# An Empirical Analysis of Tobacco-Oriented Social Networks to Characterize Social Interaction

**Pankaj Prasad Dwivedi[1], Dilip Kumar Sharma[*2]**
Jaypee University of Engineering and Technology
A.B. Road, Raghogarh, Dist. Guna – M.P. (India) – 473226,
E-mail: pankaj.career9@gmail.com[1]
E-mail: dilipsharmajiet@gmail.com[2]

**Abstract**

Tobacco "wars" are finding a new home on social media. This type of media, assessing the present situation is critical for tobacco control activism. This research defines user engagement and social influence using information-theoretic techniques and social network analysis to demonstrate the influence of tobacco-related user-generated material. Our empirical investigations show that the explosion of pro-tobacco material has long-term impacts with greater influence and more active users, as well as in the tobacco control across the world, the absence of social media tools. User involvement in the pro-tobacco community seems to be more engaged, and tobacco marketing content created by users is more successful in capturing the attention of the user. In addition, we construct an investigation and three tobacco-related social networks with their structural features. We find that the pro-tobacco systems exceed someone else in size, implying that a large number of people are exposed to pro-tobacco information. Such findings suggest that the gap between tobacco control and tobacco promotion is growing, and that tobacco control may be going in the wrong direction via social media towards tobacco promotion.

**Keywords: -** Tobacco Communities, Entropy, Tobacco Addiction, Smoking Addiction, Social Networks, Network Topology.

**Tob Regul Sci.™ 2022; 8(1): 511-521**
**DOI: doi.org/10.18001/TRS.8.1.45**

## Introduction

Tobacco usage is one of the world's most significant public health dangers, killing approximately 6,000,000 human beings every year, with tobacco use directly responsible for more than 5,000,000 of those deaths1. Tobacco smoking has been associated with the development of several major cardiovascular disease (CVD), like illnesses, respiratory diseases, including cancer, heart disease, stroke, and hypertension [2]. Out of the 4,800 chemicals in tobacco products, [61] have been proven to cause cancer, including approximately 90% of lung cancers, as well as increases the risk of at least 13 other cancers, including those of the mouth, pancreas, kidneys, bladder, larynx, esophagus, pharynx, and uterine cervix [3,5]. According to studies [6], smoking is the second largest cause of cardiovascular disease after hypertension, causing almost 10% of CVD. When Chronic Obstructive Pulmonary Disease (COPD), usually produced through smokes, progresses, breathing becomes more challenging. It has been discovered that smoking is responsible for 90% of COPD deaths and raises the chance of getting COPD in young tobacco users by decreasing the development and growth of the lungs. Tobacco usage is also connected to poor mental health in humans. Nicotine addiction has been linked to an increased risk of sadness and even suicide [7,8]. Tobacco use has long-lasting and far-reaching effects on humans, in addition to the effects on individuals. Social cites play a large role in propagating smoking behavior [9], as well as nicotine's transgenerational effect [10], provide significant obstacles to tobacco control.

Despite the fact that many smoke-free regulations are implemented, traditional tobacco control measures are finding it difficult to combat the worldwide nicotine pandemic in the era of social media. With more people, especially teenagers and young adults, accessing social networking platforms, cigarette businesses hope to gain tremendously from social media marketing. The promotion of tobacco products takes place on numerous social media sites, such as YouTube video clip [17–22], Weibo [11–16], and Facebook, as well as on smartphone apps ('ishisha' and 'Cigar Boss') [23,24]. The use of pro-tobacco material may influence young people to become regular smokers [25]. At the same time, tobacco control organizations use social media channels to emphasize that tobacco use increases the risk of disease and death, as well as to assist smokers in quitting [26,27]. In recent years, social media has proved to be a new battlefield for tobacco companies and tobacco control activists [28–30]. To assess the existing situation, potential hazards, and countermeasures to be taken in response to the tobacco "wars" in social media, it is critical to analyze user engagement with tobacco-related information in social media. An interaction analysis of user interactions can reveal tobacco product marketing in social media and assist in understanding tobacco product communications within unconventional communication venues [31]. It provides valuable information to the tobacco control community, allowing them to assess program performance, identify policy gaps, and make an informed choice [32].

