Published in final edited form as:

Addict Behav. 2019 April ; 91: 222–226. doi:10.1016/j.addbeh.2018.11.015.

Strategies to find audience segments on Twitter for e-cigarette education campaigns

Kar-Hai Chu^{a,*}, Jon-Patrick Allem^b, Jennifer B. Unger^b, Tess Boley Cruz^b, Meleeka Akbarpour^b, and Matthew G. Kirkpatrick^b

^aCenter for Research on Media, Technology, and Health, University of Pittsburgh, 230 McKee Place, Suite 600, Pittsburgh, PA 15213, United States

^bDepartment of Preventive Medicine, University of Southern California, 2001 North Soto Street, 3rd Floor, Los Angeles, CA 90032, United States

Abstract

The development of public health education campaigns about tobacco products requires an understanding of specific audience segments including their views, intentions, use of media, perceived barriers, and benefits of change. For example, identifying and targeting individuals who express ambivalence about e-cigarette use on Twitter may be helpful in devising and focusing public health campaigns to reduce e-cigarette use. This study developed a novel analytic strategy using social network analysis to identify audience segments on Twitter based on positive, negative, and neutral e-cigarette sentiment. Using Twitter data collected from April 2015 to March 2016, we identified different sub-groups of users who retweeted about e-cigarettes, and measured each subgroup's clustering coefficient (CC), which describes how tightly people cluster together. Ten high CC and ten low CC groups were randomly selected; then 100 randomly selected tweets from each group were coded for e-cigarette sentiment (positive, negative, neutral). Results indicate that differences in e-cigarette sentiment are associated with clustering of Twitter network ties. Statistical analyses revealed that high CC groups were more likely to have strong e-cigarette sentiments, suggesting that tightly clustered groups may be "echo chambers" (i.e., like-minded people repeating the same messages). By contrast, low CC groups were more likely to have neutral sentiments, and had greater fluctuation in sentiment over time, suggesting that they may be more flexible in their opinions about e-cigarettes and may be particularly receptive to targeted public health campaigns. Informatics techniques such as determination of clusters using social network analysis can be useful in identifying audience segments for future public health campaigns.

Keywords

Twitter; Social network analysis; Tobacco regulatory science; E-cigarette; Sentiment analysis

*Corresponding author.: chuk@pitt.edu (K.-H. Chu).

Conflict of interest

On behalf of all authors, the corresponding author states that there is no conflict of interest.

1. Introduction

Recent evidence suggests that the use of e-cigarettes (a non-combustible tobacco product) is increasing (Syamlal, Jamal, King, & Mazurek, 2016) and may pose significant health risks (McConnell et al., 2016; Office of the Surgeon General, 2016), including an increased risk of future combustible cigarette use in adolescents (Barrington-Trimis et al., 2016; Leventhal et al., 2015; Primack, Soneji, Stoolmiller, Fine, & Sargent, 2015; Soneji et al., 2017). As social media use continues to increase (Greenwood, Perrin, & Duggan, 2016), several stakeholders (such as the tobacco industry, public health agencies, and users of tobacco products) have used a range of social media platforms to disseminate information about these products. For example, the tobacco industry, including combustible cigarette and e-cigarette brands, has taken advantage of marketing on Twitter, Instagram and other platforms (Allem et al., 2018; Allem, Escobedo, Chu, Cruz, & Unger, 2017; Centers for Disease Control and Prevention, 2016; Chu et al., 2015; Chu, Allem, Cruz, & Unger, 2017). Additionally, recent public health campaigns have used Twitter to spread messages that stress the dangers of nicotine addiction. On the other hand, anti-tobacco messages that are widely distributed can elicit negative reactions in viewers with strong opinions (such as current e-cigarette users and independent vendors), leading to counter messaging (Allem et al., 2016; Harris et al., 2014). Overall, there is a need for the public health community to develop strategies to effectively disseminate evidence-based information regarding e-cigarettes to the populations they serve.

