



HHS Public Access

Author manuscript

J Health Commun. Author manuscript; available in PMC 2024 August 14.

Published in final edited form as:

J Health Commun. 2023 August 03; 28(8): 487–497. doi:10.1080/10810730.2023.2225447.

Community and Opinion Leadership Effects on Vaping Discourse: A Network Analysis of Online Reddit Threads

JACOB A. ROHDE,

JESSICA LIU,

VAUGHAN W. REES

Harvard T.H. Chan School of Public Health, Boston, MA, USA

Abstract

Reddit is a popular hub for discussing vaping. A deeper understanding of the factors that influence this online discourse could inform public health messaging efforts targeting this platform. Using a network analysis framework, we sought to investigate the role of opinion leaders and online communities in facilitating vaping discussions on Reddit. We collected Reddit submissions about vaping posted in May 2021 and used these submissions to generate subreddit-level ($N=261$) and thread-level ($N=8,377$) data sets. We coded subreddits into four community categories: 1) Vaping, 2) Substance use, 3) Cessation, and 4) Non-specific. We used sociometric in-degree centrality statistics to identify subreddit opinion leaders. We computed non-parametric ANOVAs and negative binomial regressions to test associations between opinion leadership and subreddit community category variables on subreddit network composition (comprised of subreddit-level network nodes and edges) and the number of commenters on Reddit threads about vaping (thread-level). Subreddit network composition was largely dependent on opinion leaders in Non-specific communities, and less so in Vaping and Substance use communities. At the thread-level, the rate of commenters was higher among threads initiated by opinion leaders than non-opinion leaders (adjusted rate ratio [aRR]=4.84). Furthermore, threads posted in Vaping (aRR=1.64), Substance use (aRR=1.92), and Cessation (aRR=1.21) communities had higher commenter rates than those posted in Non-specific communities. Communities and opinion leaders play a key role in the composition and reach of vaping discussions on Reddit. These findings provide a foundation for public health campaigns and interventions targeting Reddit and perhaps other social media platforms.

Social media platforms are popular hubs for discussing vaping and e-cigarette use. Sites such as Twitter and Reddit are frequently used to promote, market, and sell vaping devices (Barker & Rohde, 2019; Lazard et al., 2016; McCausland, Maycock, Leaver, & Jancey,

Address correspondence to Jacob A. Rohde, Cancer Prevention Fellowship Program, Health Communication and Informatics Research Branch, Division of Cancer Control and Population Sciences, National Cancer Institute, Bethesda, MD, USA. jacob.rohde@nih.gov.

Disclosure Statement

JAR & VWR have served as paid consultants in litigation against tobacco companies. Opinions expressed by the authors are their own, and this material should not be interpreted as representing the official viewpoint of the U.S. Department of Health and Human Services, the National Institutes of Health, or the National Cancer Institute.

Supplemental data

Supplemental data for this article can be accessed online at <https://doi.org/10.1080/10810730.2023.2225447>.

2019), and to discuss some of the health effects associated with vaping (Kwon & Park, 2020). In addition, studies have shown that social media platforms are being used to share misinformation about vaping (Majmundar, Allem, Unger, & Cruz, 2021; Soule et al., 2020; Suarez-Lledo & Alvarez-Galvez, 2021). For example, a recent content analysis of nicotine and vaping posts on Twitter found that roughly 40% of tweets contained potential misinformation about the addictive properties of nicotine vaping devices (Sidani et al., 2022). Systematically analyzing patterns of vaping discussions on social media could inform targeted public health messaging campaigns disseminated on these platforms.

Networked Communities

Drawing from a social network analysis framework, online communities may influence discussions about vaping on social media (Valente, 2010). Broadly, a community is a group of connected individuals who provide support to one another, and have shared interests and life experiences (Wellman & Leighton, 1979). It should be noted that the term “community” is not synonymous with social network. While a social network can be thought of as a web of interconnected individuals, the types of connections between members in a social network are not predicated on closeness; though, communities are often subcomponents of larger social networks (Berkman, Glass, Brissette, & Seeman, 2000). In the context of social media, a community might be considered a group of people who subscribe to an online vaping forum.

Online communities foster engagement and promote identity. Research has shown that social media communities across different health topics, such as cancer and inflammatory bowel disease, offer an outlet for people to connect with and support one another (Coulson, 2013; van Eenbergen, van de Poll-Franse, Krahmer, Verberne, & Mols, 2018). This is also true in certain substance use contexts. For example, a recent content analysis of an online Chinese smoking cessation community found that peers frequently provided one another with informational and emotional support based on individuals’ current stages in the quitting process (Qian, Gui, Ma, & Dong, 2021). In addition, people who vape often cite community and a sense of belongingness as key aspects of vaping culture (Langley, Bell-Williams, Pattinson, Britton, & Bains, 2019; Lucherini, 2022; Tokle & Pedersen, 2019). This suggests that people who participate in online vaping communities may seek gratification by networking with and contributing to discussions about vaping (e.g., what devices to purchase) with other community members.

Currently, there is a dearth of work investigating vaping discourse throughout online communities. Data from Reddit may offer an opportunity to expand this research. Briefly, Reddit is a popular social media site founded in 2005. In 2021, nearly one-fifth of U.S. adults reported having used Reddit, which is a substantial increase from 11% in 2019 (Pew Research Center, 2021). Scientific research investigating Reddit has also been steadily increasing, speaking to the site’s growing role as a sphere for public discourse (Proferes, Jones, Gilbert, Fiesler, & Zimmer, 2021). A key feature of this platform is its use of subreddits, which are user-created, topic-based communities (Anderson, 2015). For example, the subreddit r/stopsmoking is a public community where Redditors can provide one another with smoking cessation support and advice. Redditors (i.e., Reddit

users) predominately interact with one another in subreddits, and the content posted to subreddits are typically managed by one or more moderators to ensure participants follow community rules. Reddit uses a thread-based model to organize its participation. Posts that initiate a thread in a subreddit are called “submissions” and the content posted in response to submissions are referred to as comments.

Vaping is a popular discussion topic on Reddit, and some of the vaping-centered subreddit communities have more than 200,000 subscribed members. A content analysis found that submissions discussing vaping on Reddit varied based on the subreddit communities they were published in (Barker & Rohde, 2019). For example, submissions about buying and selling vaping products were prevalent among vendor-related and general vaping subreddits (e.g., r/electronic_cigarette), whereas submissions discussing the various health effects of vaping were prevalent among research and news subreddits (e.g., r/news). These findings suggest that social media communities may influence online vaping discourse; however, it should be noted that this study did not assess interpersonal communication across different Reddit communities (e.g., Redditors commenting on one another’s posts). Thus, it remains unclear whether patterns of online vaping discourse vary across different online communities (e.g., pro vaping community vs. vaping cessation community).

