiles, which allowed us to expand our initial database and create a larger catchment of dispensary accounts on Twitter. With this final sample of dispensary accounts, we collected the account IDs of followers and the last 3200 tweets available as of February 16, 2016. Finally, a set of 12 cyberbehaviors were derived with metadata from the accounts and fell into three broad categories: account age, posting frequency, and tweet composition (Table 1).

Table 1. Definitions for Twitter cyberbehaviors.

Cyberbehaviors	Definition
Account age	
Overall age	Number of days a Twitter account has existed
Total days tweeting	Number of days at least one tweet was sent from an account
Posting frequency	
Tweets collected	Total number of tweets collected from an account timeline
Percentage of days tweeting	Percentage of days since an account was created that there has been a tweet
Max. tweets per day	Maximum number of times an account has posted a tweet in a single day
Average tweets per day	Mean number of times an account tweets per day[a]
Median absolute deviation	Median absolute deviation (MAD) of tweets per day
Tweet composition	
Hashtag (#)	Percentage of tweets collected that contained a hashtag
Mention (@)	Percentage of tweets collected that mentioned another user directly
Retweet (RT)	Percentage of tweets collected that were retweets
Media	Percentage of tweets collected that contained embedded media[b]
Hyperlink (http://)	Percentage of tweets collected that contained a hyperlink

[a]Excludes days on which an account did not tweet.

[b]Images, videos, and documents.

Network Structure and Community Detection

With the account information for each dispensary and their followers, we created a projection from the dispensary networks with edge weights representing shared followership [29]. Because the sampled dispensary accounts had a highly right-skewed distribution of followers, we normalized follower counts between two dispensaries by calculating the ratio of shared followers to potential shared followers, where *potential shared followers* was the minimum of the follower counts of the two dispensaries. With similarities to the Jaccard index, this measurement computed a projection function that determined the shared potential followers between dispensaries [30,31]. As depicted in Figure 1, two hypothetical dispensaries share four followers out of a total of six potential shared followers, yielding a projection function of 66%.

The Louvain Method [32] was then utilized to detect dispensary communities in SFBA and GLA. This unsupervised algorithm finds communities of large networks and provides a hierarchical structure for the network through an iterative, two-stage process that maximizes modularity. The method first began by selecting a random dispensary (ie, node) and assigning that dispensary to a community of one of its neighboring dispensaries, until all existing dispensaries in the network were assigned to a community. In the second phase, each dispensary represented a community from phase one, while edges between dispensaries represented the sum of the weights of the previous connections between dispensaries in those two communities. These two phases of optimizing modularity and constructing a meta-network in each city were repeated until a network with the maximum value of modularity was found.

After creating the network data and calculating communities, we visualized each city's network with a force-directed graph drawing algorithm. This network visualization algorithm places nodes with more shared follower potential closer to each other and repulses nodes with limited or no potential. For the purposes of this study, dispensaries from the same community were visualized using colored nodes.

XSL•FO

RenderX

Figure 1. Hypothetical shared follower network.
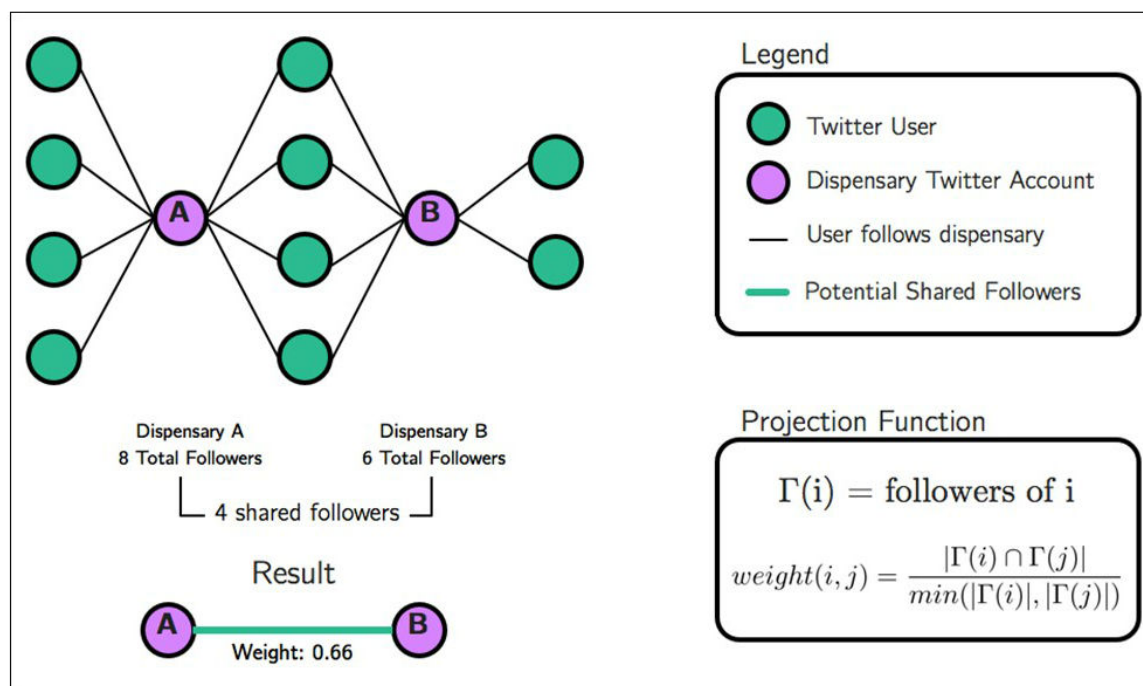
Cyberbehavioral Dimensions and Community Classification

For the 12 cyberbehaviors, descriptive statistics were computed to characterize each city's set of communities. Wilcoxon rank-sum and Kruskal-Wallis tests were also performed to explore any statistically significant differences between cities and communities for the 12 cyberbehaviors. We then conducted principal components factor analysis (PCA) with a 12x12 correlation matrix of the cyberbehaviors to extract empirically meaningful dimensions. PCA provided a method with which to address multicollinearity among the cyberbehaviors and arrive at a more parsimonious set of dimensions that account for the data variability. Lastly, we performed quadratic discriminant analysis (QDA) to determine the classification accuracy of the extracted cyberbehaviors [33]. For the purpose of this study, QDA produced classification tables for each city, which allowed for distinguishing between modularity classes with the extracted cyberbehavioral dimensions from the PCA.