Social media platforms can be used to send tailored health education messages about tobacco to groups that would benefit from these messages (Thackeray, Neiger, & Keller, 2012). This kind of targeted approach requires identification of audience segments that hold similar views and readiness to change. Social media is an arena in which users may encounter messages that can influence their behaviors and attitudes. The Pew Research Center recently found that approximately 20% of social media users changed their minds about an issue because of something they saw on social media (Duggan & Smith, 2016). Further, sending and receiving pro-smoking social media messages is positively associated with “offline” smoking intentions and attitudes in college students (Depue, Southwell, Betzner, & Walsh, 2015; Yoo, Yang, & Cho, 2016). If public health researchers could identify different audience segments on social media, then it is possible they could devise targeted public health campaigns for that specific population. Individuals who are already hold strongly anti-tobacco opinions are unlikely to need convincing, and individuals who already hold strongly pro-tobacco opinions are unlikely to be convinced by a social media campaign. Therefore, it is possible that health communication campaigns would be most effective in changing opinions (and also cost-efficient), if they focused on individuals in the middle (i.e., those that hold – or at are exposed to – neutral opinions, ambivalent opinions, or no opinions at all). This communication strategy would be consistent with recent evidence indicating that providing information about anti-tobacco social norms decreases tobacco attitudes for emerging adults exposed to ambivalent tobacco messaging (Hohman, Crano, & Niedbala, 2016). However, to our knowledge, there is a noticeable gap in the literature on methods to segment audiences in social media for public health.

In the current study, we examine a novel method of identifying segments of Twitter users based on their sentiment toward e-cigarettes (i.e., positive, negative, or neutral opinions about e-cigarettes). The majority of extant Twitter studies (including those described above) have operationalized constructs such as dissemination of messages and influence of Twitter users with metrics available from Twitter such as hashtag usage (Rattananitnont, Toyoda, & Kitsuregawa, 2012), person tagging, i.e., being mentioned by another person in a tweet (Cha, Haddadi, Benevenuto, & Gummadi, 2010), or retweets, i.e., forwarding tweets to the user's network of followers (Kupavskii et al., 2012). The number of followers or retweets might represent popularity, and a health campaign could target popular users to spread a particular message. However, relying entirely on Twitter metrics to study human behavior could be limiting as these metrics (e.g., retweets) have no basis outside of Twitter.

Here we present a novel, alternative method applying concepts and metrics from social network analysis to further understand audience segmentation on Twitter. We will examine sub-groups of Twitter users to study how clustering may be associated with positive, negative, and neutral sentiment toward e-cigarettes. Twitter tends to contain “echo chambers,” where like-minded people repeat the same messages back and forth to each other (Barberá, Jost, Nagler, Tucker, & Bonneau, 2015). As people are more likely to choose to follow friends in networks that are similar to one another (De Choudhury, 2011), we hypothesize that tightly connected groups will have a higher proportion of positive or negative sentiment than loosely connected groups; in other words, echo chambers will have a lot of one-sided sentiment. Additionally, we hypothesize that loosely connected groups will have greater fluctuation in sentiment over time.

2. Methods

2.1. Social network analysis

Social network analysis (SNA) provides many tools to understand how people or organizations are connected, find hidden structures within these interconnections, and identify potentially important actors in a network. It is a combination of theories, methods, and measurements that can be used to study social structure created by relationships between people (Wasserman & Faust, 1994), and has been applied to identify actor roles in various situations that can help advance new ideas, e.g., in the diffusion of innovations (Rogers, 2003).

Of particular relevance to the current study is an important area in SNA that may be useful in identifying audience segments: the dynamics of clustering. We use two metrics in our analysis: 1) modularity, which describes how well a network can be divided into smaller clusters, or modules, and is useful in finding community structure (Newman & Girvan, 2004). For example, a primary school divided into sub-groups of grades would have a high modularity, as students are more likely to be connected with others in their grades; 2) clustering coefficient (CC), which measures how nodes (i.e., people in a social network) tend to cluster together. A simplistic way to describe CC is to explore how many of one's friends are friends with each other. With respect to our hypotheses, we operationalize “tightly connected groups” as those with high CC.

2.2. Data source

In the first phase of this study, we analyzed 367 k retweets (based on 60 k original tweets) collected over one month (October 2015) using two search terms relevant to e-cigarettes (“vape” and “ecig”). These terms were determined based on frequency and relevance of use after exploration of over 100 search terms in Twitter messages (see (De Choudhury, 2011) for a full description). Each retweet is labeled with the Twitter name of the user who retweeted the message and the user who originally posted the message. This information was used to construct the retweet network (i.e., the social network structure of individual users, where connections between users are defined by retweets of messages from one user to another). This social network structure forms the basis of quantifying distinct clusters.