Opinion Leadership

Opinion leadership is another important concept potentially useful for understanding discussions about vaping on social media. Research has characterized opinion leaders in various ways; though, the term broadly refers to influential individuals who are perceived as trustworthy among peers (Bamakan, Nurgaliev, & Qu, 2019). Early studies about opinion leadership effects stem from the two-step flow model of communication posited by Lazarsfeld, Berelson, and Gaudet (1948). In their model, communication from mass media (e.g., newspapers) tends to reach individuals most proximal to and interested in the information; these people are referred to as “opinion leaders.” After, these opinion leaders mediate the flow of communication by relaying the information to their respective social circles, such as friends, family members, and colleagues.

Studies have since expanded opinion leadership research beyond application in the two-step flow model—most notably, in social network contexts via the diffusions of innovation theory (Rogers, 1983; Valente, 1996). Diffusions of innovation examines how change permeates across a network, such as use of a new product or service. This paradigm suggests that the adoption of ideas or behaviors across networks is influenced by interpersonal communication and social contacts (Valente & Davis, 1999). That is, the more people who have adopted a product or behavior in a network, the more opportunities there are for word of mouth to spread and influence non-adopters.

Diffusions of innovation takes into account several factors that facilitate adoption, such as different “adopter categories” (e.g., early adopters, laggards) and adoption thresholds (Valente, 1996). This paper focuses on one particular theoretical property, which concerns the thoughts and behaviors of opinion leaders in social networks. Within a diffusions of innovation paradigm, research suggests that opinion leaders act as “change agents” in the

spread of information, and are often the first individuals in a network to adopt a behavior (Valente & Davis, 1999; Valente & Pumpuang, 2007). Take for example the creator or a moderator of an online discussion forum, which can be considered a type of opinion leader. Because of their centrality and influential status in the online forum, this individual's promotion of a behavior or product is more likely to diffuse to a large number of community members and at a faster speed than general forum participants (i.e., non-opinion leaders) sharing the same information.

In the context of vaping and tobacco use, opinion leaders may be key in promoting ideas and marketing emerging products such as new e-juice flavors or new commercial vaping devices (i.e., diffusion). Such clout could spread throughout a network and encourage susceptible individuals to use these products. For example, Vassey et al. (2022) analyzed public “influencers” discussing vaping on Instagram and found that a large proportion of their followers were teenagers. Another study showed that followers of online tobacco use opinion leaders were more likely to also use tobacco products themselves compared to non-followers (Chu et al., 2019). These findings indicate that opinion leaders may influence vaping-related perceptions and behaviors.

In the past, opinion leaders have been thought of as celebrities, politicians, and those who are well-educated or who come from high socioeconomic statuses (Rogers, 1983). In social media research, however, evidence has shown little association between sociodemographic characteristics and peoples' perceptions of opinion leaders (Park & Kaye, 2017). Instead, studies have utilized network measures of centrality (e.g., in-degree) or the frequency of participation in a network to identify opinion leaders (Chu et al., 2019; Dewi, Yudhoatmojo, & Budi, 2017; Kilgo et al., 2016).

To date, few studies have examined opinion leadership effects among vaping discussions on social media. Briefly mentioned above, Chu et al. (2019), using online social networks of users tweeting about tobacco products, found that opinion leaders had the highest prevalence of current vaping, compared with followers of opinion leaders and general social media users. It should be noted that Chu and colleagues' study used social media data only to identify and enroll participants in an online survey. In other words, their research did not examine possible opinion leadership effects occurring naturally in online discourse. Moreover, their investigation did not assess the influence of opinion leaders in respect to certain network characteristics, such as whether opinion leaders play a central role in the composition of online vaping communities.

Study Aims

Drawing from the above review, this study sought to investigate whether two distinct yet interrelated concepts—networked communities and opinion leadership—affect discussions of vaping on Reddit. Specifically, we aimed to investigate 1) the influence of opinion leaders in shaping the composition of different online subreddit communities, and 2) the impact of communities and opinion leaders on the reach of vaping discussions.

Influenced by opinion leadership effects within a diffusion of innovation paradigm, we hypothesized in the first aim that a large proportion of online vaping discourse across subreddit communities would occur through opinion leaders given their influential position within their respective social networks. To address gaps in vaping and online community research, we also sought to answer whether opinion leaders had effects on community composition based on the type of subreddit communities that vaping discourse occurred in (e.g., vaping vs. non-vaping community). Our second aim was also informed, in part, by diffusion of innovation. Specifically, we hypothesized that the reach of discussion threads about vaping on Reddit would be greater when initiated by opinion leaders. Similar to aim 1, we also tested whether the reach of discussion threads varied based on the type of subreddit community that threads were initiated in.

Ultimately, the goal of our study was to better understand patterns of vaping discourse on Reddit to inform public health interventions and messaging efforts among this audience.

Methods

Data Collection and Cleaning

We collected data for the current study from Pushshift—an online archive of publicly available Reddit posts dating back to 2005 (Baumgartner, Zannettou, Keegan, Squire, & Blackburn, 2020). Our intention was to analyze a single “snapshot” in time of vaping discourse, and so we used the most recent monthly data available on Pushshift at the time of conducting this work. Using Reddit submissions (original posts) published in May 2021, we extracted approximately 80 vaping-related keywords. Keywords included terms such as “e-cig” and “pod mod,” as well as popular vaping product names such as “juul,” “sourin,” and “puff bar.” Our choice of keywords was also informed by literature examining discussions of vaping on social media and from web searches for different brand names of vaping devices (Barker & Rohde, 2019; Lazard et al., 2016). Our keywords did not restrict Reddit submissions about non-nicotine or other substance use (e.g., cannabis) vaping behaviors (Gaiha, Lin, Lempert, & Halpern-Felsher, 2022; Pearson & Villanti, 2020). Collection criteria were case-insensitive (e.g., “VAPE” and “vape” were treated as identical terms). This process yielded 27,607 Reddit submissions. Metadata variables collected in this process included the unique username of the posting Redditor, the date and time of the post, and the subreddit the post was published in.

Preliminary descriptive analyses of the text data revealed some submissions unrelated to vaping. For example, one of our vaping-related keywords was “mistic,” a pod-based vaping product. Including this keyword collected posts from Redditors talking about how “optimistic” they were. Thus, we re-assessed our list of keywords and removed those that were not unambiguously about vaping or e-cigarettes (see Supplemental Material for a full list of final keywords used). We also removed false positive posts irrelevant to the current study. For example, one of our keywords was “puff bar.” While most posts discussing this keyword were about vaping, some were instead about “cocoa puff bars,” a cereal treat.

We removed auto-generated submissions published by subreddit moderator bots (signified by author name: “AutoModerator”). We also removed submissions posted by now deleted

Redditors because they shared the same author metadata (i.e., “[deleted]”), meaning we would not be able to distinguish unique Redditors among these submissions. The size of the data set after removing false positives and applying cleaning procedures was 8,377 unique submissions about vaping.

