Deciphering Influence on Social Media: A Comparative Analysis of Influential Account Detection Metrics in the Context of Tobacco Promotion

Alex Kresovich,
Andrew H. Norris,
Chandler C. Carter,
Yoonsang Kim,
Ganna Kostygina,
Sherry L. Emery
NORC at the University of Chicago, USA

Abstract

Influencer marketing spending in the United States was expected to surpass $6 billion in 2023. This marketing tactic poses a public health threat, as research suggests it has been utilized to undercut decades of public health progress—such as gains made against tobacco use among adolescents. Public health and public opinion researchers need practical tools to capture influential accounts on social media. Utilizing X (formerly Twitter) little cigar and cigarillo (LCC) data, we compared seven influential account detection metrics to help clarify our understanding of the functions of existing metrics and the nature of social media discussion of tobacco products. Results indicate that existing influential account detection metrics are *non-harmonic* and *time-sensitive*, capturing distinctly different users and categorically different user types. Our results also reveal that these metrics capture distinctly different conversations among influential social media accounts. Our findings suggest that public health and public opinion researchers hoping to conduct analyses of influential social media accounts need to understand each metric's benefits and limitations and utilize more than one influential account detection metric to increase the likelihood of producing valid and reliable research.

Corresponding Author: Alex Kresovich, Social Data Collaboratory, NORC at the University of Chicago, Chicago, IL, 60603, USA., kresovich-alex@norc.org.

Declaration of Conflicting Interests
The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Supplemental Material
Supplemental material for this article is available online.

Over four billion people globally are social media users, and projections estimate there will be nearly six billion by 2027 (Statista, 2022). Currently, over 80% of the United States population uses social media (Triton Digital, 2021). As a result, marketers now use social media to reach target audiences and potential consumers. Spending on social media marketing, including paid advertising, owned accounts, earned media, and influencer marketing, was expected to exceed well over $50 billion in 2022 (Williamson, 2020). Recent data indicates that influencer marketing consistently outperformed other revenue-boosting tools on social media platforms, such as embedded advertisements and search engine optimization, regarding return on investment (SocialPubli, 2019). While brands often struggle to create engaging social media content, influencers specialize in capturing niche audiences, creating trends, and encouraging followers to buy the products they promote (Campbell & Farrell, 2020). The worldwide influencer marketing industry was expected to surpass $21 billion in 2023 (Geyser, 2023), with an estimated $6.16 billion spent on influencer marketing in the United States alone (Lin, 2023).

A social media influencer is defined as a person (user) who has "the power to affect the purchasing decisions of others because of his or her authority, knowledge, position, or relationship with his or her audience" (Geyser, 2022, para. 2). Social media influencers actively engage with distinct audience niches on social media platforms (Geyser, 2022), and influencers maintain active relationships and interactions with their followers, representing homogeneous consumer segments to marketers (Leung et al., 2022). These follower interactions allow influencers to leverage the network effects of social media platforms by indirectly reaching larger audiences via peer mediation (engagement or sharing of the influencer post), which creates a persuasive heuristic to pay attention to the content (Anspach, 2017; Bene, 2017).

Influencer Marketing in the Tobacco Space

Tobacco companies have been using influencer marketing to undercut the decades of public health progress against tobacco use among adolescents (Johnston et al., 2019), and influencer marketing was a core strategy in JUUL's meteoric rise in popularity in 2018 (Jackler et al., 2019; Kaplan, 2018). Commercial tobacco interests appear to mobilize on social media platforms (N. A. Silver et al., in press), and many tobacco companies hire young social media influencers to promote tobacco use as normative to an attractive or aspirational lifestyle (Czaplicki, Kostygina et al., 2020; Hejlová et al., 2019; J. Liu et al., 2020; Navarro et al., 2021). Often, these influencers have been discouraged by tobacco companies from disclosing that they were engaged in paid advertising (Campaign for Tobacco Free Kids, 2018), for which the Federal Trade Commission requires #ad or #sponsored labels to ensure youth recognize the content as an advertisement (Klein et al., 2020).

A growing body of research suggests that the exponential growth of JUUL e-cigarette-related influencer content across social media platforms was significantly correlated with the exponential uptake of JUUL product use among U.S. youth (Chu et al., 2018; Czaplicki, Kostygina, et al., 2020; Fadus et al., 2019; Huang et al., 2019; Jackler et al., 2019; Johnston et al., 2019). In 2018, JUUL undertook several voluntary actions putatively aimed at reducing JUUL-related content on social media, including eliminating its own social media accounts and monitoring and removing "inappropriate material from third-party accounts" (Scutti, 2018). Nonetheless, these voluntary measures affected no meaningful change in the amount or content of JUUL-related content on Instagram (Czaplicki, Tulsiani, et al., 2020). In 2019, over 125 public health and other organizations from 48 countries collectively called upon social media companies to prohibit the promotion of tobacco products on their platforms, including promotions by influencers (Campaign for Tobacco Free Kids, 2019). In response, Facebook and Instagram announced they would no longer allow content advertising vaping and tobacco products (Azad, 2019). Most social media platforms banned paid tobacco advertising; however, few platforms have policies that address novel strategies such as sponsored and influencer posts—content where companies collaborate with influencers to endorse products to their followers in exchange for compensation or other incentives—and few require age-gating to prevent youth exposure to tobacco promotional content (Kong et al., 2022). While the Food and Drug Administration (FDA) now requires marketers of newer products, like IQOS, to report using partners, bloggers, brand ambassadors, and influencers to promote their products (U.S. Food and Drug Administration [FDA], Center for Tobacco Products, 2020), the proliferation of international influencer accounts which challenges enforcing this policy by making it more difficult for regulatory authorities to monitor and ensure compliance when influencers are based in different countries with varying regulations and standards.

The amount tobacco companies spend on influencer marketing remains unknown. However, in 2021, the Bureau of Investigative Journalism in the United Kingdom revealed that British American Tobacco—the second-largest tobacco company in the world (Kolmar, 2023)—had recently launched a £1 billion social media influencer campaign to harness influencers on TikTok, Instagram, and Facebook to promote their tobacco products—mainly to youth—in countries including Kenya, Pakistan, Sweden, and Spain (Chapman, 2022).

Operationalizing Influential Account Detection for Tobacco Control Research

To date, tobacco social media researchers have prioritized conducting research to identify social media influencers promoting tobacco use (Gu et al., 2022; Hejlová et al., 2019; Kostygina et al., 2016; Navarro et al., 2021; Vassey et al., 2023; Wu et al., 2022), assess the reach and engagement of influencers in the tobacco space (Gu et al., 2022; Kostygina et al., 2020; Kostygina et al., 2016; Vassey et al., 2023; Wu et al., 2022), characterize the content of pro- and anti-tobacco influencer posts (Gu et al., 2022; Hejlová et al., 2019; Kostygina et al., 2016; Navarro et al., 2021; Ramadhan & Bigwanto, 2022; Vassey et al., 2023; Wu et al., 2022), describe audience engagement with influencer posts about tobacco (Gu et al., 2022; Hejlová et al., 2019; Kostygina et al., 2016, 2020; Vassey et al., 2023), explore the efficacy

of potential countermeasures to pro-tobacco influencer posts (Klein et al., 2020, 2022; Vogel et al., 2020), and measure influencers' influence on the tobacco attitudes and behaviors (Ramadhan & Bigwanto, 2022; Vogel et al., 2020). Researchers have used three broad types of influencer metrics in tobacco work: *simple descriptive metrics* (i.e., follower count) (Blakemore et al., 2020; Bonnevie et al., 2020, 2021; Borgmann et al., 2016; Drescher et al., 2021; Edgerton et al., 2016; Gough et al., 2017; Harding et al., 2019; MacKay et al., 2022; Patel et al., 2022; Zou et al., 2021), *network metrics* (i.e., degree centrality) (Jain & Sinha, 2020; I. Kim & Valente, 2021; Lutkenhaus et al., 2019; Moukarzel et al., 2020, 2021; Munoz-Acuna et al., 2021; Zheng et al., 2021), and *index metrics* (Gallagher et al., 2021).

Simple descriptive metrics serve as a "quick and dirty" way to identify influential accounts on social media. These metrics often exist in the metadata and do not require advanced analytical methods to interpret. Simple descriptive metrics can take form as follower influence (a user's number of followers, indicating audience size), mention influence (the number of @user mentions, indicating the user's ability to engage or incite others into conversation, not including mentions in retweets to avoid conflating scores for mention influence and repost influence), and repost influence (the number of times a user's message is reposted, indicating the ability of the user to create shareable content). Emerging research suggests diverging effects between popularity metrics (i.e., follower counts) and marketing effectiveness. For example, micro-influencers (10k–100k followers) are often considered more trustworthy and authentic by audiences than mega-influencers (> 1 million followers) (Enberg, 2018; Main, 2017). Another significant weakness of simple descriptive metrics is their vulnerability to validity issues resulting from fraudulent or misleading follower and engagement numbers (i.e., influencers paying for followers, likes, or shares) (Zote, 2019).