Results

Descriptive Statistics

Overall, a total of 119 dispensary accounts were examined, with 61 in SFBA and 58 in GLA. The mean values for each cyberbehavior are shown for the two cities in Table 2. Each account in SFBA and GLA was approximately three years old on average. The cyberbehaviors for posting frequency and tweet frequency were highly comparable between the two cities. Although dispensary accounts in SFBA spent a higher number of days tweeting and sent more tweets than accounts in GLA, no significant differences were found.

Network Structure and Community Formation

Figure 2 visualizes the two shared follower networks of dispensaries in SFBA and GLA. The size of the nodes corresponds to the total number of followers, the thickness of the edges indicates the shared follower potential, and the color of the nodes refers to the community classifications from the Louvain method. Overall, the distribution of shared followers between each pair of accounts differed between SFBA and GLA. The range of shared followers was .2% to 71% in SFBA compared with 3% to 46% in GLA.

Among the SFBA networks, 21% (n=13, marked in green) of dispensaries were in a weakly connected community with all members having a modest number of Twitter followers. Another 38% (n=23, marked in orange) were in a fairly centralized community of dispensaries with strong interconnections between smaller accounts. The largest community accounted for 41% of the sample (n=25, marked in purple) and had strong interconnections through the most popularly followed dispensary. In GLA, a community accounting for 38% (n=22, marked in orange) of the network had the two dispensaries with the most followers, although its members were weakly connected. A small and weakly connected network accounted for 17% of the network (n=10, marked in green), with only two that had a relatively large group of followers. Despite only a modest number of followers on Twitter, the remaining 45% of the network (n=26, marked in purple) formed the largest and most strongly connected community, indicating a substantial portion of shared followers between any given pair of members.

Table 2. Descriptive statistics for Twitter cyberbehaviors.

Cyberbehaviors	SFBA[a] (n=61)	GLA[b] (n=58)	P value[c]
	Mean	Mean	
Account Age, Days (Years)	1107.8 (3.0)	1006.2 (2.8)	.49
Total Days Tweeting	285.5	202.6	.14
Tweets Collected	965.4	590.3	.21
Max. Tweets Per Day	15.1	16.4	.87
Average Tweets Per Day	3.0	2.9	.72
MAD[d] Tweets Per Day	0.8	0.8	.92
Percentage of Days Tweeting	25.9	24.1	.34
Percentage of Tweets with Media	20.4	21.0	.98
Percentage of Tweets with #[e]	40.4	40.4	.92
Percentage of Tweets with @[f]	26.1	27.6	.54
Percentage of Tweets with RT[g]	10.2	10.5	.63
Percentage of Tweets with Hyperlink	55.9	51.8	.47

[a]SFBA: San Francisco Bay Area.

[b]GLA: Greater Los Angeles.

[c]The P values were calculated with the Wilcoxon rank-sum tests to accommodate for the nonparametric nature of the cyberbehaviors.

[d]MAD: median absolute deviation.

[e]#=hashtag.

[f]@=user mention.

[g]RT: Retweet.

In the subgraphs (Figure 3), we recalibrated the tie-strength and considered an edge to be present when the proportion of shared followers of a given pair of dispensaries was above the 95th percentile of shared follower potential between any given pair of dispensaries in each of the two cities. We tested various thresholds of shared follower potential: the median, the third quartile (75th percentile), 90th percentile and 95th percentile. As highly consistent network graphs were found, the 95th percentile was used as the final threshold to produce the subgraphs with the strongest tie-strength in the social networks with the best visual clarity.

As Figure 3 illustrates, one dispensary is particularly popular in SFBA, where it not only attracts substantially more followers than its counterparts, but also has stronger connections to the followers. In contrast, large dispensaries in GLA are not as strongly connected and centralized as those in SFBA, given that they do not share many followers (Figure 3). Although four of them seem to be very popular with a large group of followers, they occupy different network positions and attract different groups of Twitter users through a smaller but highly centralized and interconnected dispensary. Additionally, there was a cluster of well-connected dispensaries with a group of small, but mostly shared followers.

Cyberbehavioral Dimensions

The mean values for the cyberbehaviors are summarized for the extracted communities in Multimedia Appendix 1. In SFBA, the weakly connected orange community had lower rates of maximum tweets per day, average tweets per day, hashtags, and user mentions, despite having the highest number and percentage of days where a tweet was sent. In comparison, the moderately connected green and strongly connected purple communities had higher frequencies of tweets as well as more user engagement (eg, mention and retweet), hashtag usage, embedded media, and hyperlinks. Significant differences between SFBA communities were found for account age, total days tweeting, average tweets per day, and percentage of tweets with media. For the GLA dispensaries, the highly connected purple community had the highest percent of days tweeting, embedded media, hashtag usage, and hyperlinks. The weakly connected green and orange communities tended to have higher account ages, total tweet days, and total tweets. The only significant differences between GLA communities were found for account age.