There were 251 k individuals and 297 k retweet connections between individuals in the analyzed network. A modularity detection algorithm (Newman, 2006) identified 580 distinct modules, or clusters, within the network. Each individual can only belong in one cluster. For each cluster we calculated the clustering coefficient (CC range = 0 to 1; higher value denotes a more connected cluster, i.e., has more within-cluster retweets). Groups with the 10 highest clustering coefficients (high CC) and those with the 10 lowest clustering coefficients (low CC) were selected for coding of e-cigarette sentiment in order to find groups that were distinctly different.

2.3. Initial sentiment coding

For each high CC and low CC group, 100 retweets were randomly sampled, providing a total of 2000 retweets for sentiment coding. We focused exclusively on retweets during the initial coding stage in order to construct a retweet network for cluster analysis. Two trained coders explored a separate sample of data and decided on rules for determining whether retweets were positive, negative, or neutral toward e-cigarette use, or irrelevant. For each retweet, coders determined the sentiment using both the text and any referenced images or hyperlinks. Rules for coding included first identifying the theme of the tweet (in this dataset, relevant tweets included themes of taste/sensation, health benefits/concerns, social facilitation/consequences, evaluation of specific products or vaping behavior in general, and regulatory policy) and then coding the sentiment as positive, negative, or neutral based on an overall impression. Retweets were coded as irrelevant if they were non-English or did not relate to e-cigarettes or vaping (e.g., tweets about musical groups that have ‘Vape’ in the band name). After the rules were established, two coders coded all the retweets and then discussed and resolved all discrepancies. In this dataset, paraphrased examples of retweets coded for sentiment included: 1) “Those are AWESOME ecigs!” [positive]; 2) “I would never want to date anyone who used e-cigarettes” [negative]; and 3) “Science fiction could have never predicted vape pens” [neutral]. Inter-coder agreement for coding retweets was acceptable ($\alpha = 0.81$). Three high CC groups and one low CC group had predominately irrelevant retweets (i.e., > 60%). Thus, these groups were removed from the analysis. Additionally, we removed irrelevant tweets from the remaining groups, leaving a total of 1534 coded retweets for analysis.

2.4. Follow-up coding

We conducted follow-up coding to validate and expand on the initial coding results. The initial coding filtered an initial population of 251 k individuals in 580 clusters down to 16 clusters, allowing us to narrow our focus to the individual level. This approach also makes it feasible to extend the coding methodology to a longer time period of data collection for individual users in each CC group, and allows for analysis of change in sentiment over time. We included all tweets posted by any user from the initial coding for a full year, from April 2015 to March 2016 (excluding the initial month that codes were originally based on). Using the initial October 2015 period as the midpoint, this expanded period minimizes the time difference between the datasets used for initial and follow-up coding. The follow-up analysis was no longer restricted to retweets, which were necessary to create the network connections for the initial clustering. These tweets were randomly sampled from users in each of the 16 groups (9 low CC, 7 high CC) used in the initial coding according to the following two rules: 1) at least 100 tweets must be randomly sampled for each CC group; and 2) if, after retrieving 100 tweets, the total number of users who posted the tweets for the sample was < 10 , then we would continue adding tweets until we had reached 10 users; we were concerned that some groups were not adequately represented, so this added restriction prevented any one group's sentiment from being defined by a small number of users. Using this procedure, we coded a total of 2363 tweets as positive, negative, neutral, or irrelevant using the same coding system described above; inter-coder agreement was acceptable ($\alpha = 0.76$). After removing irrelevant tweets, the sample for analysis was a total of ($N = 1937$) tweets.

2.5. Statistical analysis

To assess the consistency of the two coding procedures, Pearson correlations testing the relationship between initial coding and follow-up coding were conducted.

Primary statistical procedures to test the association between CC group and sentiment were the same for both the initial coding (one month of retweets) and the follow-up coding (one-year of all tweets). First, for each coding set we further categorized the high CC and low CC groups by “strength” of sentiment. Groups that contained e-cigarette-related tweets that were predominately negative or predominately positive (i.e., the proportion of negative [or positive] tweets was greater than the proportion of positive [or negative] tweets and greater than the proportion of neutral tweets) were categorized as Strong Sentiment. Groups that contained predominately neutral were categorized as Neutral. Then for each coding set, we conducted an $N-1$ chi-square test – which is appropriate for small sample sizes with expected frequencies greater than one but less than five (Campbell, 2007) – to examine the potential association between CC (High, Low) and e-cigarette sentiment (Strong, Neutral).