However, existing studies have shown little detailed experiential research to explain user involvement in tobacco-oriented social media. Many studies have been done in recent years to reveal hidden patterns of smoking advertising activity and the widespread usage of tobacco company-sponsored video clips on social media [11,14,18,20,33, 34]. In doing so, also attempted to conduct an online survey by some researchers to clarify the impact of tobacco product consumption preferences [35]. Furthermore, the majority of extant studies are case studies with no quantitative methodologies. Because of the significant rise of tobacco-related information on social media, analysis methods and existing collection of information are not suitable for solving this problem.

To remedy this gap, we describe user engagement in cigarette groups using large-scale social media data in this research. We'll create three tobacco-related social networks in particular to demonstrate diverse patterns of user involvement in the smoking community. There are two main contributions to this paper. (1) It's the first time, that large-scale nicotine data has been gathered from social media and online to reveal patterns of tobacco interaction in the tobacco community, to our knowledge. (2) Based on a large data set, we quantitatively characterized the interaction patterns of group activities in the tobacco community and found some important results. These findings will aid us in better understanding tobacco community social interactions and making more knowledgeable tobacco control choices.

**Table 1**. Network metrics for tobacco-related social networks are compared.

| Network | Edges | Edges of GC | Nodes | Node of GC | $\|k\|$ | $C$ | $Q$ | $P_c$ |
|---|---|---|---|---|---|---|---|---|
| **Tobacco Promotion** | 658618 | 652418 | 508017 | 501881 | 2.5946 | 0.000023 | 0.35297 | 0.5 |
| **Tobacco Control** | 277447 | 264888 | 218074 | 207453 | 2.545762 | 0.000279 | 0.3327379 | 0.5 |
| **Tobacco Cessation** | 368171 | 362040 | 163498 | 158119 | 4.506034 | 0.01385576 | 0.5276513 | 0.25 |

**Outcomes.** We examine user interactions from three views in the tobacco community, including a Dynamical Patterns, user interaction based statistical Patterns, and a topological Patterns of the tobacco community.

**Dynamical Patterns.** Tobacco products have been more prominent on social media platforms like Instagram, Facebook, Twitter, and YouTube etc., which has boosted their usage. A large number of people are exposed to pro-tobacco user-generated material, which includes everything from product evaluations to smoking fetish pictures to situations using tobacco. We assess the effect of tobacco-related fan sites on Facebook to expose the consequences of such tobacco-related data to potential consumers.

In terms of social impact, [37] to demonstrate the effect of fan pages we apply the transfer entropy (TE) by using a time series of comments. To select an acceptable bin width, we must first assess the response time of comments. The time range is described as the variation in time between the timestamp of the comments and the start of the post. The pro-tobacco group received 90 percent of reader comments in 330 hours, whereas the anti-tobacco group received 197 hours and the quitting-tobacco group received 213 hours. The pro-tobacco group's wider range of time implies that user participation is more active in this group, and the material in this group has a longer lasting influence than in other groups. On the other hand, 95 percent of comments were made within 163 hours for the whole dataset, indicating that a post will receive.