Social media researchers use network metrics to capture sources and flow of information through social networks. Network metrics allow interpretation of the users' distinct roles in media conversations. I. Kim and Valente argue that there are three types of influential users on X (formerly Twitter) represented by conventional network metrics: information sources (represented by in-degree centrality), information disseminators (out-degree centrality), and information brokers (betweenness centrality) (I. Kim & Valente, 2021). Thus, network metrics inherently represent varying functions of influential accounts on social media.

Index metrics offer researchers a tool to identify influential social media accounts by detecting sustained amplification of messages or posts from a specific user around a particular topic. Public health scholars have created indexes to operationalize social media influence (Gallagher et al., 2021; Jain & Sinha, 2020). However, just as often, researchers conducting influencer studies lean on third-party tools designed by social media analytic companies for the same purpose (Albalawi & Sixsmith, 2017; Edgerton et al., 2016). Unfortunately, there is a lack of scholarly consensus over which third-party index metric best detects influential accounts on social media, best illustrated by the variation in metrics utilized in the extant literature (Albalawi & Sixsmith, 2017; Edgerton et al., 2016). Further, third-party tools threaten the replicability and transparency of influencer research as the algorithms utilized by these black-box proprietary tools to identify accounts are not shared with researchers (Albalawi & Sixsmith, 2017; Edgerton et al., 2016; Gallagher et al., 2021; Jain & Sinha, 2020).

Researcher-created index metrics, such as T factor (Bornmann and Haunschild, 2015) and *h*-index (Gallagher et al., 2021), are more transparent influential account detection metrics. The *h*-index was initially created to measure academic citations (Hirsch, 2005) and has since been adapted by Gallagher et al. to identify influential accounts, or "elites," who receive sustained amplification in a specific conversation (i.e., COVID-19) on social media (Hirsch, 2005). Retweets are used to calculate the *h*-index rather than likes or favorites because the original *h*-index metric was designed to assess the impact and influence of scholarly publications, typically measured through the number of citations received. A higher *h*-index score represents the consistent production of many posts *and* notable attention paid to each post by audiences. The emphasis on consistency also minimizes any statistical noise created by spam or outlier viral posts.

The Need to Assess Available Tools to Detect Influential Accounts on Social Media

Public health and public opinion researchers need practical tools to capture influential accounts discussing tobacco on social media and promoting tobacco use among impressionable youth and young adult audiences (Brien et al., 2020; Gu et al., 2022; Hejlová et al., 2019; Hickman, 2021; Navarro et al., 2021; Nedelman & Selig, 2018). However, before we can efficiently and reliably capture influential tobacco accounts on social media, we need to understand what approaches are available for detecting influential accounts on social media. Therefore, this study represents the vital first step in evaluating the existing methodological tools for detecting influential social media accounts. Our evaluation uses tobacco marketing content as a use-case but is generalizable across marketing, social marketing, and public opinion research domains.

As the extant social media influencer literature regularly uses "influencer" as a catch-all term for distinctly different online personas, including celebrities and customer product reviewers(Leung et al., 2022), we aim to help clarify the difference between describing a user account as an "influencer" and being able to detect influential accounts on social media. For example, an influential account on social media vested in manipulating public opinion on tobacco policy is qualitatively different from an entertainer on social media doing a viral dance holding a JUUL e-cigarette product to promote the product to their audience.

Agenda-setting theory (McCombs et al., 2013) posits that the media can shape public opinion—and, by extension, behavior—by influencing what topics get the most attention (McCombs et al., 2013). Using this agenda-setting lens and guided by the work of I. Kim and Valente (2021), we predict that the differing influential account detection metrics will detect influential accounts strategically guiding or participating in differing conversations (i.e., discussing tobacco policy versus promoting tobacco products) within their respective communication spheres. In other words, we predict that the differing analytical metrics (*simple, network*, and *index metrics*) will capture distinctly different influential tobacco accounts on X. The extant literature has established that influential accounts on X serve as sources, disseminators, and brokers of information, helping to guide different concurrent conversations (I. Kim & Valente, 2021). Therefore, we predict that

our analyses will help to reveal different types of "influencers" that serve as influential accounts in respective communication spheres (e.g., LCC policy, LCC lifestyle, and LCC product promotion discussions). We also predict that the differing influential account metrics will detect influential accounts strategically guiding or participating in differing conversations (i.e., discussing tobacco policy promoting tobacco products) within their respective communication spheres.

This research will help explicate the existing influential account detection metrics using real-world data focusing on a public health issue and, in doing so, clarify our understanding and highlight the many forms that influential accounts can take, such as those in entertainment (i.e., celebrities and social media stars) (Rosenstein, 2017), politics (i.e., commentators and politicians) (Coulter, 2020), or news media (i.e., journalists and news media company accounts) (CNN, 2022). These findings will also help social media researchers delineate which influential account detection metrics are best suited to address the aspects of social media influence they hope to address in their research—including, but not limited to, detecting likely undisclosed tobacco influencers on social media platforms. Given the exponential growth in spending on tobacco influencer marketing (Chapman, 2022; Kaplan, 2018), tobacco control researchers must integrate the extant operational definitions and delineate the differences between "being an influential account on social media" and "being an influencer" to address the conceptual overlap and to advance our understanding of social media influence. As the current social media influencer literature regularly uses "influencer" as a catch-all term for distinctly different online personas, including celebrities and customer product reviewers (Leung et al., 2022), we aim to help clarify the difference between describing a user account as an "influencer" and being able to detect influential accounts on social media.

Therefore, our study aims to address the following hypothesis and research questions:

> *Research Question 1 (RQ₁).* Do *simple descriptive*, *network*, and *index* metrics capture distinctly different types of influential accounts on X (formerly Twitter)?

> *Research Question 2 (RQ₂).* How dynamic are the influential account detection metrics within the data time frame every week?

> *Research Question 3 (RQ₃).* Which categories of influential accounts are most prominent in shaping the discourse in the LCC X space?

> *Research Question 4 (RQ₄).* How do content features (use of hashtags, hyperlinks, n-grams) in the social media messages (tweets) posted by accounts captured differ across the seven influential account detection metrics?

Method

Data Collection

To conduct these analyses, we first identified X (formerly Twitter) posts (original tweets and retweets) relevant to LCC products and used posts from the 3 weeks following the FDA's menthol ban announcement (April 29, 2021, to May 21, 2021) (U.S. FDA, 2022). We selected the FDA menthol ban announcement as a time period for data collection

because it was the most recent tobacco control milestone, which would generate public discussion at the time of data collection (American Lung Association, 2023). X was ideal for this influencer detection research because other social media platforms limit researchers' ability to extract user-level data (i.e., user descriptions). However, we plan to adapt this methodological evaluation design to other popular platforms (i.e., Instagram and TikTok) in future research. We are focusing on LCC posts because previous tobacco research on X has identified prominent celebrities and community accounts in LCC promotion (Kostygina et al., 2016, 2017). Further, we are focusing on LCC posts because research indicates that cigar product flavor bans may reduce tobacco product co-use among young adults (Glasser et al., 2023)—one of the target populations for big tobacco on social media.

We collected 647,465 Tweets from April 29, 2021, through May 21, 2021, that matched any of our LCC search rules via X's Historical PowerTrack (Twitter, 2023). Our LCC search results and search terms—including terms (e.g., cigar and cigarillo), brands (e.g., Black & Mild and ZigZag), and behaviors (e.g., "smoke a stogie" and "need a blunt") are available upon request. Historical PowerTrack allowed access to full public tweets that matched the search rules. After we detected relevant posts, utilizing the approach detailed below, we used post metadata to remove the duplicate tweets and then prepared the analytic dataset for qualitative human coding, quantitative summary statistics, and more.

As with any search rules, tweets irrelevant to LCC were collected (Y. Kim et al., 2016). For example, Phillies Cigars, a brand of traditional cigar in the United States, was captured via our keywords; however, the Phillies were also a prominent baseball team. We developed a supervised machine classifier to identify LCC-relevant posts and exclude non-relevant posts using the team's standard procedures (Czaplicki, Kostygina, et al., 2020; Kostygina et al., 2016, 2020). The development of the classifier consisted of training a model on a human-coded dataset extracted using LCC keywords. Two coders assessed a randomly selected subset of posts for their relevance to LCC, considering both the visual and linguistic aspects of each post. The sampling process was stratified by the week of posting to ensure that the training dataset encompassed the entire duration of the study. The final classifier was the result of stochastic gradient descent learning with modified Huber loss function and elasticnet regularization. This process aimed to ensure the inclusion of only tweets directly relevant to the topic of interest, in this case, little cigars and cigarillos. We removed 287,853 posts (42.7% of all "raw data" posts collected) for being classified as non-relevant, leaving us with a final analytical sample of 386,612 X posts. The classifier achieved an F1 score of 0.93 (precision 0.91, recall 0.94) from five-fold cross-validation on the training set; and an F1 score of 0.92 (precision 0.89, recall 0.95) on the held-out test set, indicating a high-performance classifier that would provide reliable predictions (Y. Kim et al., 2016).

We also considered the presence of bot-generated content. While our primary focus was on classifying and removing non-relevant posts to LCC discussions, it is worth noting that robotic accounts and automated messages could play a significant role in promoting tobacco products and related policies on social media (X. Liu, 2019). These bots contributed to the information flow in this domain, making it important to distinguish their presence. Therefore, we utilized SGBot (Yang et al., 2020, 2022) to differentiate likely X bot accounts

from non-bot accounts and to report the proportion of likely bot-generated tweets by each detection method.