Creating Thread- and Subreddit-Level Data Sets

We used unique ID values to extract relevant comments to the collected submissions (as well as their corresponding metadata) from Pushshift (Baumgartner, Zannettou, Keegan, Squire, & Blackburn, 2020). We searched for comments through the end of June 2021 to account for long, engaging threads. This approach ensured we captured comments published in June that were in response to submissions published late in May. We did not retain comments posted by now deleted Redditors and auto-generated moderator bots. This process yielded 81,071 comments corresponding to the Reddit submissions. Using both submissions and comments, we created two separate network data sets: **1**) a subreddit-level data set where each unit of analysis was a subreddit network constructed by aggregating and linking submission and comment authors within subreddits, and **2**) a thread-level data set where each unit of analysis was a single Reddit thread initiated by a submission and its corresponding comments. The size of the subreddit-level and thread-level data sets were 261 (subreddit networks) and 8,377 (threads). See Figure 1 for an example of how we constructed the two data sets.

Categorizing Subreddit Communities

We coded the 261 subreddits into four mutually exclusive community categories: 1) Vaping—subreddits about e-cigarettes or general vaping culture, 2) Substance use—subreddits about other substance use behaviors, such as marijuana and alcohol use, 3) Cessation—subreddits about smoking and vaping-related cessation, and 4) Non-specific—all other subreddits. We chose these community categories based on taxonomies from past Reddit and vaping research and from preliminary analyses of the subreddits in the data set (Barker & Rohde, 2019). Two study authors (JAR & JL) independently coded each of the subreddits, and intercoder reliability (measured using the ReCal2 software) comparing the full set of results between the two coders was high (Krippendorff’s $\alpha=.89$) (Freelon, 2010).

Categorizing Opinion Leaders

We used a post-hoc sociometric approach to identify opinion leaders in the subreddit- and thread-level data sets. For each of the 261 subreddit networks, we calculated the number of network connections and individual Redditors’ degree centrality. Degree centrality refers to the number of connections that a specific node (i.e., a member in a subreddit) has with other nodes in the same network (Wasserman & Faust, 1994). Importantly, each subreddit network was constructed as a directed graph, meaning that connections between nodes are not inherently reciprocated. For example, Redditor B commenting on Redditor A’s post signifies a directed relationship between $B \rightarrow A$, but not the other way around (i.e., $A \rightarrow B$). Thus, a node’s degree centrality within a subreddit for the current study was further distinguished by its in-degree (responses to a Redditor’s submission) and out-degree (a Redditor commenting on a submission). In the previous example, Redditor A’s in-degree centrality would be 1 and Redditor B’s would be 0.

Drawing from literature, we signified a Redditor as an opinion leader if their in-degree centrality comprised at least 10% of the sum of a subreddit's total in-degree (Valente & Pumpuang, 2007). In addition, opinion leaders were only identified among subreddits with at least 50 unique Redditors captured in the subreddit-level data set. We implemented this restriction to avoid classifying Redditors as opinion leaders if they had only one or two in-degree connections in subreddit networks with small sizes. Finally, it should be noted that although Redditors can appear in multiple subreddits, their opinion leadership status across subreddits can vary. For example, a Redditor could be classified as an opinion leader in r/Vaping but not in r/juul.

Study Outcomes

Subreddit network composition.—For the subreddit-level data set, we identified both the number of nodes (i.e., unique Redditors) and edges (i.e., connected threads among posters and commenters) in each subreddit network that were exclusively linked to an opinion leader. We did not count Redditors commenting to their own submissions when calculating edge frequencies. We normalized these outcomes by taking the proportion of the subset of nodes and edges relative to the full network. We interpret this outcome as the impact that an opinion leader has on the composition of a subreddit community.

Number of thread commenters.—For the thread-level data set, we calculated the discrete number of unique Redditors who commented on a thread initiated by a vaping submission. We did not include submission Redditors who may have commented on their own threads in this calculation. We interpret this outcome as a proxy indicator of thread “reach.”

Data Analysis

We used descriptive statistics to characterize the subreddit- and thread-level data sets. We report numerical variables (e.g., number of Redditors in a subreddit network) in terms of mean and standard deviation, as well as median and interquartile range. We used non-parametric tests to account for modest heterogeneity in the distribution of these outcomes. For the subreddit network composition outcome, we computed Kruskal–Wallis one-way ANOVAs to evaluate median differences in the proportion of subreddit nodes and edges exclusively linked to opinion leaders grouped by the subreddit community category variable. We restricted this analysis to subreddits that had at least one opinion leader. We computed post-hoc pairwise comparisons of the outcomes among community categories using Wilcoxon rank sum tests with continuity correction and *p*-value adjustment (Benjamini, Hochberg, and Yekutieli method).

We used a negative binomial regression to evaluate multi-variable associations with the number of thread commenters outcome. We used a mixed effects approach to account for correlated random intercepts among Redditors who initiated multiple threads. Predictors in the model were whether the author who initiated the thread was an opinion leader in the subreddit that the thread was published in (dichotomous; not opinion leader=reference category) and the 4-level subreddit community category variable previously discussed (categorical; Non-specific=reference category). We exponentiated model coefficients to

represent rate ratios for interpretability. For exploratory analyses, we ran a second model that included a community category by opinion leader interaction term. All statistical analyses were computed using R (version 4.1.2).

Data Visualization

We visualized example subreddits by community category (i.e., Vaping, Substance use, Cessation, and Non-specific) using the NetworkX module in Python3 (version 3.9.0) (Hagberg, Schult, & Swart, 2008). Nodes (i.e., subreddit members) were colored red if they were opinion leaders in the subreddit or gray if they were not. Network edges with opinion leaders were also colored red (or gray otherwise). We chose the example networks based on approximate matches to the median number of subreddit members and opinion leaders for their respective subreddit community category. We also used the wordcloud2 package in R to generate four separate word plots based on the text of the Reddit submissions in the data set, one each for the four subreddit community categories. We removed both common stop words and the vaping-related keywords used to collect the submissions prior to generating the word plots. The top 50 words in each community category are displayed in the plots.

Results

Descriptive Characteristics of the Subreddit- and Thread-Level Data

Characteristics of the subreddit- and thread-level data sets are in Table 1. Ninety-seven (37%) subreddits were coded as Substance use communities, 40 (15%) were coded as Vaping communities, and 7 (3%) were coded as Cessation communities; the remaining 117 (45%) subreddits were coded in the Non-specific community category. The average number of subreddit submissions and comments was 32 ($SD = 94$) and 310 ($SD = 1,038$), respectively. Subreddits had, on average, 133 ($SD = 323$) unique users and 166 ($SD = 547$) network connections. There was approximately one ($Mean = .99$, $SD = 1.43$) opinion leader per subreddit. Popular words among submissions posted in Vaping communities were “coil,” “pod,” and “juice.” Top words among submissions posted in Substance use communities were “product,” “weed,” and “CBD.” Lastly, top words among submissions posted in Cessation communities were “quit,” “nicotine,” and “day.” Example visualizations of subreddit networks and community category word plots (including for subreddits coded as Non-specific) are in Figure 2.