Then, in order to examine the potential change in sentiment over time as a function of CC group, we conducted mixed linear models for each sentiment outcome (positive, negative, neutral) that included CC Group (Low, High) and Time (11 months) as fixed effect factors. Significant CC Group $\times$ Time interactions were followed by mixed linear models examining the average rate of change (i.e., the average of month-to-month differences in sentiment

proportions) with CC Group as a fixed effect. For all analyses, p values were considered statistically significant at < 0.05 .

3. Results

3.1. Consistency of coding procedures

Overall CC groups' proportions of positive, negative, and neutral retweets and tweets were consistent between the initial and follow-up coding [positive sentiment: $r(14) = 0.97$; $p < .001$; negative sentiment: $r(14) = 0.65$; $p = .006$; neutral sentiment: $r(14) = 0.74$; $p = .001$].

3.2. Association between CC group and sentiment (initial and follow-up coding)

Table 1 shows the proportion of positive, negative, and neutral retweets and tweets as a function of each CC group for both the initial and follow-up coding. For the initial coding, based on these proportions, two low CC groups and five high CC groups were categorized as Strong Sentiment (i.e., four groups had predominately negative sentiments and three groups had predominately positive). Further, seven low CC groups and two high CC groups were categorized as Neutral Sentiment. The chi-square analysis revealed a marginally significant association between CC and e-cigarette sentiment. High CC groups were more likely to have strong e-cigarette sentiments, and low CC groups were more likely to have neutral sentiments [$\chi^2(1) = 3.63$; $p = .057$].

Results for follow-up coding were similar to those for the initial coding. Two low CC groups and six high CC groups were categorized as Strong Sentiment (five groups had predominately negative sentiments and three groups had predominately positive). Seven of the low CC groups and one high CC group were categorized as Neutral Sentiment. The chi-square analysis revealed that high CC groups were significantly more likely to have strong e-cigarette sentiments, and low CC groups were more likely to have neutral sentiments [$\chi^2(1) = 5.95$; $p = .015$].

3.3. Change in sentiment over time (follow-up coding)

Fig. 1 shows the change in neutral (left panel), negative (middle panel), and positive (right panel) sentiment over time as a function of CC group. Overall, the Low CC group (Mean = 0.45; SE = 0.05) had a greater average proportion of neutral tweets compared to the High CC group [Mean = 0.15; SE = 0.06; Main effect of CC Group: $F(1,14) = 12.9$; $p = .003$]. There was also a significant CC Group $\times$ Time interaction [$F(10,129) = 2.38$; $p = .013$], indicating that the proportion of neutral tweets in the Low CC group fluctuated from month to month, while the proportion in the High CC group was relatively stable (Fig. 1, left panel). Follow-up analyses indicated that the average rate of month-to-month change for neutral tweets was greater for the Low CC group (Mean = 0.26; SE = 0.03) compared to the High CC group (Mean = 0.07; SE = 0.03; $F(1,14) = 25.6$; $p < .001$).

Additionally, the High CC group (Mean = 0.51; SE = 0.12) had a greater average proportion of positive tweets compared to the Low CC group [Mean = 0.15; SE = 0.10; Main effect of CC Group: $F(1,14) = 5.2$; $p = .039$]. However, there was no interaction with Time for positive tweets, and there were no significant main or interactive effects for negative tweets.

4. Discussion

Our findings suggest that SNA-identified groups on Twitter with high clustering coefficients are strongly opinionated about e-cigarettes. In both the initial coding time period (one month of retweets) and the follow up (one-year of all tweets), the high CC groups had either more positive or more negative tweets, and fewer neutral ones, when compared with low CC groups. Additionally, the low CC groups were much less likely to express strong positive views overall. Because neutral views were expressed by those more loosely clustered, they were less likely to be exposed to strong positions about e-cigarettes from others in their local clusters. Interestingly, neutral views in the loosely clustered groups were unstable over time, suggesting they may be less fixed in their views and thus may be better targets for e-cigarette-related health campaign messages (Hohman et al., 2016). While previous research suggests that e-cigarette discussions on Twitter are generally positive in the aggregate (Cole-Lewis et al., 2015), our results indicate that pro-and anti-cigarette sentiment are both well represented within different segments of users.