Influential Account Detection Metric Construction

The *simple descriptive* metrics (follower influence, mention influence, and repost influence) were constructed by aggregating totals for each measure across the selected period. These metrics were created using standard X (formerly Twitter) metadata.

The *h-index* metric was constructed following Gallagher et al. (2021) using the Python package scholarmetrics. The *h*-index is a measure that identifies the largest number "h" such that there are "h" tweets, each of which has been retweeted at least "h" times. In this context, retweets are considered akin to citations in traditional scholarly impact metrics.

Our network measures were based on a retweet network, which allows for direct comparisons with the other retweet-based metrics (*h*-index, simple descriptive metrics). The retweet network structure created a directed network with connections (i.e., network edges) made when one user retweets another user's LCC-relevant post. The retweet network dataset was generated using the Python networkx package (NetworkX, 2023), with each row containing the user retweeting, the user being retweeted, and a timestamp/identifier of the occurrence. We used this data to calculate three network centrality metrics: in-degree (information sources), out-degree (information disseminators), and betweenness centrality (information brokers).

After constructing the seven influential account detection metrics, we ranked the accounts by each metric and identified the top 100 as potential influential accounts for each metric ($n = 700$). After removing the duplicate influential accounts which appeared across more than one metric, the final sample of influential tobacco accounts was coded ($n = 565$).

Influential Account Category Coding

As guided by the real-time infoveillance of Twitter health messages framework (Colditz et al., 2018), we utilized three X (formerly Twitter) coders with extensive experience analyzing tobacco conversations on X to code influential account types. Coders reviewed influential accounts using basic textual information from their posts and user metadata. The coders also reviewed the influential accounts on the X website as needed.

The three coders categorized each influential account identified from the seven influential account detection metrics into 1 of 15 mutually exclusive categories describing the type of influential account. We used a grounded theory (Strauss & Corbin, 1997) approach to create and refine these 15 account-type categories. The three coders first determined if each tobacco-related influential account represented an Individual or Non-Individual/Group/Organization to aid the subsequent category coding. Each influential account was then coded as one of the categories in Table 1. It is also worth noting that the X verification process changed since these data were collected, which we considered when categorizing users and interpreting the results. Our findings regarding *verified* and *unverified* accounts reflect the verification status before X changed verification to a paid service on November 9, 2022. Our coding manual can be found in Supplemental Online Table 1.

A random 16% of the total sample (88 of 565 accounts) was coded independently by each coder to test interrater reliability (Riff et al., 2014). Reliability analyses indicated a high level of agreement between coders for the influential account category analysis, with Gwet's AC1 (Gwet, 2008) exceeding 0.90 across all coding categories. All three coders coded all accounts in the random sample, and disagreements were resolved by group discussion.

Message Content Feature Analysis

Two coders reviewed LCC-relevant tweets made by influential accounts to code content themes. We performed Natural Language Processing (Chowdhary, 2020) to identify the most-occurring features related to a thematic analysis by reviewing top hyperlinks, generated user-level descriptives (e.g., follower, friend, and post counts), engagement statistics, and n-grams. N-grams were used for coding salient themes. N-grams connected sequences of *n* items (words) from a given sample of text or speech (Broder et al., 1997). We examined top unigrams, bigrams, and trigrams and found that the trigrams were the most interpretable for salient content themes, consistent with previous social media tobacco research (Benson et al., 2020; Czaplicki, Kostygina, et al., 2020; Malik et al., 2021; N. Silver et al., 2022). Salient themes for each of the seven influential account detection metrics were identified using a grounded theory (Strauss & Corbin, 1997) approach from trigram analysis and coded by consensus agreement between two coders.

Statistical Analyses

We evaluated RQ1 by utilizing Jaccard similarity coefficients to compare differences between the top 100 accounts captured by the seven different metrics across the 3 weeks. We evaluated RQ2 by utilizing Jaccard similarity to compare the within-metric differences/ changes of top accounts across the 3 weeks by comparing pairs of adjacent weeks (Week 1 and Week 2 vs. Week 2 and Week 3) to evaluate how dynamic each metric was during the three weeks. RQ3 was evaluated by comparing which influential account category type (e.g., entertainer, political commentator, media network, etc.) appeared most often across the seven detection metrics and, in total, across the study period. RQ4 was evaluated by cross-examination of results from both unsupervised (i.e., LDA) and traditional (e.g., n-grams and top tweet features) thematic analysis methods performed on the data.

Ethical Considerations

Human subjects were not involved in this study. Our inclusion of public X messages and metadata (including usernames) in our analyses and publication follows X's Terms of Service.

Results

RQ1 explored if simple descriptive, network, and index metrics captured distinctly different types of influential accounts on X. The highest similarity coefficient was between the retweet and out-degree centrality ("disseminator") metrics (Jaccard similarity = 31; *n* = 47 mutual accounts captured out of a possible 100). The out-degree ("disseminator") and betweenness ("broker") centrality metrics (Jaccard similarity = 13; *n* = 23 mutual accounts captured out of a possible 100) were the only other metrics with a similarity coefficient

above .10, indicating that there was minimal overlap between the metrics. There was no overlap in accounts captured between in-degree and retweets, out-degree and followers, and out-degree and retweets. Further, no single influential account was captured by all seven metrics. As seen in Table 2, our results suggest that the extant influential account metrics capture distinctly different accounts.

RQ2 explored how dynamic the influential account detection metrics were across the 3 weeks following the FDA's announcement of the menthol ban. Within each metric, we assessed the week-by-week dynamics of accounts captured by each metric by comparing the Jaccard similarity between their top 100 accounts over the 3 weeks (comparing Weeks 1 and 2 vs. Weeks 2 and 3). As seen in Table 3, our findings indicated that Jaccard similarity increased slightly for all metrics except the mention metric. The retweet metric had Jaccard similarity coefficients near 40 across both weeks, indicating less turnover in the top 100 accounts, compared to the other metrics of influential accounts captured across this period.

By contrast, the in-degree centrality ("source") and betweenness centrality ("broker") metrics were very dynamic, with Jaccard coefficients across this period generally below 10, indicating a lot of turnover in the accounts that each metric was capturing. In contrast, our findings suggest that retweets and mentions changed the least every week. At the same time, there was more turnover among the metrics capturing accounts initiating connections —in-degree centrality ("sources") and betweenness centrality ("brokers")—changed the least.

RQ3 explored which categories of influential accounts were most prominent in shaping the discourse in the LCC X space in the three weeks following the FDA menthol ban announcement. Table 4 shows the top 5 account categories by influential account detection metric. As seen in Table 4, across all accounts captured across the index, network, and simple metrics, *unverified* Other (organic/unclear) (19.6%), *unverified* NSFW (17.9%), and *unverified* LCC Users were the most prominent account-type categories. Among all accounts analyzed, 21.8% were deleted, suspended, or inaccessible, and 15.8% of the captured accounts did not post any LCC content during the study period and were thus coded as non-relevant. NSFW accounts were most likely to be classified as bot accounts. About one-quarter (24.6%) of the NSFW accounts were flagged as likely bot accounts, *unverified* X accounts (using X's previous verification system) with an SGBot probability score above .80. NSFW was in the top 3 categories for five metrics (*h*-index, in-degree [sources], out-degree [disseminators], mentions, and retweets), ranging from 14.5% of accounts to 27.1%. Other accounts appeared in the top 3 account types for four metrics (*h*-index, in-degree [sources], out-degree [disseminators]), and LCC community/user accounts tended to be more prominent in the simple metric samples (mentions = 40.2%; retweets = 23.3%) than in the network metric samples (out-degree [disseminators] = 9.4%; betweenness [brokers] = 6.2%).

Our findings indicate that the follower metric—the most used metric in most social media research—captured distinctly different influential account types than all other metrics. Most of the accounts that the follower metric captured appeared to be brands or organizations (73.5%) rather than individuals (26.8%), while total relevant accounts across all metrics

were primarily individuals (80.0%). The follower metric was also the only metric for which *verified* News Media accounts were among the top 5 account types. These accounts were the majority captured by the metric (51.8%), while *verified* News Media accounts do not appear in the top 5 account types for any other metric (0% to 5.1% for all other metrics). The second most prominent category for the follower metric is *verified* Entertainment accounts (39.3%). While *verified* Entertainment accounts are the second most prominent account type for the follower metric (39.3%), *unverified* Entertainment accounts are more prominent among the remaining metrics, particularly network and index measures (11.9% to 26.6%). *Unverified* Entertainment accounts appear in the top 5 categories for all metrics except mentions and followers, indicating that only network or index metrics were detecting them efficiently. While all other metrics captured NSFW and Other (organic/unclear) accounts in their top 5 categories, the follower metric did not capture any NSFW or Other accounts. The follower metric also had the highest proportion of accounts (44%, $n = 44$) coded as "non-relevant," while no other metric had more than 18% coded as non-relevant.