For the thread-level data set, there were more than 6,000 (72%) unique authors who initiated a thread. Of those, 327 (4%) were categorized as opinion leaders in the subreddit the thread was published in. Most (79%) threads had at least one comment. Threads were posted in Substance use (41%), Vaping (35%), Cessation (5%), and Non-specific (19%) subreddit communities. Mean comments per thread was 5.26 ($SD = 23.13$). Mean response time for thread comments was 72.23 hours ($SD = 167.94$), and median response time was 14.82 hours.

Composition of Subreddit Networks with Opinion Leaders

Ninety-five subreddits contained opinion leaders. Of those, most were categorized as Substance use (53%) communities, followed by Non-specific (31%), and Vaping (15%)

communities. Only two subreddits in the Cessation community category contained opinion leaders. The median proportion of subreddit nodes exclusively linked to opinion leaders was higher among Non-specific communities (71%) than Vaping (25%) and Substance use (39%) communities (both $p < .05$; Figure 3). Similarly, the median proportion of subreddit edges linked to opinion leaders was higher among Non-specific communities (74%) than Vaping (34%) and Substance use (44%) communities (both $p < .05$). There were no differences in the median proportion of nodes and edges linked to opinion leaders between Vaping and Substance use communities. We did not include the Cessation subreddit community category in these analyses due to the low frequency of this category.

Opinion Leadership and Subreddit Community Effects on Number of Thread Commenters

The direct effects of opinion leadership and subreddit community category on the number of thread commenters are in Table 2, Model 1. The rate of commenters was higher among threads initiated by opinion leaders than non-opinion leaders (adjusted rate ratio [aRR] = 4.84). In addition, threads posted in Vaping (aRR = 1.64), Substance use (aRR = 1.92), and Cessation (aRR = 1.21) communities had higher rates of commenters than those posted in Non-specific communities.

Opinion leadership by community category interaction effects on the number of thread commenters are in Table 2, Model 2. Only six threads posted in Cessation communities were initiated by opinion leaders, which precludes us from examining a meaningful interaction effect with this categorical variable. Thus, we did not include threads posted in Cessation communities in Model 2. Compared to Non-specific communities, the rate of commenters was lower among both Vaping (aRR=.21) and Substance use (aRR=.24) communities when the thread was initiated by an opinion leader.

Discussion

Our social media network analyses sought to identify factors associated with vaping discussions on Reddit. Drawing from nearly 90,000 submissions and comments posted in May 2021, our findings showed a sizable vaping discourse occurring across 261 different subreddits communities. Among those subreddits, the composition of vaping discussion networks was less affected by opinion leaders in communities about vaping and substance use culture (e.g., r/electronic_cigarette, r/weed) compared to Non-specific communities (e.g., r/AskReddit). Lastly, we found that both opinion leaders and subreddit community categories were correlated with the number of commenters (i.e., “reach”) on Reddit threads about vaping. Importantly, these observations help characterize network-level patterns of vaping discourse on Reddit, which is a critical early step for understanding how public health interventions or messaging campaigns could target and disseminate content on this platform.

Within a diffusion of innovation paradigm, opinion leaders are often considered central agents in the exchange of information throughout a connected social network (Valente & Pumpuang, 2007). Findings from the current study partially support this notion. As we hypothesized in our first study aim, opinion leaders facilitated a large proportion of vaping discussions in networked communities; however, the extent of this proportion varied by

the type of subreddit community. For example, opinion leaders talking about vaping in Non-specific subreddits were attributable for more than 70% of both the number of Redditors in the discussion network and the proportion of interpersonal connections between Redditors. This finding is likely due to the low level of ego (i.e., Redditor) involvement with this topic among community members in these Non-specific subreddits. Briefly, ego involvement refers to, “. . . the extent to which individuals’ self-concept is connected with their position on a particular issue . . .” and comes from group identity literature (Lapinski & Rimal, 2005, pp. 136). In the context of the current study, Redditors from Non-specific subreddits, such as those about video games, may not be interested in substance use culture, meaning that posts about vaping in said communities would appear out of context. As such, it would make sense that a majority of discussions about vaping in Non-specific communities would occur through a select few opinion leaders.

By contrast, we observed only a marginal opinion leadership effect on the composition of Vaping and Substance use subreddits, suggesting that vaping dialogue in these communities is inclusive and far-reaching. This observation may be explained by Redditors’ proximal association and involvement with vaping culture when participating in these types of subreddits. For example, it is reasonable to assume that most Redditors who join the r/electronic_cigarette subreddit are interested in vaping. As such, these individuals would likely feel comfortable not only participating in community threads about this behavior, but also in initiating their own threads about the topic.

At the thread-level, we found that the number of commenters to an individual Reddit thread about vaping varied based on the type of subreddit community it was posted in. For instance, threads published in Vaping, Substance use, and Cessation communities typically had more commenters than those posted in Non-specific communities. This finding suggests that vaping-related subreddits generate a large amount of interpersonal communication (i.e., Redditor-Redditor discourse), which is again likely explained by ego involvement with vaping and substance use culture.

As hypothesized in our second study aim, we also found that threads posted by opinion leaders tended to have more commenters than those posted by non-opinion leaders. Notably, this effect was further modified when threads were published in subreddits coded in the Non-specific community category. This finding supports our opinion leadership and subreddit network composition results discussed above. One possible explanation for this observed interaction could have to do with thread novelty relative to the subreddit community that content was posted in. For instance, opinion leaders participating in Non-specific subreddits may be some of the only Redditors discussing vaping in those communities. Thus, it is possible that these threads generate a lot of comments given the non-sequitur topic. For instance, an opinion leader Redditor posting about vaping in a news-related subreddit such as r/news would likely attract a lot of attention from other Redditors curious about the topic. By contrast, the opinion leaders who participated in vaping-related subreddits may not have been posting content out of the ordinary as perceived by other community members, meaning the content would likely attract less engagement.

Our study extends much of the current vaping and social media content analysis literature by identifying patterns of vaping-related discourse across a popular social media platform while also highlighting the role that opinion leaders play in facilitating this discourse (Barker & Rohde, 2019; Chen et al., 2020; Dai & Hao, 2017; Ketonen & Malik, 2020; Lazard et al., 2016; Martinez, Hughes, Walsh-Buhi, & Tsou, 2018). This work also adds to diffusion of innovation literature by testing the concepts of opinion leadership and networked communities together, with our results showing that opinion leaders in Non-specific subreddits may be more central in facilitating vaping discourse to other community members than vaping-related communities. Additional work testing the relationship between these two concepts in other social media contexts is still warranted.