Figure 2. Shared follower networks in the San Francisco Bay Area and Greater Los Angeles.
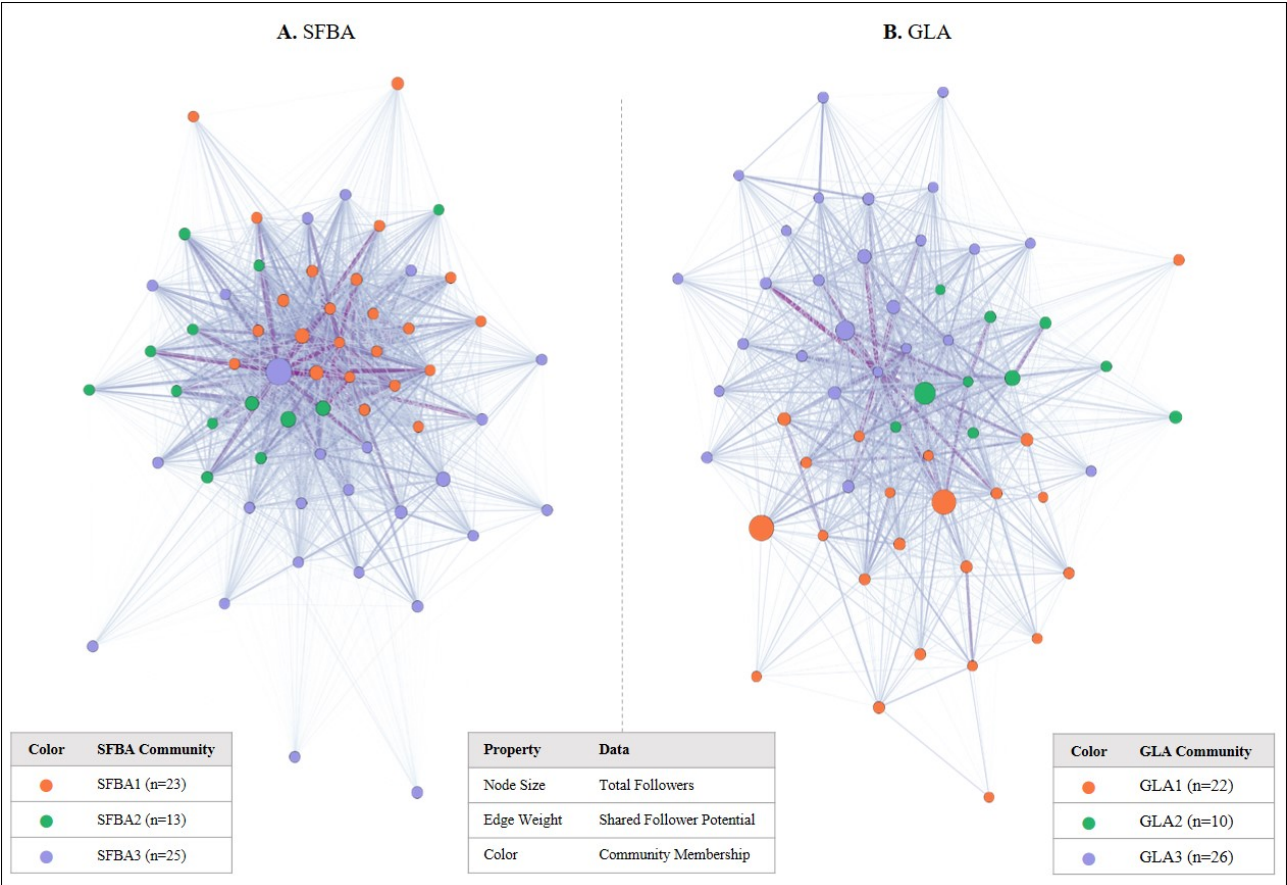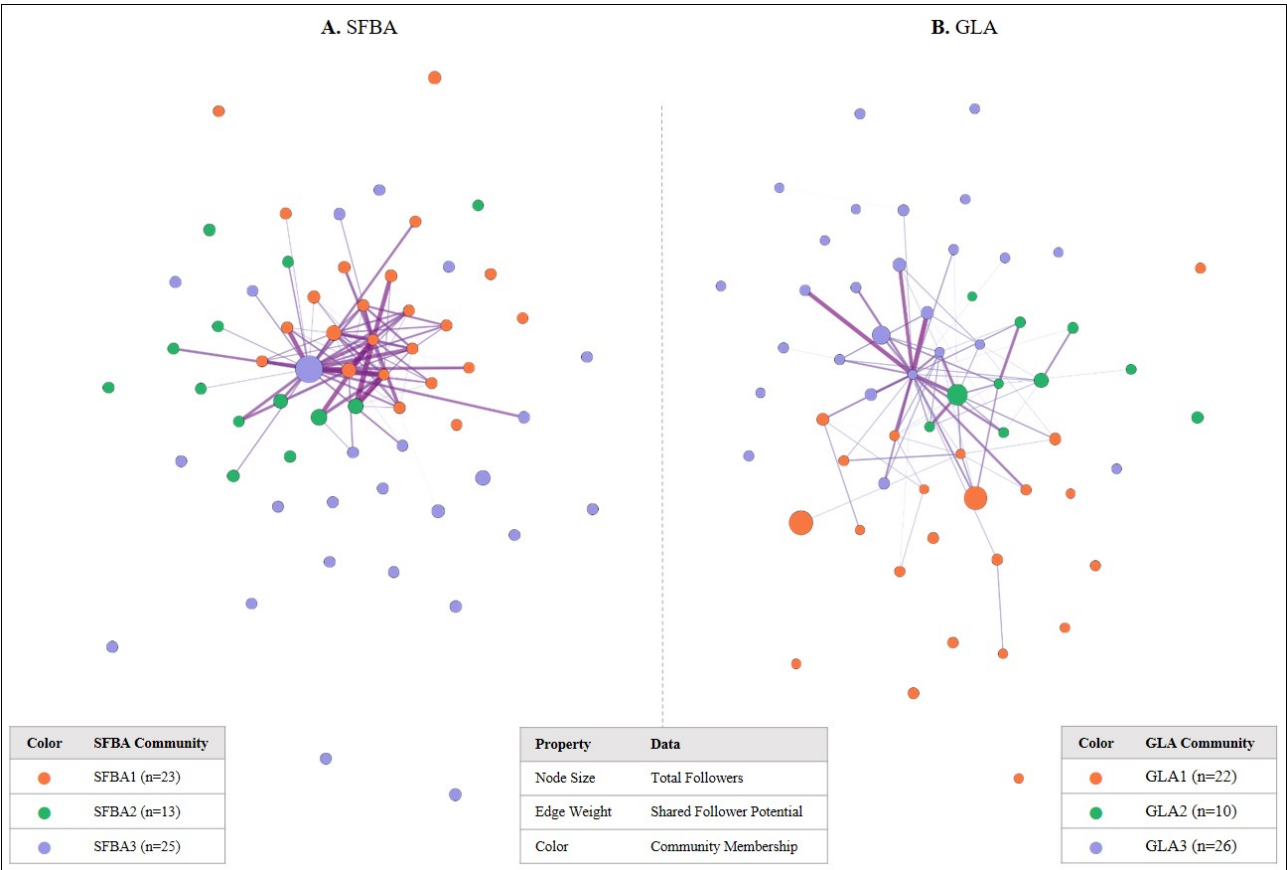

Figure 3. Shared follower network subgraphs in the San Francisco Bay Area and Greater Los Angeles.

The principal components factor analysis yielded five relevant factors (eigenvalues>1.0) that describe the Twitter behaviors of dispensaries in SFBA (Table 3). The first factor for SFBA was classified as *activity* (eigenvalue=4.3) and included three behaviors indicating the daily message frequency of users: maximum tweets per day, average tweets per day, and median absolute deviation of tweets per day. The second factor, *age* (eigenvalue=3.2), included two items: account age and percentage of tweets with media like images and videos, suggesting a specific Twitter usage pattern among SFBA dispensaries with older accounts that were less likely to include media in their tweets.