The impact of fan pages is calculated in the same way. Tobacco advertising is carried out on 43 percent of the top 35 prominent fan pages. The anti-tobacco pages are dedicated to tobacco campaigns that employ a variety of tactics. 'Animals Smoking Durrys' and 'Girls Smoking' are two instances of fetish imagery (pictures of sexual fantasy situations, young men, cartoon characters, young women smoking, smoking animals, and so on) being used to advertise smoking as stylish, fashionable, or enjoyable. Tobacco retailers ('mrhookah' and 'bnbtobacco') and online tobacco shops ('hookah-shisha' and 'smoke free online'). However, for cigarette company campaigns, fan pages with names like 'Eco Dumas. lt' and 'Espinosa cigars' have been created. On the other side, social media is also used to promote tobacco control and cessation. Tobacco-free initiatives have spawned a slew of regional groups, like 'Tobacco Free Florida' and 'Tobacco Free California.' Additionally, certain cigarette cessation programs are available to assist smokers in quitting. 'Become An EX' is a well-known community dedicated to assisting smokers through facts, therapy, and sharing of experiences.

**User interaction based Statistical Patterns**- User interaction entropy (IE) compares user involvement we look at statistical patterns of webpage attributes in the three communities, such as comment volume, page likes, and post volume. We measure by log10 of volume of page likes and the proportion of fan pages with the specified page likes are computed, and the quitting-tobacco groups attain maxima at $log_{10}$ (page likes) = 1.5, implying that the majority in those two groups of fan pages have roughly $10^{1.5} \approx 32$ supporters. It peaks at $log_{10}$ (page likes) = 3, on the other hand with a share of 21.6 percent for the pro-tobacco group, indicating that have roughly $10^3 = 1000$ supporters' and 21.6 percent of pro-tobacco fan pages. Furthermore, the shares of the quit and anti-tobacco groups decline fast when the $log_{10}$ (page likes) $\geq 2$. Furthermore, after $log_{10}$ (page likes) $\geq 3$, pro-tobacco group outnumbers, according to our findings the other two groups. This implies that tobacco advertising fan sites are more efficacious at attracting the attention of the user and eliciting users to participate. The cumulative distribution functions (CDF) growth rate for the protobacco community is the slowest. When $IE = 0$, the proportion in the pro-tobacco group is the least; when $IE \geq 5$, the percentage in the pro-tobacco group exceeds another group.

Similarly, for each group, we examined the distribution of comment volume and post volume. Pro-tobacco group outnumbers the another's, when $log_{10}$ (post) $\geq 2$. This demonstrates that several pro-tobacco webpages are and have a really higher number of postings. It has been shown that more than 70% of follower webpages in quitting tobacco and anti-tobacco organizations contain comments less than 10. For the pro-tobacco group, on the contrary, it is roughly 56%. The maximum sites with minimal post indicate that more quitting tobacco and anti-tobacco fan pages are failing to stimulate user participation. In particular, if $log_{10}$ (comment) = 0, that is, 25.6% of pro-tobacco, 46.5% of anti-tobacco, and 37.4% of quitting tobacco. A large number of sites with minimal comments suggests that the majority of these stop smoking and quitting fan pages aren't really helping tobacco control programs or helping smokers in getting social support to quit.

**Topological Patterns.** Based on Facebook's history of user interactions, we are building three social networks: Tobacco Quit Network (TQN), Pro-Tobacco Network (PTN), and Anti-Tobacco Network (ATN). Huge components are retrieved for each network, and specified metrics are assessed appropriately. PTN size accounts for almost half of the user activity in the overall dataset, demonstrating that several Facebook users view pro-cigarette material and engage with tobacco advertising content, as seen in Table 1. On average, QTN is better than the other two networks, so QTN communication is better. Three clustering coefficients (3) of the systems, on the other hand, are considerably different. People that connect with the same pro-tobacco fan webpages rarely identify someone else, and user contact is essentially random, according to the PTN's lowest clustering coefficient (0.000023). The clustering coefficient for QTN, on the other hand, is much larger than on the similar vertex that of a random graph created. The crucial probability at which information percolates across the entire system is the percolation threshold $P_c$ defined in (4). PTN and ATN have almost the same percolation threshold $P_c$ of 0.5, demonstrating that PTN and ATN work in a similar way in terms of information distribution. However, $P_c$ for QTN is 0.25, implying that QTN will aid in the transmission of knowledge.