RQ4 explored how content features (use of hashtags, hyperlinks, n-grams) in the tweets differ across the seven influential account detection metrics. We used an inductive approach to analyze trigrams based on their content. The tri gram themes that emerged were casual blunt use, cigar promotion, menthol and flavor ban, and NSFW (adult or pornographic) content. The casual blunt use trigram theme indicated that a given metric mainly captured posts of users supporting other users' casual use of popular blunt products (e.g., *Black and Mild*) and blunt preparation methods (e.g., "rolling blunts"). The cigar promotion tri gram theme indicated that a metric mainly captured commercial marketing and advertising posts promoting cigar products on X, where accounts often shared the "latest cigar news" and promoted websites where users could buy cigars from extensive collections. The menthol and flavor ban tri gram theme indicated that a metric mainly captured posts about the FDA's announcement of menthol and flavored tobacco products. The NSFW tri gram theme indicated that a metric mainly captured posts containing NSFW content, many of which utilized cigar products and smoking as promotional aspects of the pornographic content.

Table 5 shows the content features in tweets posted by accounts captured across the seven influential account detection metrics. The follower metric captured the most accounts sharing hyperlinks (81% of tweets from the top 100 accounts provided a hyperlink). Results indicate some crossover across metrics in the salient themes that emerged based on the top trigrams. The follower and in-degree ("sources") metrics captured influential accounts discussing the FDA menthol ban ("ban menthol cigarettes" and "menthol cigarettes flavored," respectively), the retweet and *h*-index metrics captured influential accounts sharing NSFW content involving cigar culture ("join beautiful girls"), and the mentions and out-degree metrics ("disseminators") captured accounts promoting cigar and LCC products ("eat drink smoke" and "latest cigars daily," respectively). The betweenness ("brokers") metric captured accounts discussing casual blunt use ("smoke pairing enjoy"). The relevant accounts captured by the follower and in-degree ("sources") metrics used far fewer hashtags than the other metrics.

Discussion

The findings from this study have broad methodological implications for social media research. Our metric similarity analysis revealed that influential account detection metrics capture distinctly different influential accounts. The analyses of types of captured accounts further illustrated the substantial differences across each metric. While there was some congruence within the network metrics, these metrics were also the most dynamic (had the most turnover in top accounts captured) across the sampled period. Our findings suggested that accounts that typically retweet others' messages or consistently mention other users (often functions of bots) were consistent across the observed period. It is worth noting that the *h*-index metric is not designed for weekly analysis as it is intended to capture sustained amplification, so the observation of *h*-index change over time should be taken with caution.

Given that these metrics are all designed to capture and quantify influential accounts on social media, our findings reveal that influential accounts *take on different forms* such that social media research requires varying sensitivities when making methodological decisions. To use an earlier example, an influential account on social media vested in manipulating public opinion on tobacco policy is qualitatively different from an entertainer on social media doing a viral dance holding a JUUL e-cigarette product to promote the product to their audience. Thus, a vital takeaway from this study is that influential account detection metrics are *non-harmonic* and *time-sensitive.* The results of studies built around measures quantifying and characterizing influential accounts on social media may vary widely based on which influence detection metric the researchers select and what period they use for data collection.

Our results also revealed that these metrics capture distinctly different *conversations.* Our thematic content analysis of the content posted by the top captured accounts indicated a distinct difference in the types of conversations these influential accounts had. Researchers tapping into the LCC conversation after the FDA ban announcement using "sources" (in-degree centrality) and follower metrics would only be analyzing event-focused discussion during this sample period, while researchers utilizing mention or "disseminator" (out-degree centrality) metrics—designed to capture accounts making "outward" connections—would have mostly detected LCC marketing. During this same study period, researchers using retweet or *h*-index metrics—metrics designed to capture users whose content is being shared widely—would have discovered NSFW (pornographic or adult) LCC content. Combined with the account similarity and categorical coding findings, these results suggest that researchers studying influential social media accounts *need to utilize more than one influential account detection metric* to increase the likelihood of producing valid and reliable results. Further, to select the appropriate metric for analysis, researchers need to consider the *type of content* they are interested in (i.e., marketing vs. regular people vs. policy/advocacy) to make better-informed methodological decisions to meet their research goals.

Our findings also have significant implications for using follower count as a metric for detecting influential accounts going forward. While follower count has often been a default influencer detection measure (Blakemore et al., 2020; Bonnevie et al., 2020, 2021; Borgmann et al., 2016; Drescher et al., 2021; Edgerton et al., 2016; Gough et al., 2017;

Harding et al., 2019; MacKay et al., 2022; Patel et al., 2022; Zou et al., 2021), our account-type coding indicates that the Follower metric may be ill-fitted for rigorous influential account social media analyses. The follower metric mostly captured *verified* News Media accounts, and it was the only metric with this category type among its top 5 account types captured. Further, the follower count metric had the highest proportion of accounts coded as "non-relevant." Considered in concert with the growing evidence that follower count is not a valid representation of influence (Enberg, 2018; Main, 2017; Zote, 2019), we hope that our results further drive the nail into the coffin of the assertion that follower count can provide valid or reliable results on its own in research attempting to identify influential social media accounts.

We were surprised to observe that there was not much policy-centric discussion in the wake of the FDA menthol and flavor ban announcement among influential accounts discussing LCC products. While the ban was thematically represented by the top trigrams from the followers and "sources" (in-degree centrality) metrics, the general conversation around the ban was sparse. Although there was substantial similarity in the top hashtags captured by the different metrics, the top trigrams indicate that there was also substantial variation in the topics of the discussions. Our results suggest that distinctly different conversations were occurring in the messaging in the LCC tobacco space on X in the three weeks after the announcement of the FDA menthol ban, and researchers looking to explore the conversation on X occurring during this time would identify different conversations depending on which influential account detection metric they used. Unlike previous research exploring e-cigarette conversations on social media (Kostygina et al., 2021, 2022), we found zero users in the sample labeled by human coders as anti-tobacco regulation advocates among this sample. Indeed, we observed only two prominent conversations during this period—an LCC marketing conversation (e.g., promoting cigars) and an LCC behavior (norms) conversation (e.g., rolling blunts, sharing cigar suggestions and recommendations). This observation may be explained by the fact that tobacco-flavored (i.e., unflavored) products comprise most of the LCC product market (estimates indicate 50–80% market share) (Wang et al., 2021), so LCC users were less concerned with the ban than other tobacco-use product groups would be. Nonetheless, analyses exploring the prominence of differing tobacco conversations on social media are warranted.

The prominence of influential users coded as *unverified* Entertainment accounts in this sample is important. While the intention of this study was not to identify likely undisclosed tobacco influencers, we would like to "flag" this account type as one of interest to researchers. Social media users associate verification with "celebrity" rather than authenticity. Consumers are less likely to trust *verified* accounts than *unverified* ones (Dumas and Stough, 2022). Future research should not overlook *unverified* accounts when examining potential influencer accounts, as our results indicate that these accounts were among the most central accounts to the LCC discussion during our study period. It is important to note that in light of Elon Musk's changes to the X verification system, which allows users to pay for verification badges (Conger, 2023), ideas about trust on X are likely to shift. However, verification on other platforms that do *not* use the pay-for-badge system likely still confers meaning.

Acknowledging the prevalence of likely bot accounts sharing and discussing contentious subjects on social media is vital. Tobacco researchers have suggested that the tobacco industry may use bot accounts to create a climate of false consensus and circulate misinformation regarding tobacco products (Jun et al., 2022). Our results indicate that almost 1 in 5 accounts captured by influential account metrics were likely bots (SGBot score > .80). Is it likely not a coincidence that the metrics that captured the most likely bot accounts (> 30% of all accounts captured) were "disseminators" (out-degree centrality) and retweet metrics, metrics specifically designed to capture accounts that are most effectively sharing messages on social media. Further, NSFW (adult or pornographic) accounts had the highest proportion of likely bot scores. The growing prevalence of porn bots on social media threatens to further hamper social media data validity (Tiidenberg, 2021). Therefore, consistent with the findings of previous social media tobacco research (Allem et al., 2017; Allem & Ferrara, 2018; Ferrara, 2018), public health and public opinion researchers conducting social media research would be well advised to utilize bot detection software to supplement their detection strategy.

Some limitations of this study are worth noting. First, while our findings provide valuable insights into how different metrics identify influential accounts discussing tobacco, we acknowledge the potential biases inherent in the data collection process. Our data collection process could potentially introduce selection bias in the sample of accounts analyzed. We collected X data during a specific time frame, from April 29, 2021, to May 21, 2021, in the wake of the FDA's menthol ban announcement. While this timeframe was chosen to capture the most recent and policy-relevant discussions on LCC products and use, it is possible that the discourse during this period may not fully represent the broader, long-term discussions on X. Moreover, our data collection was limited to publicly available tweets, which may not account for the entire spectrum of conversations on LCC.

In addition, it is worth noting that our study relied on X (formerly Twitter) data, and X has changed drastically since Elon Musk acquired the company. His implemented changes may significantly affect public health research using this platform (Mole, 2022). For example, verification status was a meaningful variable in this research and was previously assigned by X to "notable" accounts, companies, and brands. However, the new X model allows users to pay for *verified* status. In other words, the platform post-Musk acquisition has changed drastically such that some of our observations and findings using pre-Musk acquisition data may not be relevant to the Twitter/X platform post-Musk acquisition. While problematic at first glance, this realization makes the findings from this work even more valuable as the observed differences across metrics stand up irrespective of the process for achieving *verified* status. Although our categorical analysis separating *verified* and *unverified* status within categories would look different, the results still suggest distinct differences in the types of accounts each influential metric captured. Nonetheless, future research is needed to develop helpful heuristics for identifying tobacco-sharing accounts of interest. The ongoing changes to Twitter X only add to the evidence that the shifting sands of social media platforms' terms of service, data-sharing policies, and operations will always present challenges to social media researchers.