We categorized the third factor for SFBA dispensaries as *longevity* (eigenvalue=1.4) with both total days tweeting and percentage of days tweeting loading on to this dimension, followed by *engagement* (eigenvalue=1.2) with both percentage of tweets that were retweets (ie, tweets being forwarded or shared with others by users who read the original tweet) and mentions (ie, including another user account in the tweet) being loaded on this factor. The engagement dimension captures how users interact with each other on Twitter. Lastly, *referencing* (eigenvalue=1.0) included two items that link the tweet to additional information sources: percentage of tweets with hashtags (#) and hyperlinks (http://).

Table 3. Results from principal components factor analysis of the 12 cyberbehaviors in the San Francisco Bay Area.

Cyberbehaviors	Activity	Age	Longevity	Engagement	Referencing
Eigenvalues[a,b]	4.3	2.2	1.4	1.2	1.0
Account Age	0.02	**0.60**	−0.02	0.01	−0.18
Total Days Tweeting	−0.06	0.29	**0.54**	−0.05	−0.13
Percentage of Days Tweeting	−0.04	−0.20	**0.74**	0.06	0.12
Tweets Collected	0.28	0.19	0.32	−0.11	−0.03
Max. Tweets Per Day	**0.46**	0.03	0.07	0.10	−0.01
Average Tweets Per Day	**0.64**	−0.03	−0.15	−0.06	0.02
MAD[c] Tweets Per Day	**0.53**	−0.05	0.05	0.06	0.02
Percentage of Tweets with Media	0.05	**−0.64**	0.14	−0.07	−0.29
Percentage of Tweets with #[d]	−0.03	−0.21	−0.06	0.16	**0.55**
Percentage of Tweets with @[e]	0.01	−0.01	0.02	**0.68**	−0.10
Percentage of Tweets with RT[f]	−0.01	0.07	0.02	**0.68**	0.05
Percentage of Tweets with Hyperlinks	0.03	0.14	0.10	-0.11	**0.73**

[a]The presence of dimensionality was supported when eigenvalues were 1.0 or greater. Values for each cyberbehavior are expressed as varimax-rotated factor loadings.

[b]Bold factor loadings denote values greater than or equal to .40.

[c]MAD: median absolute deviation.

[d]#=hashtag.

[e]@=user mention.

[f]RT: Retweet.

We found highly similar cyberbehavioral dimensions among the GLA dispensaries (Table 4). The factor that explained the most variance was *activity* (eigenvalue=3.0), including average tweets per day and median absolute deviation of tweets per day with the largest loadings. We classified the second factor as *longevity* (eigenvalue=2.4), given that account age, total days tweeting, and total number of tweets collected significantly loaded on to this dimension. The third factor found from the GLA data was *engagement* (eigenvalue=1.8), with the same two behavioral items being loaded on to this dimension. A similar *referencing* dimension was found (eigenvalue=1.3) in GLA, although hashtags were accompanied with a significant loading for percent tweets with multimedia content when compared with hyperlinks in SFBA. For GLA, a significant loading for *hyperlinks* formed its own dimension (eigenvalue=1.3).

Table 4. Results from principal components factor analysis of the 12 cyberbehaviors in Greater Los Angeles.

Cyberbehaviors	Activity	Longevity	Engagement	Referencing	Hyperlinks
Eigenvalues[a,b]	3.0	2.4	1.8	1.3	1.2
Account Age	0.14	**0.48**	0.05	0.26	0.29
Total Days Tweeting	0.10	**0.66**	0.01	0.08	0.08
Percentage of Days Tweeting	0.16	0.20	0.25	0.39	0.29
Tweets Collected	0.22	**0.54**	0.00	0.02	0.01
Max. Tweets Per Day	0.33	0.05	0.14	0.26	0.32
Average Tweets Per Day	**0.64**	0.03	0.02	0.06	0.12
MAD[c] Tweets Per Day	**0.59**	0.03	0.05	0.08	0.09
Percentage of Tweets with Media	0.11	0.01	0.01	**0.62**	0.21
Percentage of Tweets with #[d]	0.08	0.00	0.19	**0.54**	0.09
Percentage of Tweets with @[e]	0.01	0.00	**0.66**	0.13	0.01
Percentage of Tweets with RT[f]	0.00	0.03	**0.66**	0.07	0.01
Percentage of Tweets with Hyperlinks	0.11	0.00	0.01	0.04	**0.80**

[a]The presence of dimensionality was supported when eigenvalues were 1.0 or greater. Values for each cyberbehavior are expressed as varimax-rotated factor loadings.

[b]Bold factor loadings denote values greater than or equal to .40.

[c]MAD: median absolute deviation.

[d]#=hashtag.

[e]@=user mention.

[f]RT: Retweet.

Classification Accuracy of Cyberbehavioral Dimensions

Table 5 illustrates how the communities classified by the cyberbehavioral dimensions (ie, columns) corresponded to the true communities identified through the Louvain Method (ie, rows). As depicted by the bolded diagonals, the dimensions correctly classified 75% (46/61) of the dispensary communities in SFBA. The orange community had the best classification precision, followed by the green and purple communities.

In GLA, the dimensions correctly classified 71% (41/58) of the dispensary communities, with high classification precision among the orange and purple communities (Table 6). Only 20% of the green community was correctly classified, most likely due to limited sample size. Additional loading statistics for the dimensions in the QDA may be found for each city in Multimedia Appendix 2 and interpreted like the factor loadings from the PCA.

Table 5. Classification table for the communities of dispensaries in the San Francisco Bay Area.

San Francisco Bay Area (N=61)	Classified community		
True community	Orange	Green	Purple
Orange (n=23)	**20** **(87%)[a]**	0 (0%)	3 (13%)
Green (n=13)	2 (15%)	**9** **(69%)**	2 (15%)
Purple (n=25)	3 (12%)	5 (20%)	**17** **(68%)**

[a]Bold diagonals illustrate correctly classified communities.

Table 6. Classification tables from the quadratic discriminant analysis of dispensaries in Greater Los Angeles.

Greater Los Angeles (N=58)	Classified community		
True community	Orange	Green	Purple
Orange (n=22)	**18**	0	4
	(82%)[a]	(0%)	(18%)
Green (n=10)	3	**2**	5
	(30%)	**(20%)**	(50%)
Purple (n=26)	4	1	**21**
	(15%)	(4%)	**(81%)**

[a]Bold diagonals illustrate correctly classified communities.