Importantly, our findings could help inform communication efforts targeting discussions about vaping on Reddit. For example, our results support the reasonable supposition that health communication messaging campaigns, if strategically disseminated across vaping-related communities on Reddit, will have a larger reach and exposure rate than those not targeting vaping communities. However, if the goal of such a campaign is to diffuse information across subreddit communities unrelated to vaping culture, researchers should aim to work with key individuals (such as subreddit moderators) when sharing these messages with community members. It should be noted, though, that any such messaging efforts should be pilot tested to ensure that exposure minimizes the potential for unintended consequences, such as elevating erroneous beliefs about the risks of vaping relative to combusted cigarettes which may dissuade current smokers from switching to lower risk vaping devices or quitting completely (Wright et al., 2021).

The network methods in the present study have noteworthy implications for future research directions, particularly in the context of intervention design. For instance, our findings suggest that explicit mentions of an emerging vaping topic on Reddit (e.g., the harms/benefits of synthetic nicotine in vaping devices) can be tracked, and that those data can be used to model discussion networks in real time. These networks could further be used to inform participant recruitment or intervention dissemination strategies. For example, researchers could use targeted methods to deliver educational information about synthetic nicotine on Reddit to only those who have either shared or have participated in conversations about the topic. This approach emphasizes participant relevance, which would likely increase intervention impact.

Our methods could also inform studies examining the role of opinion leaders in social media intervention contexts. For example, a study could track highly central actors within Reddit networks discussing vaping as a harm-free product and provide them with informational content about the accurate risks of vaping. Researchers could then test whether intervening on those individuals might result in network-level effects, such as changes to content being posted by other Redditors who interacted with the opinion leaders. These proposed study applications offer novel opportunities for advancing vaping prevention and control research, and would contribute much-needed empirical evidence about the use of social networks for informing health interventions (Shelton et al., 2019; Valente, 2012).

Notably, our study only characterized opinion leaders using in-degree network centrality scores among discussion networks with 50 or more Redditors. Theories such as diffusions of innovation do not restrict opinion leaders based on network size; however, this was a methodological choice we made to avoid over-classifying Redditors as opinion leaders in small discussion networks. Currently, there is little consensus on methods for identifying social media opinion leaders. Thus, we suggest future Reddit studies build on this work by investigating other opinion leader classification methods, such as by using different cutoff network sizes or by characterizing gate keepers and authorities (e.g., subreddit moderators) as opinion leaders.

Strengths of this study include our large sample of Reddit data and our application of computational methods to construct vaping discussion networks across different subreddit communities. A limitation is that we only collected and analyzed Reddit threads explicitly mentioning one of our vaping-related keywords. Thus, our approach to modeling subreddit networks does not completely capture the connections among individuals not explicitly posting submissions about vaping in the subreddits. It should also be noted that vaping culture is dynamic and future studies may need to refine or modify some of the keywords used in the current study prior to extracting data. Another limitation was the low frequency of opinion leaders in cessation communities, which precluded us from evaluating the relationship between these variables in our thread-level and subreddit-level analyses. Lastly, our work only analyzed a single month (May 2021) of Reddit threads in an observational study, meaning that our findings cannot be generalized to how vaping is discussed on the platform as a whole or at different months. While the goal of this study was to investigate vaping discussions in a small snapshot of time, future work should consider analyzing longitudinal data sets adjusting for the temporal characteristics of the Reddit threads. Studies may also wish to extend our methods to investigate discussions of vaping on other online platforms, such as Twitter or YouTube.

Conclusion

Data from social media platforms such as Reddit provide key insight into how discussions about vaping proliferate across a social media network. Findings from this study showed that communities and opinion leaders shaped both the composition and reach of vaping discussions among online networks. The methodological approach described here may support further research on the dissemination of vaping information throughout social media networks. In particular, further research will be needed to model communication networks on specific vaping-related topics, such as posts mentioning vaping misinformation. Future investigations should also consider using the methods described in this study as a foundation for designing interventions to communicate about vaping to these online groups.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Funding

JL was funded by the Cancer Prevention Fellowship from the National Cancer Institute – National Institutes of Health grant number 2T32CA057711–27.

Data availability statement

The data underlying this article will be shared on reasonable request to the corresponding author.

References

- Anderson KE (2015). Ask me anything: What is Reddit? *Library Hi Tech News*, 32(5), 8–11. doi:10.7282/T3D220BR
- Bamakan SMH, Nurgaliev I, & Qu Q. (2019). Opinion leader detection: A methodological review. *Expert Systems with Applications*, 115, 200–222. doi:10.1016/j.eswa.2018.07.069
- Barker JO, & Rohde JA (2019). Topic clustering of e-cigarette submissions among Reddit communities: A network perspective. *Health Education & Behavior*, 46(2_suppl), 59S–68S. doi:10.1177/1090198119863770
- Baumgartner J, Zannettou S, Keegan B, Squire M, & Blackburn J. (2020). The Pushshift Reddit Dataset. *Proceedings of the International AAAI Conference on Web and Social Media*. 14(1). 830–839. doi:10.1609/icwsm.v14i1.7347
- Berkman LF, Glass T, Brissette I, & Seeman TE (2000). From social integration to health: Durkheim in the new millennium. *Social Science & Medicine*, 51(6), 843–857. doi:10.1016/S0277-9536(00)00065-4 [PubMed: 10972429]
- Chen L, Lu X, Yuan J, Luo J, Luo J, Xie Z, & Li D. (2020). A social media study on the associations of flavored electronic cigarettes with health symptoms: Observational study. *Journal of Medical Internet Research*, 22(6), e17496. doi:10.2196/17496 [PubMed: 32568093]
- Chu KH, Majmundar A, Allem JP, Soto DW, Cruz TB, & Unger JB (2019). Tobacco use behaviors, attitudes, and demographic characteristics of tobacco opinion leaders and their followers: Twitter analysis. *Journal of Medical Internet Research*, 21(6), e12676. doi:10.2196/12676 [PubMed: 31165716]
- Coulson NS (2013). How do online patient support communities affect the experience of inflammatory bowel disease? An online survey. *Journal of the Royal Society of Medicine Short Reports*, 4(8), 204253331347800. doi:10.1177/2042533313478004
- Dai H, & Hao J. (2017). Mining social media data for opinion polarities about electronic cigarettes. *Tobacco Control*, 26(2), 175–180. doi:10.1136/tobaccocontrol-2015-052818 [PubMed: 26980151]
- Dewi FK, Yudhoatmojo SB, & Budi I. (2017). Identification of opinion leader on rumor spreading in online social network twitter using edge weighting and centrality measure weighting *International Conference on Digital Information Management (ICDIM)*. Fukuoka, Japan.
- Freelon D. (2010). ReCal: Intercoder reliability calculation as a web service. *International Journal of Internet Science*, 5(1), 20–33.
- Gaiha SM, Lin C, Lempert LK, & Halpern-Felsher B. (2022). Use patterns, flavors, brands, and ingredients of nonnicotine e-cigarettes among adolescents, young adults, and adults in the United States. *JAMA Network Open*, 5(5), e2216194–e2216194. doi:10.1001/jamanetworkopen.2022.16194
- Hagberg AA, Schult DA, & Swart PJ (2008). Exploring network structure, dynamics, and function using NetworkX. 7th Python in Science Conference (SciPy2008), Pasadena, CA.
- Ketonen V, & Malik A. (2020). Characterizing vaping posts on Instagram by using unsupervised machine learning. *International Journal of Medical Informatics*, 141, 104223–104229. doi:10.1016/j.ijmedinf.2020.104223 [PubMed: 32623330]
- Kilgo DK, Yoo J, Sinta V, Geise S, Suran M, & Johnson TJ (2016). Led it on Reddit: An exploratory study examining opinion leadership on Reddit. *First Monday*, 21(9). doi:10.5210/fm.v21i9.6429