**Uncertainty in User Interaction.** Using information entropy [36], On the given fan pages, we assess whether or not user involvement is active. The average amount of information is called an entropy, in information theory, it included every communication and is better described as a calculation of unpredictability. For online tobacco-related groups, in this study, to calculate the unpredictability of user activity, we utilize entropy. Active users are those who have a greater entropy. Quitting tobacco and anti-tobacco groups develop faster than the overall dataset of the cumulative distribution functions (CDF), but the CDF of the pro-tobacco group stands out as having the slowest growing frequency. It is demonstrated that in the pro-tobacco group user involvement is more active. The response rates for the quitting tobacco and anti-tobacco groups are also above 80%, when using $IE \leq 6$. The pro-tobacco faction, on the other hand, has a figure of around 60%. The difference in $IE$ implies that fan pages for nicotine marketing have attracted much people than the other fan sites.
The dispensations of $IE$ for several groups are shown in the inset. When $IE \leq 5$ is used, it is clear that the numbers for both quitting and anti-tobacco organizations outnumber those for pro-tobacco ones. Pro-tobacco group percentage is the least when $IE \leq 0$, compared to over 25% group for the quitting tobacco and over 20% for the anti-tobacco. The pro-tobacco group, on the other contrary, outperforms the other two groups significantly, when $IE > 5$. The growth of pro-tobacco fan pages with maximum interaction entropy suggests that tobacco promotion fan pages are more successful for user engagement.

**Classification of User Interaction.** Transfer entropy (TE) to characterize user engagement in tobacco-related groups in this research as well as we show how to utilize interaction entropy (IE). The figure 1 shown below might be used to categorize user interactions. We divide user interaction into four categories: High transfer entropy High interaction entropy (HH), Low transfer entropy Low interaction entropy (LL), High transfer entropy Low interaction entropy (HL), Low transfer entropy High interaction entropy (LH). The administrators of HH sites have a lot of power, and user participation on these pages is highly lively. For HL pages, this indicates that user engagement is minimal, but the administrator's effect is great. While the administrators of LH sites aren't particularly powerful, the user participation on these pages is rather lively. LL page admins, on the surface, appear to be inactive and have little impact over their followers.

We discovered that the number of LL pages in the sample is overwhelming, implying that while several fan pages are tobacco-related formed on Facebook, the majority of them are inactive and fewer important, with transfer entropy and low interaction. LL pages can appear for a variety of reasons: (1) Abandoned pages: Because the admins of fan sites didn't adequately manage the webpages, they were neglected or removed. The number of posts on the LL pages is really minimal according to figure 1. Poor upkeep can manifest itself in a variety of ways. Some pages are made with the intention of gaining more page likes but providing no engaging post updates. As a result, they are only active for a brief time period; (2) newly-born pages: the page is formed extremely close to the data gathering break. Only a few postings are issued for newly-created pages, with few pages likes and comments. As a result, LL pages labeled as newly-born pages.
The HL pages are generally unusual. The conventional perception that prominent users are more likely to influence or affect [38,39] is contradicted by the definition of HL pages. When user involvement is minimal, it suggests that the page is not appealing to Facebook users. As a result, gaining strong influence on sites with low IE is challenging. Launching advertising ads on HH sites, on the other hand, is a breeze. A high level of user participation exposes to user-generated content a large number of users, ultimately increasing the visibility of occupational items, ultimately increasing the visibility of commercial items. The considerable influence of the administrators, on the other hand, enhances information diffusion outside the local community. Tobacco advertising, as portrayed on HH pages, is overpowering. 'Animals Smoking Durrys,' for example, portrays smoking as a joyful activity with smoking animals.