Discussion

Principal Findings

As a popular social network platform that enables rapid information exchange about controversial social phenomena, Twitter represents an unregulated domain where cannabis dispensaries can form communities through regular communication and engagement with large audiences. In this study, the networks in SFBA and GLA both included sets of highly influential dispensaries with large groups of shared followers. However, the network structure of SFBA was more strongly connected and centralized than that of GLA, which had four large dispensaries that occupied relatively separate network spaces. The most strongly connected dispensaries in both cities had newer accounts, higher daily activity, more frequent user engagement, and increased usage of embedded media, keywords, and hyperlinks. As such, both network structure and cyberbehaviors significantly distinguished between the communities in each city, which provides evidence for contextual indicators that can be utilized for the surveillance of information exchange among dispensaries on Twitter.

Cyberbehaviors and Distinguishable Communities

Among the large and interconnected dispensary communities, the cyberbehaviors indicated regular tweets to shared followers that may include patient, consumer, and cannabis industry populations with strong mutual interests. The younger age of these highly active dispensaries may also demonstrate the emergence of new marketing strategies that streamline product promotions, share information, and develop brand loyalty within a larger sharing economy on Twitter [34]. In addition, these communities exhibited comparatively higher user engagement and referencing, two dimensions that may reciprocate collective consumption of cannabis through Twitter-mediated interactions and cooperative cyberbehaviors that rapidly disseminate cannabis-related information [35]. Together, the structural and dimensional characteristics of these communities indicate that influential dispensaries may use Twitter to boost social traffic to their websites and grow their social networks [18,36-41].

In contrast, the dispensaries on the network periphery had lower shared follower potential and exhibited more generic cyberbehaviors (eg, text-only tweets) that do not provide followers with engaging content or links to additional resources. As populations with greater socioeconomic status are

significantly more likely to send and receive hyperlinks [42-45], these dispensaries may lack the resources to engage in cyberbehaviors that place them in more densely connected network spaces characterized by regular communications and strategic engagements with larger populations of shared followers. The *referencing* and *hyperlink* dimensions found in this study may therefore serve as key contextual measures of social capital among cannabis markets in California. Indeed, several large dispensaries were able to occupy their own network spaces outside of the center cores in both cities through increased *referencing* and *hyperlinks*, which may help attract shared followers with regional preferences, motivations, and norms related to cannabis consumption [46]. As several California studies have found that dispensaries are more likely to cluster in communities with higher levels of cannabis demand, consumption, and morbidity [47-50], follow-up analyses that incorporate geospatial data will be better suited to determine how network position corresponds to the geographic distribution of dispensary communities in California and other states [51].

With regard to community formation among dispensaries, we conceptualized shared followers as a form of affinity that signals mutual interest and affiliation with dispensaries they choose to jointly follow. By incorporating this feature into a social network to understand interconnections between dispensaries, shared followers represent a potential resource that may flow between dispensaries and help form communities in response to unique patterns of cyberbehavior among dispensaries [52-54]. In the larger context of public health surveillance, the social networks constructed in SFBA and GLA may serve as the foundation for more rigorous studies to evaluate how new social policies and regulations disrupt or facilitate community formation and cyberbehavior. Moreover, the rapidly growing presence of dispensaries on the Internet suggests that the cyberbehaviors identified in this study may be useful measures to capture the frequency and types of communication that occur on Twitter [55,56]. Coordinated efforts to engage with researchers, policymakers, and stakeholders will be necessary to better understand the utility of these measures and develop scalable strategies to monitor large-scale industry practices on Twitter and other social media platforms [57,58].

Limitations

Although this study utilized shared follower potential to understand network structure and community formation among dispensaries, we acknowledge the multiple ways in which social

networks may be represented. Instead of shared follower networks, a strict flow network using shared or liked tweets among dispensaries may demonstrate different dynamics of social interaction and information propagation [59]. Indeed, exploratory analyses revealed very low levels of message diffusion among dispensaries in SFBA and GLA, which suggests that shared followers typically do not exchange or directly engage with dispensary tweets. Considering the referencing dimension, the content from dispensary tweets may also be constructed as a semantic network that not only illustrates conceptual connections between phrases, keywords, and hashtags, but also classifies how cannabis products are priced and promoted [60,61]. While such analyses were beyond the scope of this paper, rigorous content analyses will provide the framework with which to create a classification system that can be systematically trained to identify direct-to-consumer advertising of cannabis products and other specific types of tweets, such as health claims, industry events, scientific studies, and sentiment towards state and federal policies [61-63].

Similarly, the ability of the cyberbehavioral dimensions to distinguish between communities suggests that metadata can provide additional insights into dispensary and consumer behaviors on Twitter. Larger studies that leverage these dimensions with metadata like dispensary type (eg, nonprofit, delivery, and health services), provisions for state and local laws, and geospatial characteristics may improve the detection of dispensary communities [64]. For example, computational methods like stochastic block modeling can improve the accuracy of community detection with metadata without *a priori* assumptions about their correlations [65]. In other words, these

methods can learn (eg, unsupervised and semi-supervised) whether important correlations exist and subsequently use or ignore metadata depending on whether they provide useful information to network structure and community formation [65]. Finally, integrative techniques like exploratory graph analysis, latent network modeling, and residual network modeling represent new and exciting approaches that can help derive more parsimonious cyberbehavioral dimensions when compared with PCA and other latent variable modeling approaches [66,67].

Conclusions

The findings from this study indicate that network structure and multiple dimensions of dispensary behavior on Twitter shape two of California's largest cannabis markets. As California successfully passed Proposition 64 on the November 2016 ballot, the legalization of recreational cannabis use for adults aged 21 years and older further stresses the need to determine the policy implications of online cannabis marketing and monitor community activity through contextual measures of cyberbehavior that may influence population-based consumption. In addition, the emergence of online marketplaces and mobile apps demonstrates how the digitization of dispensaries has started to shift consumers away from storefronts to high-tech collaborative consumption platforms that personalize product choices and automate purchases. With Twitter as a key part of this digital paradigm shift, the scalable methodology used in this study will serve as the basis for more rigorous designs that longitudinally track community formation and patterns of cyberbehavior among dispensaries.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Descriptive statistics for cyberbehaviors among communities of dispensaries.