- Kwon M, & Park E. (2020). Perceptions and sentiments about electronic cigarettes on social media platforms: Systematic review. *JMIR Public Health and Surveillance*, 6(1), e13673. doi:10.2196/13673 [PubMed: 31939747]
- Langley T, Bell-Williams R, Pattinson J, Britton J, & Bains M. (2019). 'I felt welcomed in like they're a little family in there, I felt like I was joining a team or something': Vape shop customers' experiences of e-cigarette use, vape shops and the vaping community. *International Journal of Environmental Research and Public Health*, 16(13), 2341. doi:10.3390/ijerph16132341 [PubMed: 31269741]
- Lapinski MK, & Rimal RN (2005). An explication of social norms. *Communication Theory*, 15(2), 127–147. doi:10.1111/j.1468-2885.2005.tb00329.x
- Lazard AJ, Saffer AJ, Wilcox GB, Chung AD, Mackert MS, & Bernhardt JM (2016). E-cigarette social media messages: A text mining analysis of marketing and consumer conversations on Twitter. *JMIR Public Health and Surveillance*, 2(2), e6551. doi:10.2196/publichealth.6551
- Lazarsfeld PF, Berelson B, & Gaudet H. (1948). *The people's choice*. New York, NY: Columbia University Press.
- Lucherini M. (2022). The affective economy of vaping: A qualitative analysis of responses to an online questionnaire. *Critical Public Health*, 32(5), 737–746. doi:10.1080/09581596.2021.1950638
- Majmundar A, Allem JP, Unger JB, & Cruz TB (2021). Vaping and COVID-19: Insights for public health and clinical care from Twitter. *International Journal of Environmental Research and Public Health*, 18(21), 11231. doi:10.3390/ijerph182111231 [PubMed: 34769751]
- Martinez LS, Hughes S, Walsh-Buhi ER, & Tsou MH (2018). "Okay, we get it. You vape": An analysis of geocoded content, context, and sentiment regarding e-cigarettes on Twitter. *Journal of Health Communication*, 23(6), 550–562. doi:10.1080/10810730.2018.1493057 [PubMed: 29979920]
- McCausland K, Maycock B, Leaver T, & Jancey J. (2019). The messages presented in electronic cigarette-related social media promotions and discussion: Scoping review. *Journal of Medical Internet Research*, 21(2), e11953. doi:10.2196/11953 [PubMed: 30720440]
- Park CS, & Kaye BK (2017). The tweet goes on: Interconnection of Twitter opinion leadership, network size, and civic engagement. *Computers in Human Behavior*, 69, 174–180. doi:10.1016/j.chb.2016.12.021
- Pearson JL, & Villanti AC (2020). It is past time to consider cannabis in vaping research. *Nicotine & Tobacco Research*, 22(5), 597–598. doi:10.1093/ntr/ntaa012 [PubMed: 31956918]
- Pew Research Center. (2021). Social media fact sheet. <https://www.pewresearch.org/internet/fact-sheet/social-media/>
- Proferes N, Jones N, Gilbert S, Fiesler C, & Zimmer M. (2021). Studying Reddit: A systematic overview of disciplines, approaches, methods, and ethics. *Social Media + Society*, 7(2), 1–14. doi:10.1177/20563051211019004
- Qian Y, Gui W, Ma F, & Dong Q. (2021). Exploring features of social support in a Chinese online smoking cessation community: A multidimensional content analysis of user interaction data. *Health Informatics Journal*, 27(2), 146045822110214. doi:10.1177/14604582211021472
- Rogers EM (1983). *Diffusion of innovations*. New York, NY: Free Press.
- Shelton RC, Lee M, Brotzman LE, Crookes DM, Jandorf L, Erwin D, & Gage-Bouchard EA (2019). Use of social network analysis in the development, dissemination, implementation, and sustainability of health behavior interventions for adults: A systematic review. *Social Science & Medicine*, 220, 81–101. doi:10.1016/j.socscimed.2018.10.013 [PubMed: 30412922]
- Sidani JE, Hoffman BL, Colditz JB, Melcher E, Taneja SB . . . Chu KH (2022). E-cigarette-related nicotine misinformation on social media. *Substance Use & Misuse*, 57(4), 588–594. doi:10.1080/10826084.2022.2026963 [PubMed: 35068338]
- Soule EK, Mayne S, Snipes W, Guy MC, Breland A, & Fagan P. (2020). Impacts of COVID-19 on electronic cigarette purchasing, use and related behaviors. *International Journal of Environmental Research and Public Health*, 17(18), 6762. doi:10.3390/ijerph17186762 [PubMed: 32948084]
- Suarez-Lledo V, & Alvarez-Galvez J. (2021). Prevalence of health misinformation on social media: Systematic review. *Journal of Medical Internet Research*, 23(1), e17187. doi:10.2196/17187 [PubMed: 33470931]