LH page administrators have less power than HH page administrators. We look at comment trends and post sources to see how they change across HH and LH sites. We define R as the ratio of the administrator's post volume to the entire post volume on a specific page, in terms of post sources. The value $R_{LH} = 0.2656$, for LH pages, whereas for HH pages, $R_{HH} = 0.7151$. This suggests that the pages are maintained by fans rather than administrators of the LH sites. It also shows the promotional methods of two different types of fan pages. The administrators are in charge of maintaining the HH pages. The primary mode is the strategy's name. It is the potential users, in the case of the LH pages, not the administrators, who are in charge of providing updates. The crowdsourcing mode is what it's called. The average comment volume (CVP) per post differed significantly between the two types of pages ($CVP_{HH} = 6.2281$, $CVP_{LH} = 1.6730$). The poor transfer entropy is due to the lower CVP for LH pages.

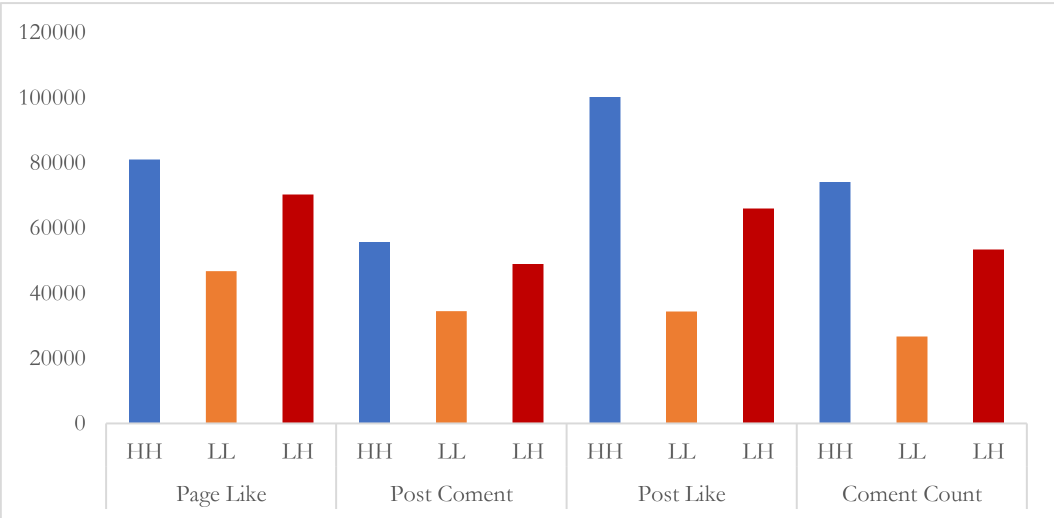


**Figure 1. Comparison in different groups of page features.**

## Discussion

This research shows that the proliferation of pro-tobacco content has a continuing effect, resulting in much engaged users and more influential inspiration, and illustrate the global tobacco control deficit of social media resources. In terms of user engagement and social impact, we discovered that the pro-tobacco group has more active user interaction, and tobacco promotion fan sites are more successful in attracting user consideration. Such data point to a widening gap among tobacco control, cigarette promotion, and that tobacco control is losing impetus on social media to tobacco promotion. However, there are certain limitations to the work presented in this publication.

To begin, the dataset's coverage is critical in revealing the disparities amongst tobacco-related networks. To address this issue, we created a database of tobacco-related keywords by combining three different information origins to optimize the dataset's attention: existing knowledge, tobacco brands, and synthetic tobacco terms. To include as many cigarette brands as possible in our dataset, we combine tobacco brands from tobacco review websites and online tobacco stores. We collected a significant number of smoking concerning words in synthetic tobacco terms, which include both tobacco marketing and smoke-free operations. Also featured are a variety of tobacco goods like as tobacco accessories such as pipes, snuff, hookah and cigars etc. Although none of these strategies are sufficient to assure the keyword set's completeness, they do help to widen the set's coverage. Furthermore, with the inclusion of new phrases, our system is scalable. Because of this characteristic, the dataset is suitable for revealing user involvement in tobacco-related forums.