Our mutually exclusive coding category design also presented challenges. Like other analyses of media (Guo & Li, 2016), we faced the challenge of interpreting influential users' intended meanings of their self-categorizations as part of our coding process. While we were able to establish a reliable coding process, the clarity and harmony gaine by using a mutually exclusive coding scheme potentially came at the expense of nuanced aspects of influential user accounts. For example, the line between entertainment and commercial accounts was sometimes blurred as accounts would offer services while intending to provide entertainment products (e.g., a self-described blogger selling unrelated products on their website). Future research could utilize a multi-faceted coding scheme to help tease out nuanced differences among influential accounts discussing tobacco on social media.

We hope that these findings constitute vital first steps in improving methodological decision-making for research analyzing influential accounts on social media. Our comparison of social media metrics provides four vital pieces of guidance for public health and public opinion researchers hoping to conduct valid and reliable analyses of influential accounts on social media. First, our findings highlight the importance of considering various metrics for capturing influential accounts. Researchers should take this into account and consider using a combination of these metrics to gain a more comprehensive view of the landscape of influential accounts in tobacco-related discussions on social media. Second, researchers should consider the temporal dimension of influential account detection metrics when analyzing influential accounts and be aware of which metrics are better suited for capturing stable or evolving accounts in their studies. Third, our findings indicate that different metrics capture different types of accounts, including *unverified* accounts, NSFW content, LCC users, and more. Understanding these categories can help researchers identify the key players in the tobacco-related discourse and tailor their research to focus on specific account types or content. Fourth, our analysis provides insights into the content features associated with influential accounts across different metrics. Public health and public opinion researchers can leverage this information to gain a deeper understanding of the topics and themes that resonate with influential accounts in the tobacco discussion. This knowledge can inform the creation of more targeted content and campaigns to ensure appropriate and more effective strategies for message design and placement among these influential accounts.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

Biographies

Alex Kresovich (Ph.D., University of North Carolina at Chapel Hill) is a Research Scientist in the Social Data Collaboratory (SDC) at NORC at the University of Chicago. His research interests include media effects of health messaging, and he is passionate about implementing behavior change theories in health communication interventions aimed at improving outcomes related to youth mental health, tobacco use, and opioid use.

Andrew H. Norris (M.S., New York University) is a Data Scientist at Brightside Health. His research interests include data science and utilizing big data to answer complex questions related to health.

Chandler C. Carter (M.A., University of North Carolina at Chapel Hill) is a public health researcher at NORC at the University of Chicago. Her research interests include youth tobacco use, tobacco control, and the impact of social media on body image and mental health.

Yoonsang Kim (Ph.D., University of Illinois at Chicago) is a Principal Data Scientist with NORC at the University of Chicago. She oversees study design, statistical analysis, machine learning, data harmonization, and other data science practices. At NORC's Social Data Collaboratory, she serves as a lead biostatistician for social media research to examine the effects of exposure to social media on health behaviors, focusing on data quality assessment and the development of social media-derived measures.

Ganna Kostygina (Ph.D., University of Southern California) is a Principal Research Scientist at the Social Data Collaboratory, NORC, at the University of Chicago. Her research agenda centers on advancing regulatory and communication science and technology for tobacco control and health promotion. Specifically, she has conducted research on topics related to tobacco and substance use prevention, as well as tobacco product marketing and counter-marketing on traditional and digital media channels.

Sherry L. Emery (Ph.D., University of North Carolina at Chapel Hill) is the Director of the Social Data Collaboratory (SDC) and a Senior Fellow at NORC at the University of Chicago. Her interdisciplinary research applies the approaches of health communication, data science, and public policy to understand how both traditional and new media influence health behavior.

References

Albalawi Y, & Sixsmith J (2017). Identifying Twitter influencer profiles for health promotion in Saudi Arabia. Health Promotion International, 32(3), 456–463. 10.1093/heapro/dav103 [PubMed: 26516181]

Allem J-P, & Ferrara E (2018). Could social bots pose a threat to public health? American Journal of Public Health, 108(8), 1005–1006. [PubMed: 29995482]

Allem J-P, Ferrara E, Uppu SP, Cruz TB, & Unger JB (2017). E-cigarette surveillance with social media data: Social bots, emerging topics, and trends. JMIR Public Health and Surveillance, 3(4), e8641.

American Lung Association. (2023). Tobacco control milestones. https://www.lung.org/research/sotc/tobacco-timeline

Anspach NM (2017). The new personal influence: How our Facebook friends influence the news we read. Political Communication, 34(4), 590–606. 10.1080/10584609.2017.1316329

Azad A (2019, July 24). First on CNN: Facebook and Instagram to restrict content related to alcohol, tobacco and e-cigarettes. CNN. https://www.cnn.com/2019/07/24/health/facebook-instagram-alcohol-tobacco-bn/index.html

Bene M (2017). Go viral on the Facebook! Interactions between candidates and followers on Facebook during the Hungarian general election campaign of 2014. Information, Communication & Society, 20(4), 513–529. 10.1080/1369118X.2016.1198411

Benson R, Hu M, Chen AT, Nag S, Zhu S-H, & Conway M (2020). Investigating the attitudes of adolescents and young adults towards JUUL: Computational study using Twitter data. JMIR Public Health and Surveillance, 6(3), e19975. [PubMed: 32876579]

Blakemore JK, Bayer AH, Smith MB, & Grifo JA (2020). Infertility influencers: An analysis of information and influence in the fertility webspace. Journal of Assisted Reproduction and Genetics, 37(6), 1371–1378. 10.1007/s10815-020-01799-2 [PubMed: 32382959]

Bonnevie E, Rosenberg SD, Kummeth C, Goldbarg J, Wartella E, & Smyser J (2020). Using social media influencers to increase knowledge and positive attitudes toward the flu vaccine. PLOS ONE, 15(10), Article e0240828. 10.1371/journal.pone.0240828 [PubMed: 33064738]

Bonnevie E, Smith SM, Kummeth C, Goldbarg J, & Smyser J (2021). Social media influencers can be used to deliver positive information about the flu vaccine: Findings from a multi-year study. Health Education Research, 36(3), 286–294. 10.1093/her/cyab018 [PubMed: 34252187]

Borgmann H, Loeb S, Salem J, Thomas C, Haferkamp A, Murphy DG, & Tsaur I (2016). Activity, content, contributors, and influencers of the twitter discussion on urologic oncology. Urologic Oncology, 34(9), 377–383. 10.1016/j.urolonc.2016.02.021 [PubMed: 27036217]

Bornmann L, & Haunschild R (2015). t factor: A metric for measuring impact on Twitter (arXiv preprint arXiv:1508.02179). https://arxiv.org/abs/1508.02179

Brien EK, Hoffman L, Navarro MA, & Ganz O (2020). Social media use by leading U.S. e-cigarette, cigarette, smokeless tobacco, cigar and hookah brands. Tobacco Control, 29(e1), e87. 10.1136/tobaccocontrol-2019-055406 [PubMed: 32217772]

Broder AZ, Glassman SC, Manasse MS, & Zweig G (1997). Syntactic clustering of the web. Computer Networks and ISDN Systems, 29(8–13), 1157–1166.

Campaign for Tobacco Free Kids. (2018, August 27). New investigation exposes how tobacco companies market cigarettes on social media in the U.S. and around the world. https://www.tobaccofreekids.org/press-releases/2018_08_27_ftc

Campaign for Tobacco Free Kids. (2019, May 22). Over 125 organizations call on social media companies to end all tobacco advertising, including by paid influencers. https://www.tobaccofreekids.org/press-releases/2019_05_21_socialmedia_advertising

Campbell C, & Farrell JR (2020). More than meets the eye: The functional components underlying influencer marketing. Business Horizons, 63(4), 469–479.

Chapman M (2022, July 21). New products, old tricks? Concerns Big Tobacco is targeting youngsters. The Bureau of Investigative Journalism. https://www.thebureauinvestigates.com/stories/2021-02-21/new-products-old-tricks-concerns-big-tobacco-is-targeting-youngsters

Chowdhary KR (2020). Natural language processing. In Fundamentals of artificial intelligence (pp. 603–649). Springer. 10.1007/978-81-322-3972-7_19

Chu K-H, Colditz JB, Primack BA, Shensa A, Allem J-P, Miller E, Unger JB, & Cruz TB (2018). JUUL: Spreading online and offline. Journal of Adolescent Health, 63(5), 582–586.