Twitter studies have tracked hashtag usage (Rattananitnont et al., 2012) and retweet counts (Kupavskii et al., 2012) to study information diffusion, and while these metrics document Twitter-specific behavior, they may not generalize to real-world behavior. Additionally, these studies tend to study Twitter users as a single group and ignore differences in opinion. We do not discount these metrics, but believe combining them with SNA and sentiment analysis in our study offers an improved approach to target specific people and groups through social media. By understanding nuances in social media that includes homo-philous behavior such as echo chambers along with audience segmentation, we may be able to better address generalizability under a framework shown to be predictive of actual human behavior, i.e. social network analysis (Ennett & Bauman, 1994).

Our findings show that organically formed clusters based on information sharing (i.e., retweets) can be identified via known metrics, and then associated with user opinions (i.e., sentiment). In defining the different groups, public health officials can take a targeted approach: (1) For tightly connected pro-e-cigarette groups, it is unlikely that a generic campaign can change their minds. In fact, anti-e-cigarette messages could potentially be used against the campaign (Allem et al., 2016). It could be worthwhile to identify the types of profiles discussing pro-e-cigarette content in order to better address specific audiences and topics; (2) For tightly connected anti-e-cigarette groups, there is less need to promote ads to this population. However, campaigns could still advocate for these people to take active measures in spreading messages or take a “door-to-door” approach; (3) For the loosely connected groups with moderate and/or fluctuating opinions, it is here where people might be more open to persuasion through traditional anti-e-cigarette messages.

The current data provide support for the notion that health communication planners may benefit from partnering with specialists in data analytics to identify clustering characteristics. To hopefully facilitate this type of collaboration and to make our SNA approach to social media easy to implement, our long-term goal is to develop a custom software package that follows the current approach and that will be accessible to the larger research community. The open-source development platform R (R Development Core Team,

2015) allows researchers to create self-contained packages that anyone can download and use in R. SNA researchers have built packages using R that can help identify opinion leaders (Jacobs, Khanna, Madduri, & Bader, 2015), study diffusion (Vega Yon, Pitts, Hayes, & Valente, 2016), or stochastic actor modeling (Ripley, Boitmanis, & Snijders, 2013). The current study is the first step in developing a similar software package that can be used to identify population segments on Twitter.

There are several limitations to this study to consider. First, there might exist bias in the data that stems from automated (e.g., robot) accounts, leading to overrepresented subpopulations (Allem & Ferrara, 2016; Clark et al., 2015). Second, we did not know the particulars about users in each group (e.g., current e-cigarette users or not). While analysis of user profiles was outside the scope of the current study, this is an interesting avenue for future research, as data about individual Twitter users could help elucidate who is clustering and whether clustering about one topic (such as e-cigarettes) is related to clustering on other topics. Last, we did not know if these results were generalizable to the offline world, as we were limited by the accessibility of Twitter data, where only public data is available. Similarly, we chose very specific keywords for our data collection process. More, or different, search terms might yield different results but the keywords in this study were informed by prior research (Chu, Allem, Cruz, & Unger, 2016).

5. Conclusion

This study demonstrated SNA and sentiment coding to identify and delineate segments of Twitter users discussing e-cigarettes. Our results revealed that groups with tight connectedness were more likely to have strong sentiments (i.e., either positive or negative) compared to those who were loosely networked. The discovery of these clusters could be very important for health communication strategies as audience segmentation is a cornerstone of effective, targeted social marketing (Thackeray et al., 2012).

Acknowledgments

Funding for this research was supported by grant number P50CA180905 from the National Cancer Institute and FDA Center for Tobacco Products.

References

- Allem J-P, Cruz TB, Unger JB, Toruno R, Herrera J, & Kirkpatrick MG (2018 7). Return of cartoon to market e-cigarette-related products. *Tob. Control* p. tobaccocontrol-2018-054437.
- Allem J-P, Escobedo P, Chu K-H, Cruz TB, & Unger JB (2017). Images of little cigars and cigarillos on instagram identified by the Hashtag #Swisher: Thematic analysis. *Journal of Medical Internet Research*, 19(7).
- Allem J-P, Escobedo P, Chu K-H, Soto DW, Cruz TB, & Unger JB (2016). Campaigns and counter campaigns: reactions on Twitter to e-cigarette education. *Tob. Control* p. tobaccocontrol-2015-052757.
- Allem J-P, & Ferrara E (Aug. 2016). The importance of debiasing social media data to better understand e-cigarette-related attitudes and behaviors. *Journal of Medical Internet Research*, 18(8), e219. [PubMed: 27507563]