Secondly, we employ the interactions and transfer entropies to show the range of user participation and its impact. Many types of user actions, such as post likes, comments, and sharing, are offered for the diversity of user involvement. Only comment records are counted, in the transfer entropy. On the one hand, a liked post is missing on the timestamp. Its lack of post-like information types revealing the changing dynamics of social networks impossible. On the other side, reputable businesses are inundated with bogus Facebook likes [40,41,42,43,44]. Fake likes lead to webpages that are overrun with inactive followers, resulting in skewed stats. Like a fake (http://likefake.com), for example, is a simple technique for boosting popularity with fake likes. We solely use comment data to gauge the effect of fan pages to resolve this issue. The comments on the postings are presumed to be genuine user interactions rather than false, abusive accounts. We will obtain a better understanding of how many individuals are genuinely engaged if we remove bogus likes. Furthermore, we will have a better understanding of their impact in tobacco communities.

Although this study has certain limitations, it might be useful in a variety of situations. We can establish the breadth of tobacco production debates and dialogues in non-traditional venues by analyzing user engagement in tobacco-related social networks, according to the Food and Drug Administration's research goals [31]. Nicotine businesses, stand to gain a lot of social media without running the danger of being accused of breaking any laws [11] from the marketing possibilities. Tobacco-related sale efforts on social media must be automatically removed from a large amount of user-generated material. The study of user engagement exposes the frameworks of user interaction in the selling campaigns, allowing them to be distinguished from other tobacco-related information.

## Methods

**Data Collection.** The major focus of this article is on user activity on tobacco-related Facebook fan sites. To locate tobacco-related fan sites, we first create a list of tobacco-related keywords that includes synthetic tobacco-terms, tobacco brands, and existing information. We incorporate tobacco brands from four different sources diversity of tobacco brands: the official smoking brand list, online cigarette shops, tobacco review web sites, and tobacco-brand related wiki webpages. We collected 186 cigarette brands in total and calculated how many times each brand appeared in each data source. Finally, the representative brands were chosen based on the average frequency of occurrence. We obtained a total of 70 well-known brands. We create synthetic keywords by combining tobacco-related terms with various origins (as shown in Table 2). Over, certain well-known tobacco-related fan webpages like as 'quintet,' 'VchangeU,' and 'BecomeAnEX,' as well as tobacco-related items such as 'hookah', 'beedi,' 'cigar,' 'snuff,' and 'pipe smoking,' have been complemented. RestFB [45], a versatile and open-source solution for accessing the Facebook Graph APIs [46], is used to fetch fan pages related to the supplied keywords.

**Table 2.** Statistical patterns for tobacco-related fan webpages

| Group | Comment Volume | Page like | Post Volume | Post Like |
|---|---|---|---|---|
| **Tobacco Cessation** | 191004 | 1569490 | 837938 | 757 |
| **Tobacco Control** | 81236 | 1232153 | 489853 | 684 |
| **Tobacco Promotion** | 521091 | 2909532 | 14660450 | 708 |
| **Total** | 793331 | 5711175 | 15988241 | 2149 |

**Table 3.** Instructions for synthetic tobacco-related keywords.

| Prefix (Suffix) | Roots | Keywords |
|---|---|---|
| **Anti-<br>Free Stop Quit<br>Prevent** | Smoking<br>Tobacco<br>Cigarette | Nicotine, cigarette free, prevent tobacco, prevent cigarette, smoking free, anti-tobacco, anti-cigarette, quit cigarette, quit smoking, anti-smoking, tobacco free, stop tobacco, quit tobacco, prevent smoking, stop cigarette, stop smoking. |
| **Addiction<br>Cessation<br>Prevention** | Tobacco<br>Nicotine<br>Cigarette<br>Smoking | Cigarette cessation, nicotine addiction, tobacco cessation, tobacco addiction, smoking cessation, cigarette addiction, smoking addiction, tobacco prevention, nicotine cessation, cigarette prevention, smoking prevention, nicotine prevention |