CNN. (2022, February 8). Tweet message. https://twitter.com/CNN/status/1491073135285841923

Colditz JB, Chu K-H, Emery SL, Larkin CR, James AE, Welling J, & Primack BA (2018). Toward real-time infoveillance of Twitter health messages. American Journal of Public Health, 108(8), 1009–1014. 10.2105/AJPH.2018.304497 [PubMed: 29927648]

Conger K (2023, April 20). Twitter begins removing Blue check marks from verified accounts. The New York Times. https://www.nytimes.com/2023/04/20/technology/twitter-blue-check-marks.html

Coulter A (2020, January 10). Tweet message. https://twitter.com/AnnCoulter/status/1215638144210685957

Czaplicki L, Kostygina G, Kim Y, Perks SN, Szczypka G, Emery SL, Vallone D, & Hair EC (2020). Characterising JUUL-related posts on Instagram. Tobacco Control, 29(6), 612–617. 10.1136/tobaccocontrol-2018-054824 [PubMed: 31266903]

Czaplicki L, Tulsiani S, Kostygina G, Feng M, Kim Y, Perks SN, Emery S, & Schillo B (2020). #toolittletoolate: JUUL-related content on Instagram before and after self-regulatory action. PLOS ONE, 15(5), Article e0233419. 10.1371/journal.pone.0233419 [PubMed: 32437397]

Drescher LS, Roosen J, Aue K, Dressel K, Schär W, & Götz A (2021). The spread of COVID-19 crisis communication by German public authorities and experts on Twitter: Quantitative content analysis. JMIR Public Health and Surveillance, 7(12), e31834. 10.2196/31834 [PubMed: 34710054]

Dumas JE, & Stough RA (2022). When influencers are not very influential: The negative effects of social media verification. Journal of Consumer Behaviour, 21(3), 614–624. 10.1002/cb.2039

Edgerton E, Reiney E, Mueller S, Reicherter B, Curtis K, Waties S, & Limber SP (2016). Identifying new strategies to assess and promote online health communication and social media outreach: An application in bullying prevention. Health Promotion Practice, 17(3), 448–456. 10.1177/1524839915620392 [PubMed: 26921363]

Enberg J (2018, July 31). Marketers around the world are banking on microinfluencers. eMarketer. https://www.insiderintelligence.com/content/why-marketers-are-banking-on-microinfluencers

Fadus MC, Smith TT, & Squeglia LM (2019). The rise of e-cigarettes, pod mod devices, and JUUL among youth: Factors influencing use, health implications, and downstream effects. Drug and Alcohol Dependence, 201, 85–93. [PubMed: 31200279]

Ferrara E (2018). Measuring social spam and the effect of bots on information diffusion in social media. In Complex spreading phenomena in social systems: Influence and contagion in real-world social networks (pp. 229–255). https://arxiv.org/abs/1708.08134

Gallagher RJ, Doroshenko L, Shugars S, Lazer D, & Foucault Welles B (2021). Sustained online amplification of COVID-19 elites in the United States. Social Media + Society, 7(2). 10.1177/20563051211024957

Geyser W (2022, April 4). What is an influencer? Social media influencers defined [updated 2022]. Influencer Marketing Hub. https://influencermarketinghub.com/what-is-an-influencer

Geyser W (2023, October 30). The state of influencer marketing 2023: Benchmark report. Influencer Marketing Hub https://influencermarketinghub.com/influencer-marketing-benchmark-report

Glasser AM, Nemeth JM, Quisenberry AJ, Shoben AB, Trapl ES, & Klein EG (2023). The role of cigarillo flavor in the co-use of cigarillos and cannabis among young adults. Substance Use & Misuse, 58, 717–727. [PubMed: 36861952]

Gough A, Hunter RF, Ajao O, Jurek A, McKeown G, Hong J, Barrett E, Ferguson M, McElwee G, McCarthy M, & Kee F (2017). Tweet for behavior change: Using social media for the dissemination of public health messages [Original paper]. JMIR Public Health and Surveillance, 3(1), e14. 10.2196/publichealth.6313 [PubMed: 28336503]

Gu J, Abroms LC, Broniatowski DA, & Evans WD (2022). An investigation of influential users in the promotion and marketing of heated tobacco products on Instagram: A social network analysis. International Journal of Environmental Research and Public Health, 19(3), 1686. 10.3390/ijerph19031686 [PubMed: 35162709]

Guo T-C, & Li X (2016). Positive relationship between individuality and social identity in virtual communities: Self-categorization and social identification as distinct forms of social identity Cyberpsychology, Behavior, and Social Networking, 19(11), 680–685. [PubMed: 27831758]

Gwet KL (2008). Computing inter-rater reliability and its variance in the presence of high agreement. British Journal of Mathematical and Statistical Psychology, 61(1), 29–48. 10.1348/000711006X126600 [PubMed: 18482474]

Harding K, Pérez-Escamilla R, Carroll G, Aryeetey R, & Lasisi O (2019). Four dissemination pathways for a social media-based breastfeeding campaign: Evaluation of the impact on key performance indicators. JMIR Nursing, 2(1), e14589. 10.2196/14589 [PubMed: 34345773]

Hejlová D, Schneiderová S, Klabíková Rábová T, & Kulhánek A (2019). Analysis of presumed IQOS influencer marketing on Instagram in the Czech Republic in 2018–2019. Adiktologie, 19(1), 7–15.

Hickman A (2021, February 22). "Big tobacco" using COVID-19 messaging and influencers to market products. PRWeek. https://www.prweek.com/article/1683314/big-tobacco-using-covid-19-messaging-influencers-market-products

Hirsch JE (2005). An index to quantify an individual's scientific research output. Proceedings of the National Academy of Sciences of the United States of America, 102(46), 16569–16572. 10.1073/pnas.0507655102 [PubMed: 16275915]

Huang J, Duan Z, Kwok J, Binns S, Vera LE, Kim Y, Szczypka G, & Emery SL (2019). Vaping versus JUULing: How the extraordinary growth and marketing of JUUL transformed the U.S. retail e-cigarette market. Tobacco Control, 28(2), 146–151. 10.1136/tobaccocontrol-2018-054382 [PubMed: 29853561]

Jackler RK, Chau C, Getachew BD, Whitcomb MM, Lee-Heidenreich J, Bhatt AM, Kim-O'Sullivan SHS, Hoffman ZA, Jackler LM, & Ramamurthi D (2019). JUUL advertising over its first three years on the market [White paper for Stanford research into the impact of tobacco advertising, issue]. https://tobacco-img.stanford.edu/wp-content/uploads/2021/07/21231836/JUUL_Marketing_Stanford.pdf

Jain S, & Sinha A (2020). Identification of influential users on Twitter: A novel weighted correlated influence measure for Covid-19. Chaos, Solitons & Fractals, 139, 110037. 10.1016/j.chaos.2020.110037 [PubMed: 32834597]

Johnston LD, Miech RA, O'Malley PM, Bachman JG, Schulenberg JE, & Patrick ME (2019). Monitoring the future national survey results on drug use, 1975–2018: Overview, key findings on adolescent drug use. Institute for Social Research.

Jun J, Zhang N, Zain A, & Mohammadi E (2022). Social media discussions on the FDA's modified risk tobacco product authorization of IQOS. Substance Use & Misuse, 57(3), 472–480. [PubMed: 34986075]

Kaplan S (2018, August 24). Big tobacco's global reach on social media. The New York Times. https://www.nytimes.com/2018/08/24/health/tobacco-social-media-smoking.html

Kim I, & Valente TW (2021). COVID-19 health communication networks on Twitter: Identifying sources, disseminators, and brokers. Connections, 40(1), 129–142.

Kim Y, Huang J, & Emery S (2016). Garbage in, garbage out: Data collection, quality assessment and reporting standards for social media data use in health research, infodemiology and digital disease detection. Journal of Medical Internet Research, 18(2), e41. 10.2196/jmir.4738 [PubMed: 26920122]

Klein EG, Czaplicki L, Berman M, Emery S, & Schillo B (2020). Visual attention to the use of #ad versus #sponsored on e-cigarette influencer posts on social media: A randomized experiment. Journal of Health Communication, 25(12), 925–930. 10.1080/10810730.2020.1849464 [PubMed: 33238805]

Klein EG, Kierstead E, Czaplicki L, Berman ML, Emery S, & Schillo B (2022). Testing potential disclosures for e-cigarette sponsorship on social media. Addictive Behaviors, 125, 107146. 10.1016/j.addbeh.2021.107146 [PubMed: 34688982]

Kolmar C (2023, April 16). The 15 largest tobacco companies in the world. Zippia. https://www.zippia.com/advice/largest-tobacco-companies

Kong G, Laestadius L, Vassey J, Majmundar A, Stroup AM, Meissner HI, Taleb ZB, Cruz TB, Emery SL, & Romer D (2022). Tobacco promotion restriction policies on social media. Tobacco Control. Advance online publication. 10.1136/tc-2022-057348

Kostygina G, Feng M, Czaplicki L, Tran H, Tulsiani S, Perks SN, Emery S, & Schillo B (2021). Exploring the discursive function of hashtags: A semantic network analysis of JUUL-related Instagram messages. Social Media+ Society, 7(4).