[PDF File (Adobe PDF File), 173KB-Multimedia Appendix 1]

Multimedia Appendix 2

Canonical loadings for Twitter cyberbehaviors among communities of dispensaries.

[PDF File (Adobe PDF File), 559KB-Multimedia Appendix 2]

References

1. Hasin DS, Saha TD, Kerridge BT, Goldstein RB, Chou SP, Zhang H, et al. Prevalence of marijuana use disorders in the United States between 2001-2002 and 2012-2013. JAMA Psychiatry 2015 Dec;72(12):1235-1242 [FREE Full text] [doi: 10.1001/jamapsychiatry.2015.1858] [Medline: 26502112]

2. Whiting PF, Wolff RF, Deshpande S, Di Nisio M, Duffy S, Hernandez AV, et al. Cannabinoids for medical use: a systematic review and meta-analysis. JAMA 2015 Jun;313(24):2456-2473. [doi: 10.1001/jama.2015.6358] [Medline: 26103030]

3. Janichek JL, Reiman A. Clinical service desires of medical cannabis patients. Harm Reduct J 2012 Mar 13;9:12 [FREE Full text] [doi: 10.1186/1477-7517-9-12] [Medline: 22414074]

4. Reinarman C, Nunberg H, Lanthier F, Heddleston T. Who are medical marijuana patients? Population characteristics from nine California assessment clinics. J Psychoactive Drugs 2011;43(2):128-135. [doi: 10.1080/02791072.2011.587700] [Medline: 21858958]

5. Schauer GL, King BA, Bunnell RE, Promoff G, McAfee TA. Toking, vaping, and eating for health or fun: marijuana use patterns in adults, U.S., 2014. Am J Prev Med 2016 Jan;50(1):1-8. [doi: 10.1016/j.amepre.2015.05.027] [Medline: 26277652]

6. Lau N, Sales P, Averill S, Murphy F, Sato SO, Murphy S. A safer alternative: Cannabis substitution as harm reduction. Drug Alcohol Rev 2015 Nov;34(6):654-659. [doi: 10.1111/dar.12275] [Medline: 25919477]

7. Pacula RL, Jacobson M, Maksabedian EJ. In the weeds: a baseline view of cannabis use among legalizing states and their neighbours. Addiction 2016 Jun;111(6):973-980. [doi: 10.1111/add.13282] [Medline: 26687431]

8. Bestrashniy J, Winters KC. Variability in medical marijuana laws in the United States. Psychol Addict Behav 2015 Sep;29(3):639-642. [doi: 10.1037/adb0000111] [Medline: 26415061]

9. McLaren J, Swift W, Dillon P, Allsop S. Cannabis potency and contamination: a review of the literature. Addiction 2008 Jul;103(7):1100-1109. [doi: 10.1111/j.1360-0443.2008.02230.x] [Medline: 18494838]

10. Goldsmith RS, Targino MC, Fanciullo GJ, Martin DW, Hartenbaum NP, White JM, et al. Medical marijuana in the workplace: challenges and management options for occupational physicians. J Occup Environ Med 2015 May;57(5):518-525 [FREE Full text] [doi: 10.1097/JOM.0000000000000454] [Medline: 25951421]

11. Kepple NJ, Mulholland E, Freisthler B, Schaper E. Correlates of Amount Spent on Marijuana Buds During a Discrete Purchase at Medical Marijuana Dispensaries: Results from a Pilot Study. J Psychoactive Drugs 2016 Jan;48(1):50-55. [doi: 10.1080/02791072.2015.1116719] [Medline: 26757234]

12. Novak SP, Peiper NC, Wiley J. Linking animal models to human self-administration practices among medical cannabis patients: a daily diary study. Drug Alcohol Depend 2017 Feb;171:e153-e154. [doi: 10.1016/j.drugalcdep.2016.08.424]

13. Cavazos-Rehg PA, Krauss M, Fisher SL, Salyer P, Grucza RA, Bierut LJ. Twitter chatter about marijuana. J Adolesc Health 2015 Feb;56(2):139-145 [FREE Full text] [doi: 10.1016/j.jadohealth.2014.10.270] [Medline: 25620299]

14. Cavazos-Rehg P, Krauss M, Grucza R, Bierut L. Characterizing the followers and tweets of a marijuana-focused Twitter handle. J Med Internet Res 2014 Jun;16(6):e157 [FREE Full text] [doi: 10.2196/jmir.3247] [Medline: 24974893]

15. Daniulaityte R, Nahhas RW, Wijeratne S, Carlson RG, Lamy FR, Martins SS, et al. "Time for dabs": analyzing Twitter data on marijuana concentrates across the U.S. Drug Alcohol Depend 2015 Oct 01;155:307-311 [FREE Full text] [doi: 10.1016/j.drugalcdep.2015.07.1199] [Medline: 26338481]

16. Lamy FR, Daniulaityte R, Sheth A, Nahhas RW, Martins SS, Boyer EW, et al. "Those edibles hit hard": Exploration of Twitter data on cannabis edibles in the U.S. Drug Alcohol Depend 2016 Jul;164(1):64-70. [doi: 10.1016/j.drugalcdep.2016.04.029] [Medline: 27185160]

17. Clark EM, Jones CA, Williams JR, Kurti AN, Norotsky MC, Danforth CM, et al. Vaporous marketing: uncovering pervasive electronic cigarette advertisements on Twitter. PLoS One 2016 Jul;11(7):e0157304 [FREE Full text] [doi: 10.1371/journal.pone.0157304] [Medline: 27410031]

18. Liang BA, Mackey TK. Prevalence and global health implications of social media in direct-to-consumer drug advertising. J Med Internet Res 2011 Aug;13(3):e64 [FREE Full text] [doi: 10.2196/jmir.1775] [Medline: 21880574]

19. Mackey TK, Cuomo RE, Liang BA. The rise of digital direct-to-consumer advertising?: comparison of direct-to-consumer advertising expenditure trends from publicly available data sources and global policy implications. BMC Health Serv Res 2015 Jun;15:236 [FREE Full text] [doi: 10.1186/s12913-015-0885-1] [Medline: 26084705]

20. Watters JK, Biernacki P. Targeted sampling: options for the study of hidden populations. Soc Probl 1989 Oct;36(4):416-430. [doi: 10.2307/800824]

21. Miller PG, Sønderlund AL. Using the Internet to research hidden populations of illicit drug users: a review. Addiction 2010 Sep;105(9):1557-1567. [doi: 10.1111/j.1360-0443.2010.02992.x] [Medline: 20626378]