According to content of fan pages and page profiles, all recovered fan pages are manually divided into four groups. To be more specific, we manually classified the retrieved results into four types by two coders: (0: unrelated to smoking; 1: cessation; 2: control; 3: promotion), and then used the Cohen's Kappa ($K$=0.9519) coefficient [47], which has a high coding reliability, to measure the similarity of these two coders' arrangement outcomes. When the first two coders cannot agree, we delegate the fan pages to the third coder. The fan webpages are eliminated with each of the first two coders if the third coder disagrees. We found 2149 tobacco-related fan pages in total from post like for tobacco (757 cessation, 684 control, and 708 promotion).
Fan page characteristics such as post volume, page like, comment volume, and post like time are gathered for tobacco-related fan sites. Additionally, the user interaction actions are recorded. When user $A$ responds to a post made by user $B$ by liking or sharing it, we have defined one interaction between the two users. The data was acquired in the spring of 2013 [12] by our newly built tobacco surveillance system. We create $\boldsymbol{G} = \{V, E\}$ which an undirected weighted interaction graph based on the user interaction data, where users that interact with posts denoted by $V = \{v_1, v_2, \ldots v_i, \ldots v_n\}$, and the set of edges denoted by $E = \{e_1, e_2, \ldots e_i, \ldots e_m\}$ connecting $v_i$and $v_j$with interaction frequency $w_{ij}$. We were able to create three interaction graphs based on the labeling of fan pages. Table 3 shows the diameters of three different interaction graphs.

**Measuring Network Topology.** Using the metrics described in (1), (2), (3), (4), we analyze the connectedness, transitivity, and resilience of the three tobacco-related groups. The degree distribution (1) illustrates the network's connection. The first instant, $|k|$, represents the network's average degree. The degree $k$ of node $i$ is determined by the number of edges that intersect it, and $P(k)$ is characterized as the fraction of network vertices having degree $k$ [48]. To calculate the probability that two node neighbors are linked to one another, the clustering coefficient $C$ is proposed to assess the transitivity of one social network. The degree to which nodes cluster into

communal groupings is shown by modularity. Assume the network's nodes are divided into neighborhoods, with $c_i$ recording node $n_i$ s community membership. Modularity of the partitioning shows in (2), where number of edges denoted by $m$, and the adjacency matrix denoted by $A$, The degrees of nodes $i$ and $j$ are denoted by $k_i$ and $k_j$, respectively, while $r_i r_j$= 1 specifies whether node $i$ and $j$ join to the similar group, and -1 otherwise [49].
The capacity of a system to preserve its connection qualities following a chance omission of a percentage of its edges and nodes is characterized as robustness (also known as resilience). In most cases, these difficulties are solved by employing percolation theory [50,51] analytically to locate the critical point, denoted by $P_c$. For some lattices, the critical point value may be computed accurately. $P_c$ is precisely computeable for the Bethe lattice [52,53]. The vertex number of immediate neighbors denoted by $t$ and $P_c$ for the Bethe lattice is dominated by $t$, as seen in the following equation. According to [54], the average degree of a social network may be approximated by $t$.

**Degree Distribution** (1)

$$|k^n| = \sum_k k^n \, P(k)$$

**Modularity** (2)

$$Q = \frac{1}{2m} \sum_{ij} \left[ A_{ij} - \frac{k_i k_j}{2m} \right] \cdot \frac{r_i r_j + 1}{2}$$

**Clustering Coefficients** (3)

$$C = \frac{1}{N} \sum_{i=1}^{N} c_i \text{ where } \quad c_i = \frac{e_i}{k_i(k_i - 1)/2}$$

**Percolation Threshold** (4)

$$P_c = \frac{1}{t-1} \sum_{i=1}^{N} c_i \text{ where } \ t = \sum_k k^n \, P(k)$$