Kostygina G, Huang J, & Emery S (2017). TrendBlendz: How Splitarillos use marijuana flavours to promote cigarillo use. Tobacco Control, 26(2), 235–236. 10.1136/tobaccocontrol-2015-052710 [PubMed: 27052097]

Kostygina G, Tran H, Binns S, Szczypka G, Emery S, Vallone D, & Hair E (2020). Boosting health campaign reach and engagement through use of social media influencers and memes. Social Media + Society, 6(2). 10.1177/2056305120912475

Kostygina G, Tran H, Kim Y, Czaplicki L, Kierstead E, Kreslake J, Emery S, & Schillo B (2022). Heating up: How early Twitter marketing gave rise to organic word-of-mouth about heated tobacco products. Social Media+ Society, 8(4).

Kostygina G, Tran H, Shi Y, Kim Y, & Emery S (2016). "Sweeter Than a Swisher": Amount and themes of little cigar and cigarillo content on Twitter. Tobacco Control, 25(Suppl. 1), i75. 10.1136/tobaccocontrol-2016-053094 [PubMed: 27697951]

Leung FF, Gu FF, & Palmatier RW (2022). Online influencer marketing. Journal of the Academy of Marketing Science, 50(2), 226–251. 10.1007/s11747-021-00829-4

Lin Y (2023). U.S. influencer marketing spend (2019–2024). Oberlo.

Liu J, McLaughlin S, Lazaro A, & Halpern-Felsher B (2020). What does it meme? A qualitative analysis of adolescents' perceptions of tobacco and marijuana messaging. Public Health Reports, 135(5), 578–586. [PubMed: 32791026]

Liu X (2019). A big data approach to examining social bots on Twitter. Journal of Services Marketing, 33, 369–379.

Lutkenhaus RO, Jansz J, & Bouman MPA (2019). Tailoring in the digital era: Stimulating dialogues on health topics in collaboration with social media influencers. Digital Health, 5. 10.1177/2055207618821521

MacKay M, Ford C, Colangeli T, Gillis D, McWhirter JE, & Papadopoulos A (2022). A content analysis of Canadian influencer crisis messages on Instagram and the public's response during COVID-19. BMC Public Health, 22(1), Article 763. 10.1186/s12889-022-13129-5 [PubMed: 35428287]

Main S (2017, March 30). Micro-influencers are more effective with marketing campaigns than highly popular accounts. Adweek. https://www.adweek.com/performance-marketing/micro-influencers-are-more-effective-with-marketing-campaigns-than-highly-popular-accounts

Malik A, Khan MI, Karbasian H, Nieminen M, Ammad-Ud-Din M, & Khan SA (2021). Modeling public sentiments about JUUL flavors on Twitter through machine learning. Nicotine & Tobacco Research, 23(11), 1869–1879. 10.1093/ntr/ntab098 [PubMed: 33991191]

McCombs ME, Shaw DL, & Weaver DH (2013). Communication and democracy: Exploring the intellectual frontiers in agenda-setting theory. Routledge.

Mole B (2022, November 30). Musk's Twitter abandons COVID misinfo policy, shirking "huge responsibility." Ars Technica. https://arstechnica.com/science/2022/11/musks-twitter-abandons-covid-misinfo-policy-shirking-huge-responsibility

Moukarzel S, Caduff A, Rehm M, Del Fresno M, Pérez-Escamilla R, & Daly AJ (2021). Breastfeeding communication strategies, challenges and opportunities in the Twitter-verse: Perspectives of influencers and social network analysis. International Journal of Environmental Research and Public Health, 18(12), 6181. 10.3390/ijerph18126181 [PubMed: 34201000]

Moukarzel S, Rehm M, & Daly AJ (2020). Breastfeeding promotion on Twitter: A social network and content analysis approach. Maternal & Child Nutrition, 16(4), e13053. 10.1111/mcn.13053 [PubMed: 32638522]

Munoz-Acuna R, Leibowitz A, Hayes M, & Bose S (2021). Analysis of top influencers in critical care medicine "Twitterverse" in the COVID-19 era: A descriptive study. Critical Care, 25(1), 254. 10.1186/s13054-021-03691-6 [PubMed: 34284805]

Navarro MA, O'Brien EK, Ganz O, & Hoffman L (2021). Influencer prevalence and role on cigar brand Instagram pages. Tobacco Control, 30(e1), e33–e36. [PubMed: 33046583]

Nedelman R, & Selig R (2018, December 19). JUUL: How social media hyped nicotine for a new generation. CNN https://www.cnn.com/2018/12/17/health/juul-social-media-influencers/index.html

NetworkX. (2023, October). NetworkX—NetworkX documentation. https://networkx.org

Patel VR, Gereta S, Blanton CJ, Chu AL, Reddy NK, Mackert M, Nortjé N, & Pignone MP (2022). #ColonCancer: Social media discussions about colorectal cancer during the COVID-19 pandemic. JCO Clinical Cancer Informatics, 6, e2100180. 10.1200/cci.21.00180 [PubMed: 35025670]

Ramadhan FA, & Bigwanto M (2022). Online promotion of e-cigarettes in Indonesia: One-month monitoring in selected platforms. https://seatca.org/dmdocuments/Eng_Online%20Promotion%20of%20ENDS%20in%20Indonesia_23022022.pdf

Riff D, Lacy S, & Fico F (2014). Analyzing media messages: Using quantitative content analysis in research. Routledge.

Rosenstein J (2017, May 16). Why is everyone smoking again? Harper's BAZAAR. https://www.harpersbazaar.com/beauty/a9554018/smoking-on-instagram

Scutti S (2018, November 15). Juul to eliminate social media accounts, stop retail sales of flavors. CNN. https://www.cnn.com/2018/11/13/health/juul-flavor-social-media-fda-bn/index.html

Silver N, Kucherlapaty P, Kostygina G, Tran H, Feng M, Emery S, & Schillo B (2022). Discussions of flavored ends sales restrictions: Themes related to circumventing policies on Reddit. International Journal of Environmental Research and Public Health, 19(13), 7668. [PubMed: 35805325]

Silver NA, Feng M, Kierstead EC, Tran H, Binns S, Emery SL, & Schillo BA (in press). Unraveling public opinion from industry agenda—setting: Characterizing the most amplified, influential, and connected voices driving Twitter discourse about tobacco regulatory policy from September 2019—July 2021. Social Media + Society.

SocialPubli. (2019). 2019 influencer marketing report: A marketer's perspective. https://socialpubli.com/blog/2019-influencer-marketing-report-a-marketers-perspective/

Statista. (2022). Number of social media users worldwide from 2018 to 2027 (in billions). https://www-statista-com.proxy.uchicago.edu/statistics/278414/number-of-worldwide-social-network-users/

Strauss A, & Corbin JM (1997). Grounded theory in practice. Sage.

Tiidenberg K (2021). Sex, power and platform governance. Porn Studies, 8(4), 381–393.

Triton Digital. (2021). Percentage of U.S. population who currently use any social media from 2008 to 2021. Statista. https://www-statista-com.proxy.uchicago.edu/statistics/273476/percentage-of-us-population-with-a-social-network-profile/

Twitter. (2023, April 24). Historical PowerTrack API. https://developer.twitter.com/en/docs/twitter-api/enterprise/historical-powertrack-api/overview

U.S. Food and Drug Administration. (2022, June 27). FDA proposes rules prohibiting menthol cigarettes and flavored cigars to prevent youth initiation, significantly reduce tobacco-related disease and death. https://www.fda.gov/news-events/press-announcements/fda-proposes-rules-prohibiting-menthol-cigarettes-and-flavored-cigars-prevent-youth-initiation

U.S. Food and Drug Administration, Center for Tobacco Products. (2020, December 7). Marketing granted order: Philip Morris products. https://www.fda.gov/media/144700/download

Vassey J, Valente T, Barker J, Stanton C, Li D, Laestadius L, Cruz TB, & Unger JB (2023). E-cigarette brands and social media influencers on Instagram: A social network analysis. Tobacco Control, 32, e184–e191. 10.1136/tobaccocontrol-2021-057053 [PubMed: 35131947]

Vogel EA, Guillory J, & Ling PM (2020). Sponsorship disclosures and perceptions of e-cigarette Instagram posts. Tobacco Regulatory Science, 6(5), 355–368. 10.18001/trs.6.5.5 [PubMed: 33778107]

Wang X, Kim Y, Borowiecki M, Tynan MA, Emery S, & King BA (2021). Trends in cigar sales and prices, by product and flavor type—The United States, 2016–2020. Nicotine & Tobacco Research, 24(4), 606–611. 10.1093/ntr/ntab238

Williamson DA (2020, November 24). U.S. social media advertising in 2021. Insider Intelligence. https://www.insiderintelligence.com/content/us-social-media-advertising-in-2021

Wu J, Harlow AF, Wijaya D, Berman M, Benjamin EJ, Xuan Z, Hong T, & Fetterman JL (2022). The impact of influencers on cigar promotions: A content analysis of large cigar and swisher sweets videos on TikTok. International Journal of Environmental Research and Public Health, 19(12), 7064. https://www.mdpi.com/1660-4601/19/12/7064 [PubMed: 35742315]

Yang K-C, Ferrara E, & Menczer F (2022). Botometer 101: Social bot practicum for computational social scientists (arXiv preprint arXiv:2201.01608). https://arxiv.org/abs/2201.01608

Yang K-C, Varol O, Hui P-M, & Menczer F (2020). Scalable and generalizable social bot detection through data selection. In Proceedings of the AAAI conference on artificial intelligence. https://ojs.aaai.org/index.php/AAAI/article/view/5460

Zheng C, Wang W, & Young SD (2021). Identifying HIV-related digital social influencers using an iterative deep learning approach. AIDS, 35(Suppl. 1), S85–S89. 10.1097/qad.0000000000002841 [PubMed: 33867491]

Zote J (2019, July 29). What are fake influencers and how can you spot them? Sprout Social. https://sproutsocial.com/insights/fake-influencers

Zou W, Zhang WJ, & Tang L (2021). What do social media influencers say about health? A theory-driven content analysis of top ten health influencers' posts on sina weibo. Journal of Health Communication, 26(1), 1–11. 10.1080/10810730.2020.1865486 [PubMed: 33372857]

Table 1.