- Tokle R, & Pedersen W. (2019). “Cloud chasers” and “substitutes”: E-cigarettes, vaping subcultures and vaper identities. *Sociology of Health & Illness*, 41(5), 917–932. doi:10.1111/1467-9566.12854 [PubMed: 30677161]
- Valente TW (1996). Social network thresholds in the diffusion of innovations. *Social Networks*, 18(1), 69–89. doi:10.1016/0378-8733(95)00256-1
- Valente TW (2010). *Social networks and health: Models, methods, and applications*. Oxford University Press. doi:10.1093/acprof:oso/9780195301014.001.0001
- Valente TW (2012). Network interventions. *Science*, 337(6090), 49–53. doi:10.1126/science.1217330 [PubMed: 22767921]
- Valente TW, & Davis RL (1999). Accelerating the diffusion of innovations using opinion leaders. *The Annals of the American Academy of Political and Social Science*, 566(1), 55–67. doi:10.1177/000271629956600105
- Valente TW, & Pumpuang P. (2007). Identifying opinion leaders to promote behavior change. *Health Education & Behavior*, 34(6), 881–896. doi:10.1177/1090198106297855 [PubMed: 17602096]
- van Eenbergen MC, van de Poll-Franse LV, Kraemer E, Verberne S, & Mols F. (2018). Analysis of content shared in online cancer communities: Systematic review. *JMIR Cancer*, 4(1), e7926. doi:10.2196/cancer.7926
- Vassey J, Valente T, Barker J, Stanton C, Li D. . . . Unger JB (2022). E-cigarette brands and social media influencers on Instagram: A social network analysis. *Tobacco Control*, tobaccocontrol-2021–057053. doi:10.1136/tobaccocontrol-2021-057053
- Wasserman S, & Faust K. (1994). *Social network analysis: Methods and applications*. Cambridge University Press.
- Wellman B, & Leighton B. (1979). Networks, neighborhoods, and communities: Approaches to the study of the community question. *Urban Affairs Quarterly*, 14(3), 363–390. doi:10.1177/107808747901400305
- Wright C, Williams P, Elizarova O, Dahne J, Bian J, Zhao Y, & Tan ASL (2021). Effects of brief exposure to misinformation about e-cigarette harms on twitter: A randomised controlled experiment. *British Medical Journal Open*, 11(9), e045445. doi:10.1136/bmjopen-2020-045445

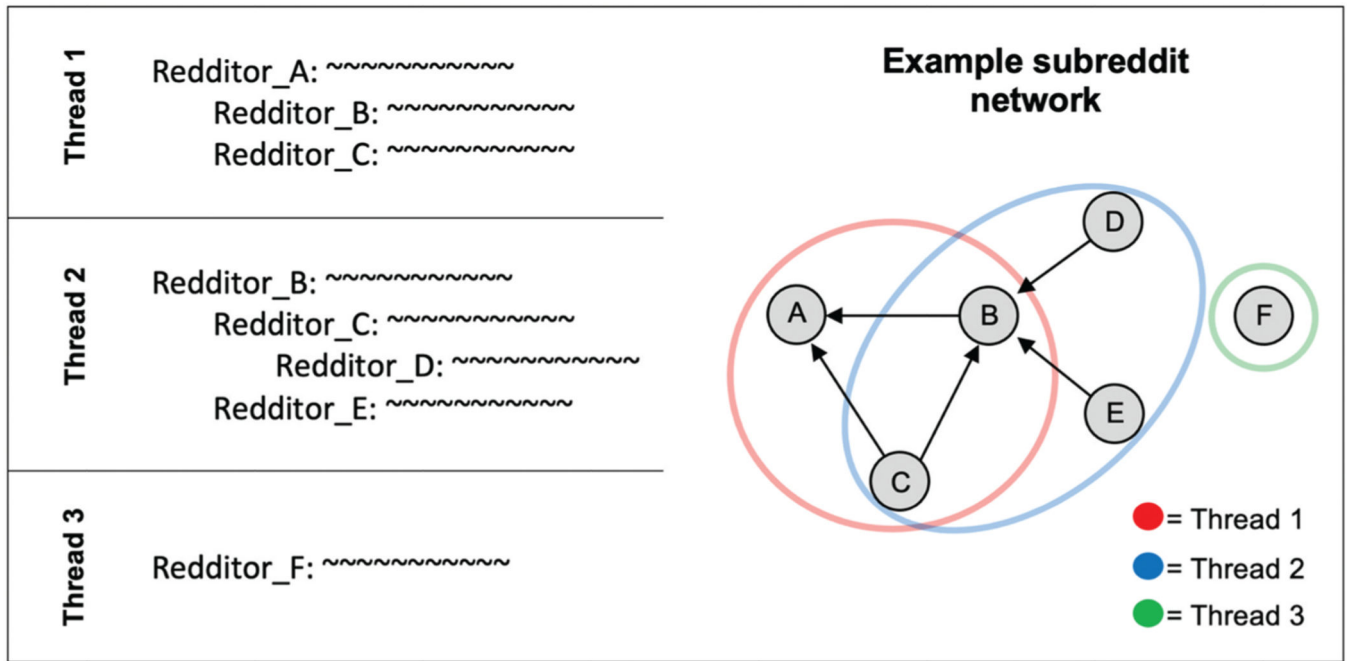


Figure 1.
Constructing an example thread- and subreddit-level data set.