22. Kral AH, Malekinejad M, Vaudrey J, Martinez AN, Lorvick J, McFarland W, et al. Comparing respondent-driven sampling and targeted sampling methods of recruiting injection drug users in San Francisco. J Urban Health 2010 Sep;87(5):839-850. [doi: 10.1007/s11524-010-9486-910.1007/s11524-010-9486-9] [Medline: 20582573]

23. Kral AH, Wenger L, Carpenter L, Wood E, Kerr T, Bourgois P. Acceptability of a safer injection facility among injection drug users in San Francisco. Drug Alcohol Depend 2010 Jul 01;110(1-2):160-163 [FREE Full text] [doi: 10.1016/j.drugalcdep.2010.02.009] [Medline: 20303679]

24. Kral AH, Wenger L, Novak SP, Chu D, Corsi KF, Coffa D, et al. Is cannabis use associated with less opioid use among people who inject drugs? Drug Alcohol Depend 2015 Aug 01;153:236-241 [FREE Full text] [doi: 10.1016/j.drugalcdep.2015.05.014] [Medline: 26051162]

25. Bowes L, Joinson C, Wolke D, Lewis G. Peer victimisation during adolescence and its impact on depression in early adulthood: prospective cohort study in the United Kingdom. BMJ 2015 Jun 02;350:h2469. [doi: 10.1136/bmj.h2469]

26. Stopka TJ, Lutnick A, Wenger LD, Deriemer K, Geraghty EM, Kral AH. Demographic, risk, and spatial factors associated with over-the-counter syringe purchase among injection drug users. Am J Epidemiol 2012 Jul 01;176(1):14-23 [FREE Full text] [doi: 10.1093/aje/kwr492] [Medline: 22562660]

27. Shapiro BJ, Lynch KL, Toochinda T, Lutnick A, Cheng HY, Kral AH. Promethazine misuse among methadone maintenance patients and community-based injection drug users. J Addict Med 2013;7(2):96-101 [FREE Full text] [doi: 10.1097/ADM.0b013e31827f9b43] [Medline: 23385449]

28. Novak SP, Håkansson A, Martinez-Raga J, Reimer J, Krotki K, Varughese S. Nonmedical use of prescription drugs in the European Union. BMC Psychiatry 2016 Aug 04;16:274 [FREE Full text] [doi: 10.1186/s12888-016-0909-3] [Medline: 27488186]

29. Caldarelli G, Chessa A. Data Science and Complex Networks: Real Case Studies with Python. USA: Oxford University Press; 2016.

30. Real R, Vargas JM. The probabilistic basis of Jaccard's Index of similarity. Syst Biol 1996;45(3):380-385. [doi: 10.2307/2413572]

31. Hamers L, Hemeryck Y, Herweyers G, Janssen M, Keters H, Rousseau R, et al. Similarity measures in scientometric research: the Jaccard index versus Salton's cosine formula. Inf Process Manag 1989;25(3):315-318. [doi: 10.1016/0306-4573(89)90048-4]

32. Blondel VD, Guillaume JL, Lambiotte R, Lefebvre E. Fast unfolding of communities in large networks. J Stat Mech 2008 Oct;2008:P10008. [doi: 10.1088/1742-5468/2008/10/P10008]

33. Schott JR. Dimensionality reduction in quadratic discriminant analysis. Comput Stat Data Anal 1993 Aug;16(2):161-174. [doi: 10.1016/0167-9473(93)90111-6]

34. Hamari J, Sjöklint M, Ukkonen A. The sharing economy: why people participate in collaborative consumption. J Assoc Inf Sci Technol 2016 Sep;67(9):2047-2059. [doi: 10.1002/asi.23552]

35. Belk R. You are what you can access: sharing and collaborative consumption online. J Bus Res 2014 Aug;67(8):1595-1600. [doi: 10.1016/j.jbusres.2013.10.001]

36. Valenzuela S, Park N, Kee KF. Is there social capital in a social network site?: Facebook use and college students' life satisfaction, trust, and participation. J Comput Mediat Commun 2009 Jul;14(4):875-901. [doi: 10.1111/j.1083-6101.2009.01474.x]

37. Wakefield R, Wakefield K. Social media network behavior: a study of user passion and affect. J Strateg Inf Syst 2016 Jul;25(2):140-156. [doi: 10.1016/j.jsis.2016.04.001]

38. Panigrahy R, Najork M, Xie Y. How user behavior is related to social affinity. USA; 2012 Presented at: 5th ACM International Conference on Web Search and Data Mining (WSDM); February 2012; Seattle, WA p. 713-722 URL: http://dl.acm.org/citation.cfm?id=2124379 [doi: 10.1145/2124295.2124379]

39. Mackey TK, Liang BA. Global reach of direct-to-consumer advertising using social media for illicit online drug sales. J Med Internet Res 2013 May;15(5):e105 [FREE Full text] [doi: 10.2196/jmir.2610] [Medline: 23718965]

40. Friedman SR, Mateu-Gelabert P, Curtis R, Maslow C, Bolyard M, Sandoval M, et al. Social capital or networks, negotiations, and norms? A neighborhood case study. Am J Prev Med 2007 Jun;32(6 Suppl):S160-S170 [FREE Full text] [doi: 10.1016/j.amepre.2007.02.005] [Medline: 17543707]

41. Moore S, Bockenholt U, Daniel M, Frohlich K, Kestens Y, Richard L. Social capital and core network ties: a validation study of individual-level social capital measures and their association with extra- and intra-neighborhood ties, and self-rated health. Health Place 2011 Mar;17(2):536-544. [doi: 10.1016/j.healthplace.2010.12.010] [Medline: 21208822]

42. Gonzalez-Bailon S. Opening the black box of link formation: social factors underlying the structure of the web. Soc Networks 2009 Oct;31(4):271-280. [doi: 10.1016/j.socnet.2009.07.003]

43. Shumate M, Dewitt L. The North/South divide in NGO hyperlink networks. J Comput Mediat Commun 2008;13(2):405-428. [doi: 10.1111/j.1083-6101.2008.00402.x]

44. Pilny A, Shumate M. Hyperlinks as extensions of offline instrumental collective action. Inf Commun Soc 2011;15(2):260-286. [doi: 10.1080/1369118X.2011.606328]

45. Shumate M. The evolution of the HIV/AIDS NGO hyperlink network. J Comput Mediat Commun 2012;17(2):120-134. [doi: 10.1111/j.1083-6101.2011.01569.x]