- Barberá P, Jost JT, Nagler J, Tucker JA, & Bonneau R (Oct. 2015). Tweeting from left to right: Is online political communication more than an echo chamber? *Psychological Science*, 26(10), 1531–1542. [PubMed: 26297377]
- Barrington-Trimis JL, et al. (2016). E-cigarettes and future cigarette use. *Pediatrics*, 138(1).
- Campbell I (2007 8). Chi-squared and Fisher–Irwin tests of two-by-two tables with small sample recommendations. *Statistics in Medicine*, 26(19), 3661–3675. [PubMed: 17315184]
- Centers for Disease Control and Prevention (2016). E-cigarette ads and youth VitalSigns, CDC.
- Cha M, Haddadi H, Benevenuto F, & Gummadi KP (2010). Measuring user influence in Twitter: The million follower fallacy.
- Chu K-H, Allem J-P, Cruz TB, & Unger JB (Sep. 2016). Vaping on Instagram: cloud chasing, hand checks and product placement. *Tob. Control* p. tobaccocontrol-2016–053052.
- Chu K-H, Allem J-P, Cruz TB, & Unger JB (2017). Vaping on instagram: Cloud chasing, hand checks and product placement. *Tobacco Control*, 26(5).
- Chu K-H, et al. (Dec. 2015). Diffusion of messages from an electronic cigarette brand to potential users through Twitter. *PLoS One*, 10(12).
- Clark EM, et al. (2015 8). Vaporous marketing: Uncovering pervasive electronic cigarette advertisements on Twitter.
- Cole-Lewis H, Varghese A, Sanders A, Schwarz M, Pugatch J, & Augustson E (2015 8). Assessing electronic cigarette-related tweets for sentiment and content using supervised machine learning. *Journal of Medical Internet Research*, 17(8), e208. [PubMed: 26307512]
- De Choudhury M (2011). Tie formation on twitter: homophily and structure of egocentric networks. 2011 IEEE Third Int'l Conference on Privacy, Security, Risk and Trust and 2011 IEEE Third Int'l Conference on Social Computing (pp. 465–470)..
- Depue JB, Southwell BG, Betzner AE, & Walsh BM (Mar. 2015). Encoded exposure to tobacco use in social media predicts subsequent smoking behavior. *American Journal of Health Promotion*, 29(4), 259–261. [PubMed: 24670071]
- Duggan M, & Smith A (2016). Americans, politics and social media. Pew Research Center6.
- Ennett ST, & Bauman KE (1994). The contribution of influence and selection to adolescent peer group homogeneity: The case of adolescent cigarette smoking. *Journal of Personality and Social Psychology*, 67(4), 653–663. [PubMed: 7965611]
- Greenwood S, Perrin A, & Duggan M (2016). Demographics of social media users. Pew Research Center2.
- Harris JK, Moreland-Russell S, Choucair B, Mansour R, Staub M, & Simmons K (Oct. 2014). Tweeting for and against public health policy: Response to the Chicago department of public health's electronic cigarette Twitter campaign. *Journal of Medical Internet Research*, 16(10).
- Hohman ZP, Crano WD, & Niedbala EM (Mar. 2016). Attitude ambivalence, social norms, and behavioral intentions: Developing effective antitobacco persuasive communications. *Psychology of Addictive Behaviors*, 30(2), 209–219. [PubMed: 26460476]
- Jacobs S, Khanna A, Madduri K, & Bader D (2015). Software tools to quantify structural importance of nodes in a network.
- Kupavskii A, et al. (2012). Prediction of retweet cascade size over time. 2335–2338.
- Leventhal AM, et al. (Aug. 2015). Association of electronic cigarette use with initiation of combustible tobacco product smoking in early adolescence. *JAMA*, 314(7), 700. [PubMed: 26284721]
- McConnell R, et al. (2016). Electronic-cigarette use and respiratory symptoms in adolescents. *American Journal of Respiratory and Critical Care Medicine*, 195, 1043–1049.
- Newman ME (2006). Modularity and community structure in networks. *Proceedings of the National Academy of Sciences*, 103, 8577–8582.
- Newman ME, & Girvan M (2004). Finding and evaluating community structure in networks. *Physical Review E*, 69.
- Office of the Surgeon General (2016). E-cigarette use among youth and young adults: A report of the surgeon general, Washington, DC.