Influential Account Coding Category Descriptions.

Coding Category	
Commercial (LCC/Tobacco)	Accounts that are a business or an individual selling LCC products. This includes LCC brands, manufacturers, apps, trade press, cigar bars or cigar smoking lounges, and vendors, e.g., @CaponeCigarillo, and @CigarDojo.
Commercial (Non-Tobacco, Non-Marijuana)	Non-LCC/non-marijuana (e.g., smokeless, e-cigarettes, cigarettes, alcohol, restaurants, hotels, etc.) accounts such as brands, manufacturers, apps, trade press (non-tobacco industry), vendors, etc.
Political/Government	Accounts for political organizations and government agencies, current or former politicians, government agency employees (non-health), policymakers, and aspiring candidates, excluding political commentators.
Health/Medical	Public health/healthcare organizations, healthcare providers, and health researchers/scientists (i.e., CDC, a doctor, and a cancer researcher).
Political Commentator	Accounts who identify as political commentators for news media organizations. They most likely will promote content on other platforms, and their podcast/T.V./other show name may contain the individual's name. They are often described as commentators (vs. journalists) on sites such as Wikipedia.
News Media	Newspaper, radio, T.V., other news outlets (e.g., news blogs), broadcasters, journalists, and reporters, excluding political commentators, and non-news (entertainment news/lifestyle magazines).
LCC/Tobacco Policy Counter-Advocate	Accounts that advocate against tobacco regulation.
LCC Community (Enthusiast, Aficionado)	Individual users (e.g., enthusiast, aficionado, and hobbyist) such as @GeniusPothead and @RollBlunts, where the majority of content is LCC-related. This includes blunts, cigars, cigarillos, little cigars, or other LCC products.
LCC Community (Aggregator)	Community accounts (e.g., aggregators) are primarily dedicated to LCC content (e.g., @StonerClips), where most content is LCC-related. This includes blunts, cigars, cigarillos, little cigars, or other LCC products.
Anti-Tobacco/LCC	Accounts that advocate for tobacco regulation/against tobacco/LCC use.
Adult or Pornographic (NSFW)	Accounts primarily dedicated to pornographic content/adult NSFW content.
Entertainment	Musicians/artists/performers/comedians, athletes/personal trainers, bloggers/vloggers (non-news), gamers, gamer/fan community accounts, entertainment news, lifestyle magazines, etc.
Marijuana Business or Advocate	Accounts for marijuana businesses (e.g., dispensaries and lounges), marijuana advocates.
Marijuana Community or User	Marijuana enthusiasts/users, marijuana community accounts, marijuana aggregator accounts. This includes anything that involves marijuana itself (not marijuana in a hollowed-out cigar, which LCC would cover).
Other	Accounts that do not fit into these categories, mostly organic unverified X (formerly Twitter) users.

Note. LCC = little cigar and cigarillo; CDC = Centers for Disease Control and Prevention.

Table 2.

Pairwise Jaccard Similarity Coefficients Across the Seven Influential Account Detection Methods.

	H-Index	"Sources"	"Disseminators"	"Brokers"	Mentions	Retweets	H-Index
H-Index	-	0.064	0.031	0.058	0.020	0.026	0.020
"Sources"	0.064	-	0.015	0.081	0.005	-	0.015
"Disseminators"	0.031	0.015	-	0.130	0.036	0.307	-
"Brokers"	0.058	0.081	0.130	-	0.042	0.081	0.005
Mentions	0.020	0.005	0.036	0.042	-	0.058	0.015
Retweets	0.026	-	0.307	0.081	0.058	-	-
Followers	0.020	0.015	-	0.005	0.015	-	-

Note. "Sources" = In-Degree Centrality. "Disseminators" = Out-degree Centrality. "Brokers" = Betweenness Centrality.

Table 3.

Jaccard Similarity Across the 3 Weeks Following the FDA Menthol Ban Announcement.

Metric	Weeks 1 and 2	Weeks 2 and 3
Retweets	379	408
Mentions	205	190
h-Index	198	282
"Disseminators"	105	212
Followers	081	149
"Brokers"	081	105
"Sources"	064	099

Note. FDA = U.S. Food and Drug Administration.

Table 4.

Top 5 Account Category Types by Influential Account Detection Metric Among Relevant Active Accounts.

Total	Index	Network			Simple		
	H-Index	"Sources"	"Disseminators"	"Brokers"	Mentions	Followers	Retweets
Other[UV] 19.6% (71)	NSFW[UV] 27.1% (16)	Other[UV] 22.0% (18)	Other[UV] 21.9% (20)	Other[UV] 24.7% (24)	LCC Users[UV] 40.2% (39)	News Media[V] 51.8% (29)	LCC Users[UV] 23.3% (20)
NSFW[UV] 17.9% (65)	Other[UV] 13.6% (8)	Entertainment[U.V.] 13.4% (11)	NSFW[UV] 18.8% (18)	Entertainment[U.V.] 17.5% (17)	NSFW[UV] 13.4% (13)	Entertainment[V] 39.3% (22)	NSFW[U.V.] 20.9% (18)
LCC Users[UV] 14.6% (53)	Entertainment[UV] 11.9% (7)	NSFW[UV] 9.8% (8)	Entertainment[U.V.] 11.5% (11)	Marijuana Users[UV] 6.2% (6)	Entertainment[V] 9.3% (9)		LCC Commercial[UV] 8.1% (7)
Entertainment[V] 12.7% (46)	Commercial Other[UV] 10.2% (6)	Entertainment[V] 8.5% (7)	LCC Users[UV] 9.4% (9)	LCC Users[UV] 6.2% (6)	Other[U.V.] 6.2% (6)		Entertainment[U.V.] 8.1% (7)
Entertainment[UV] 10.2% (37)	Marijuana Users[U.V.] 10.2% (6)		LCC Commercial[U.V.] 5.2% (5)	NSFW[UV] 5.2% (5)	LCC Commercial[UV] 5.2% (5)		Other[UV] 8.1% (7)
Likely Bot (%, n) 19.3% (70)	18.6% (11)	14.5% (8)	34.7% (25)	20.3% (13)	21.3% (17)	0.0% (0)	30.4% (21)

Note. Only top account categories are reflected, so categories do not add up to 100%. There were no Anti-LCC or Pro-LCC accounts captured in the total sample. Based on X (formerly Twitter)'s outlined criteria, accounts that received a blue verified badge before November 9, 2022, are considered authentic, notable, and active. "Total" accounts have been de-duplicated. Relevant active accounts include unverified accounts labeled as likely bots (SGBot probability score above .80). For the "Sources" detection metric, the prevalence of the fifth and sixth most common categories, Other Commercial and News Media, each made up 7.3% of relevant active accounts. For the Follower detection metric, all subsequent account categories (Other Commercial, Political/Government, Political Commentator, Health/ Medical, and Unverified Entertainment) comprised 1.8% of relevant active accounts. LCC = little cigar and cigarillo; NSFW = pornographic or adult; "Sources" = In-Degree Centrality; "Disseminators" = Out-degree Centrality; "Brokers" =betweenness Centrality; U.V. = unverified account; V = verified account.

Table 5.

Content Features in the Social Media Messages (Tweets) Posted by Accounts Captured Across the Seven Influential Account Detection Metrics.

Metrics	Top 5	Hyperlink Percentage	Top Tri-grams (n)	Salient Themes
Followers	#covid19 (3), #coronavaccine (2), #smackdown (2)	81	Ban menthol cigarettes (20)	Menthol and Flavor Ban
Mentions	#cigars (631), #botl (446), #nowsmoking (219), #sotl (160), #cigar (149)	45	Eat drink smoke (60)	Cigar Promotion
Retweets	#cigars (1637), #cigar (1201), #botl (676), #sotl (438), #smoking (363)	66	Join beautiful girls (264)	NSFW
H-Index	#cigar (212), #cigars (149), #patreon (115), #smoking (115), #sexy (115)	58	Join beautiful girls (115)	NSFW
"Brokers"	#cigars (399), #botl (308), #cigar (171), #mmemberville (150), #sotl (121)	40	Smoke pairing enjoy (15)	Casual Blunt Use
"Sources"	#buyile (5), #ivanshishkin (3), #realism (3)	46	Menthol cigarettes flavored	Menthol and Flavor Ban
"Disseminators"	#cigars (1046), #cigar (744), #botl (483), #sotl (299), #cigarlife (249)	60	Latest cigars daily (113)	Cigar Promotion

Note. NSFW = pornographic or adult.