46. Csermely P, London A, Wu LY, Uzzi B. Structure and dynamics of core/periphery networks. J Complex Netw 2013;1:93-123. [doi: 10.1093/comnet/cnt016]

47. Thomas C, Freisthler B. Examining the locations of medical marijuana dispensaries in Los Angeles. Drug Alcohol Rev 2016 May;35(3):334-337. [doi: 10.1111/dar.12325] [Medline: 26423794]

48. Freisthler B, Kepple NJ, Sims R, Martin SE. Evaluating medical marijuana dispensary policies: spatial methods for the study of environmentally-based interventions. Am J Community Psychol 2013 Mar;51(1-2):278-288 [FREE Full text] [doi: 10.1007/s10464-012-9542-6] [Medline: 22821130]

49. Mair C, Freisthler B, Ponicki WR, Gaidus A. The impacts of marijuana dispensary density and neighborhood ecology on marijuana abuse and dependence. Drug Alcohol Depend 2015 Sep;154:111-116. [doi: 10.1016/j.drugalcdep.2015.06.019]

50. Morrison C, Gruenewald PJ, Freisthler B, Ponicki WR, Remer LG. The economic geography of medical cannabis dispensaries in California. Int J Drug Policy 2014 May;25(3):508-515. [doi: 10.1016/j.drugpo.2013.12.009]

51. Beletsky L, Arredondo J, Werb D, Vera A, Abramovitz D, Amon JJ, et al. Utilization of Google enterprise tools to georeference survey data among hard-to-reach groups: strategic application in international settings. Int J Health Geogr 2016 Jul 28;15(1):24 [FREE Full text] [doi: 10.1186/s12942-016-0053-9] [Medline: 27468764]

[XSL•FO]

RenderX

52. Borgatti SP. Centrality and network flow. Soc Networks 2005 Jan;27(1):55-71. [doi: 10.1016/j.socnet.2004.11.008]

53. Smith M, Giraud-Carrier C, Purser N. Implicit affinity networks and social capital. Inf Technol and Management 2009 Jul;10(2-3):123-134. [doi: 10.1007/s10799-009-0057-2]

54. Contractor N. The emergence of multidimensional networks. J Comput Mediat Commun 2009 Apr;14(3):743-747. [doi: 10.1111/j.1083-6101.2009.01465.x]

55. DeAndrea DC, Vendemia MA. How affiliation disclosure and control over user-generated comments affects consumer health knowledge and behavior: a randomized controlled experiment of pharmaceutical direct-to-consumer advertising on social media. J Med Internet Res 2016 Jul 19;18(7):e189 [FREE Full text] [doi: 10.2196/jmir.5972] [Medline: 27435883]

56. Bierut T, Krauss MJ, Sowles SJ, Cavazos-Rehg PA. Exploring Marijuana advertising on Weedmaps, a popular online directory. Prev Sci 2017 Feb;18(2):183-192. [doi: 10.1007/s11121-016-0702-z] [Medline: 27534665]

57. Bietz MJ, Bloss CS, Calvert S, Godino JG, Gregory J, Claffey MP, et al. Opportunities and challenges in the use of personal health data for health research. J Am Med Inform Assoc 2016 Apr;23(e1):e42-e48. [doi: 10.1093/jamia/ocv118] [Medline: 26335984]

58. Spencer K, Sanders C, Whitley EA, Lund D, Kaye J, Dixon WG. Patient perspectives on sharing anonymized personal health data using a digital system for dynamic consent and research feedback: a qualitative study. J Med Internet Res 2016 Apr 15;18(4):e66 [FREE Full text] [doi: 10.2196/jmir.5011] [Medline: 27083521]

59. Mishori R, Singh LO, Levy B, Newport C. Mapping physician Twitter networks: describing how they work as a first step in understanding connectivity, information flow, and message diffusion. J Med Internet Res 2014 Apr 14;16(4):e107 [FREE Full text] [doi: 10.2196/jmir.3006] [Medline: 24733146]

60. Martin MK, Pfeffer J, Carley KM. Network text analysis of conceptual overlap in interviews, newspaper articles and keywords. Soc Netw Anal Min 2013 Dec;3(4):1165-1177. [doi: 10.1007/s13278-013-0129-5]

61. Tighe PJ, Goldsmith RC, Gravenstein M, Bernard HR, Fillingim RB. The painful tweet: text, sentiment, and community structure analyses of tweets pertaining to pain. J Med Internet Res 2015;17(4):e84 [FREE Full text] [doi: 10.2196/jmir.3769] [Medline: 25843553]

62. Hamad EO, Savundranayagam MY, Holmes JD, Kinsella EA, Johnson AM. Toward a mixed-methods research approach to content analysis in the digital age: the combined content-analysis model and its applications to health care Twitter feeds. J Med Internet Res 2016 Mar 08;18(3):e60 [FREE Full text] [doi: 10.2196/jmir.5391] [Medline: 26957477]

63. Ramo DE, Popova L, Grana R, Zhao S, Chavez K. Cannabis Mobile Apps: A Content Analysis. JMIR Mhealth Uhealth 2015 Aug;3(3):e81 [FREE Full text] [doi: 10.2196/mhealth.4405] [Medline: 26268634]

64. Cardillo A, Gómez-Gardeñes J, Zanin M, Romance M, Papo D, del Pozo F, et al. Emergence of network features from multiplexity. Sci Rep 2013;3:1344 [FREE Full text] [doi: 10.1038/srep01344] [Medline: 23446838]

65. Newman ME, Clauset A. Structure and inference in annotated networks. Nat Commun 2016 Jun 16;7:11863 [FREE Full text] [doi: 10.1038/ncomms11863] [Medline: 27306566]

66. Golino HF, Demetriou A. Estimating the dimensionality of intelligence like data using Exploratory Graph Analysis. Intelligence 2017 May;62:54-70 (forthcoming) [FREE Full text] [doi: 10.1016/j.intell.2017.02.007]

67. Golino HF, Epskamp S. Exploratory graph analysis: A new approach for estimating the number of dimensions in psychological research. PLoS One 2017 Jun;12(6):e0174035 [FREE Full text] [doi: 10.1371/journal.pone.0174035]

Abbreviations

GLA: Greater Los Angeles
PCA: Principal components factor analysis
QDA: Quadratic discriminant analysis
SFBA: San Francisco Bay Area