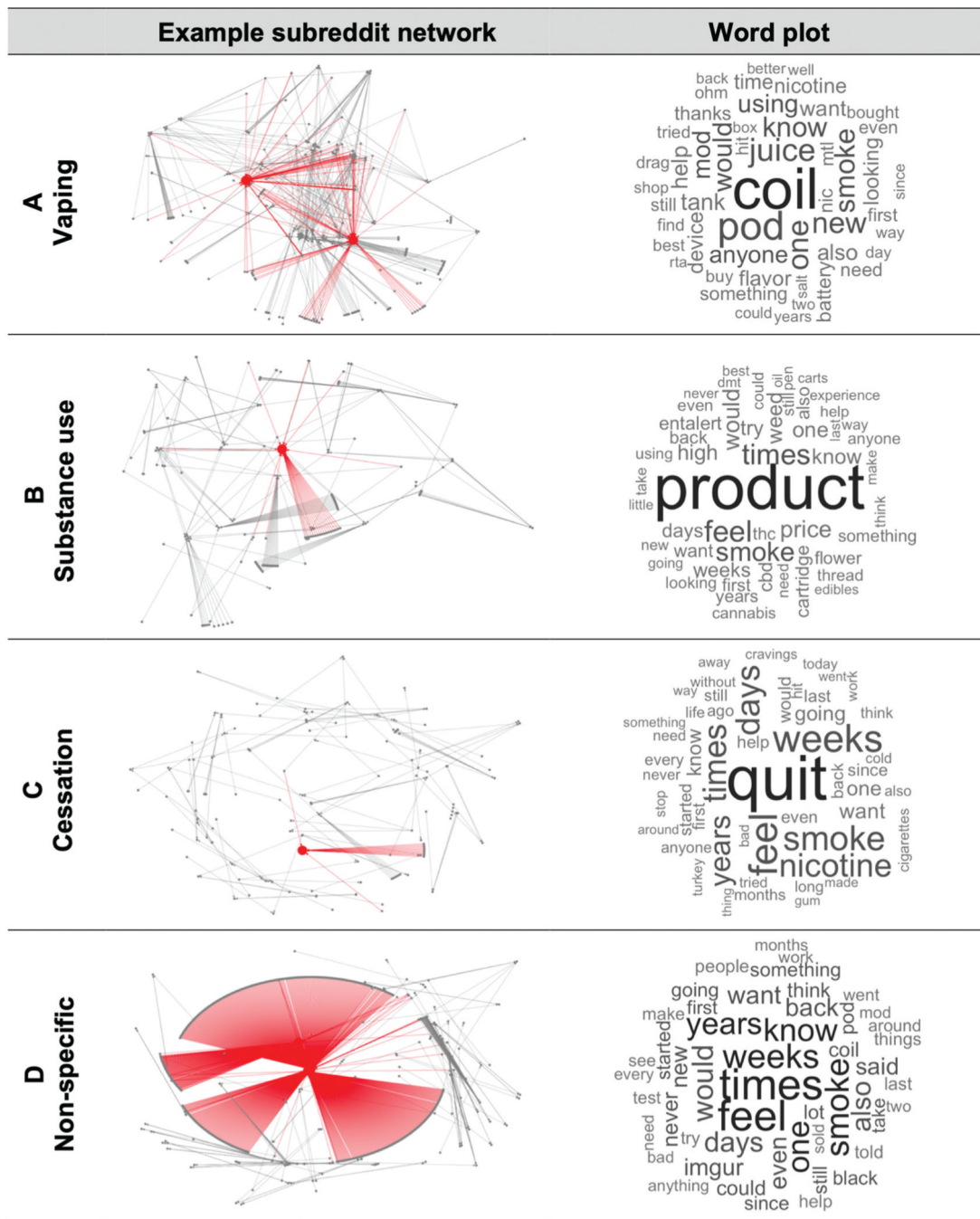


Figure 2.
Example networks and word plots corresponding to subreddit community categories.

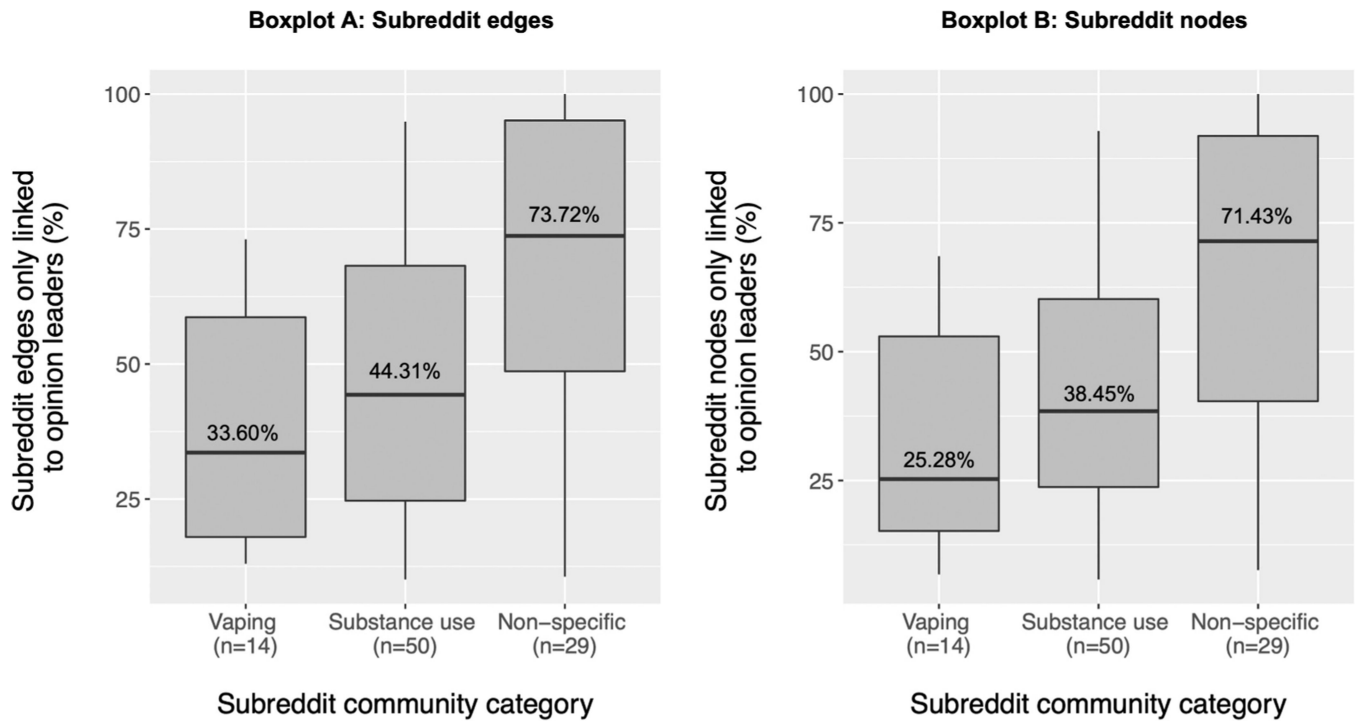


Figure 3. Median proportion of subreddit edges and nodes exclusively linked to opinion leaders by subreddit category.

Table 1.

Descriptive characteristics of the social media data

Variable	<i>n</i> (%) or <i>M</i> ± <i>SD</i>	Median (<i>IQR</i>)
Subreddit-level data set; <i>N</i> = 261		
Subreddits by community category		
Vaping	40 (15%)	–
Substance use	97 (37%)	–
Cessation	7 (3%)	–
Non-specific	117 (45%)	–
Number of submissions	32.09 ± 94.11	9 (17)
Number of comments	310.62 ± 1,038.76	53 (188)
Redditors	133.33 ± 323.35	37 (102)
Graph edges	166.64 ± 547.98	30 (112)
Number of opinion leaders in community	.99 ± 1.43	0 (1)
Thread-level data set; <i>N</i> = 8,377		
Unique authors	6,040 (72%)	–
Author is OL	327 (4%)	–
Submissions with comments	6,636 (79%)	–
Threads by community category		
Vaping	2,949 (35%)	–
Substance use	3418 (41%)	–
Cessation	415 (5%)	–
Non-specific	1,595 (19%)	–
Comments per submission	5.26 ± 23.13	3 (5)
Mean response time (hours)	72.23 ± 167.94	14.82 (38.97)

n=sample size; OL=opinion leader; IQR=interquartile range; *M*=mean; *SD*=standard deviation.

Table 2.

Correlates of the number of thread commenters

	Model 1			Model 2		
	aRR	95% CI	se	aRR	95% CI	se
Author is OL	4.84	4.28–5.48	.06	13.84	10.85–17.65	.04
<i>Subreddit category</i>						
Non-specific	REF	REF	REF	REF	REF	REF
Vaping	1.64	1.51–1.77	.04	1.80	1.66–1.95	.04
Substance use	1.92	1.78–2.07	.04	2.14	1.97–2.32	.04
Cessation	1.21	1.06–1.39	.07	–	–	–
<i>Subreddit*OL interaction</i>						
OL*Non-specific	–	–	–	REF	REF	REF
OL*Vaping	–	–	–	.21	.14–.30	.19
OL*Substance use	–	–	–	.24	.18–.33	.15

aRR=adjusted rate ratio; CI=confidence interval; se=standard error; OL=opinion leader; REF=reference level.