- Primack BA, Soneji S, Stoolmiller M, Fine MJ, & Sargent JD (Nov. 2015). Progression to traditional cigarette smoking after electronic cigarette use among US adolescents and young adults. *JAMA Pediatrics*, 169(11), 1018. [PubMed: 26348249]
- R Development Core Team (2015). R: the R project for statistical computing (Online). Available: <https://www.r-project.org>.
- Rattananitnont G, Toyoda M, & Kitsuregawa M (2012). Characterizing topic-specific hashtag cascade in twitter based on distributions of user influence. *Berlin Heidelberg: Springer* 735–742.
- Ripley R, Boitmanis K, & Snijders T (2013). CRAN - package RSiena (Online). Available: <https://cran.r-project.org/web/packages/RSiena/>.
- Rogers E (2003). *Diffusion of innovations* (5th Edition). New York: Free Press.
- Soneji S, et al. (Aug. 2017). Association between initial use of e-cigarettes and subsequent cigarette smoking among adolescents and young adults. *JAMA Pediatrics*, 171(8), 788. [PubMed: 28654986]
- Syamlal G, Jamal A, King BA, & Mazurek JM (2016 6). Electronic cigarette use among working adults — United States, 2014. *MMWR. Morbidity and Mortality Weekly Report*, 65(22), 557–561. [PubMed: 27281058]
- Thackeray R, Neiger BL, & Keller H (Mar. 2012). Integrating social media and social marketing: a four-step process. *Health Promotion Practice*, 13(2), 165–168. [PubMed: 22382492]
- Vega Yon G, Pitts S, Hayes T, & Valente T (2016). netdiffuseR: Analysis of diffusion and contagion processes on networks (Online). Available: <https://cran.r-project.org/package=netdiffuseR>.
- Wasserman S, & Faust K (1994). *Social network analysis: Methods and applications (structural analysis in the social sciences)*. Cambridge University Press.
- Yoo W, Yang J, & Cho E (2016). How social media influence college students' smoking attitudes and intentions. *Computers in Human Behavior*, 64, 173–182. [PubMed: 27956757]

HIGHLIGHTS

- Audience delineation improve e-cigs public health campaigns.
- Network analysis can support audience segmentation on social media.
- Tightly clustered groups are more likely to have strong sentiment.
- Public health officials can leverage these methods for targeted health messaging.

Fig. 1. Mean (± 1 SEM) proportion of neutral (left panel), negative (middle panel), and positive (right panel) sentiment over time (April 2015–March 2016) as a function of CC group (low, high). CC = clustering coefficient.

Table 1

E-cigarette sentiment (proportion of tweets) by CC group (initial coding and follow-up coding).

Initial Coding				Follow-up Coding				
	N	Positive	Negative	Neutral	N	Positive	Negative	Neutral
Low CC (group ID)								
20	100	0.03	0.78	0.19	94	0.11	0.35	0.54
38	99	0.05	0.67	0.28	93	0.20	0.33	0.46
11	98	0.03	0.07	0.90	97	0.22	0.38	0.40
23	98	0.02	0.05	0.93	95	0.12	0.31	0.58
28	100	0.03	0.03	0.94	98	0.11	0.61	0.28
21	99	0.03	0.02	0.95	92	0.07	0.42	0.51
29	96	0.02	0.02	0.96	95	0.27	0.29	0.43
48	99	0.03	0.01	0.96	100	0.13	0.32	0.55
56	94	0.01	0.01	0.98	105	0.10	0.52	0.37
High CC (group ID)								
98	100	0.99	0.01	0.00	108	1.00	0.00	0.00
145	99	1.00	0.00	0.00	92	1.00	0.00	0.00
176	100	0.00	1.00	0.00	104	0.00	1.00	0.00
70	100	0.11	0.83	0.06	200	0.10	0.87	0.04
17	78	0.91	0.01	0.08	380	1.00	0.00	0.00
42	89	0.18	0.40	0.42	100	0.13	0.46	0.41
51	85	0.05	0.08	0.87	84	0.36	0.19	0.45

Note: Groups are presented by Neutral sentiment for the initial coding (ascending order). Bolded proportions indicate the individual groups that were categorized as Strong Sentiment. N = 1534 total tweets for initial coding (data collected in October 2015), N = 1937 total tweets for follow-upcoding (data collected from April 2015 to March 2016, excluding October 2015).