**Quantifying User Interaction.** To determine if the user engagement is active or inactive, we employ the IE [36], which may be used to evaluate the variety of user interaction on fan sites. The higher the interaction entropy on fan sites, more active user engagement. We specifically gathered user interaction history for each fan page. If a post is shared on page $L$ and user $i$ comments (likes or shares), it is considered an interaction between them. To collect the connection behaviors and regularity among users, we used a triple, denoted by $I = (L, u_i, w_i)$, where $u_i$ are the unique IDs of user $i$ and $L$ is unique IDs of page $L$, and interaction frequency between them are denoted by $w_i$. As a result, (5) may be used to calculate the interaction entropy of page $L$, where $p_i$ is the likelihood of contact among user $i$ and page $L$. It demonstrates the diversity of user involvement on fan sites in terms of user interaction. More users engage with page $L$ when the interaction entropy is large and the users who interact with page denoted by $n$.

$$H(L) = -\sum_{i=1}^{n} p_i \log p_i, \quad p_i = \frac{w_i}{\sum_{i=1}^{n} w_i} \tag{5}$$

**Identifying Social Influence.** Transfer entropy (TE) is used in this research to disclose the quite well connection between peers and to investigate the formation of the system from a macro perspective. To begin, we consider user activities in chronological order to be a stochastic procedure. Then, in online social networks, for each user (identified as $Y$), we create a stochastic procedure $Y_s$ that records the history of interaction activities. By (6) indicated by $H(Y)$, and probability of $H(Y = y_i)$ is $p(y)$, it defines the dynamics of user $Y$. Consider $Y$ and $Z$ are two random users. In the following (7) and (8) determine the conditional entropy $H(Y|Z)$ and joint entropy $H(Y, Z)$ for $Y$ and $Z$, respectively, where $p(y|z)$ and $p(y, z)$ are the conditional possibility and joint possibility, respectively. Now we'll look at stochastic procedures. Two stochastic processes, $Y_s$ and $Z_s$, reflect the behaviors of two users, $Y$ and $Z$. The knowledge on the previous state $Z_s^{(s-k)}$ of $Z$, given by $Z_s^{(s-k)} = Z_s, Z_{s-1}, \dots Z_{s-k}$, reduces uncertainty regarding $Y_{s+1}$, in addition to previous $Y_s^{(s-k)}$ states of $Y$, as indicated by $Y_s^{(s-k)} = Y_s, Y_{s-1}, \dots Y_{s-k}$, TE from $Z$ to $Y$, given as [55], is used to calculated by following equations

$$H(Y) = -\sum_y p(y)\log p(y) \quad (6)$$
$$H(Y \mid Z) = -\sum_{y,z} p(y,z)\log p(y \mid z) \quad (7)$$
$$H(Y,Z) = -\sum_{y,z} p(y,z)\log p(y,z) \quad (8)$$
$$TE(Z \to Y) = H\left(Y_s \mid Y_{s-1}^{(s-k)}\right) - H\left(Y_s \mid Y_{s-1}^{(s-k)}, Z_{s-1}^{(s-l)}\right) \quad (9)$$

The first component in (9) denotes the uncertainty regarding $Y_s$ based only on $Y'$s past. When $Z$'s history is also provided, the second component expresses the uncertainty. $TE(Z \to Y)$ has the same definition. We'll use $l = k$ from now on for simplicity's sake. The asymmetrical transfer entropy across distinct random procedures is accompanied by a decreased uncertainty about the other in one process due to the understanding [56,57]. Transfer entropy, which was introduced in [58], is a measure of peer-to-peer influence across people on social media. In tobacco-related social networks, we also looked at peer influence. Then, from a network viewpoint, we looked at the effect of a certain fan page. Considering a system $\boldsymbol{G} = \{V, E\}$, with $V$ denoting users that interact with posts and $E$ denoting the set of edges linking $v_i$ and $v_j$, the impact of node $v_i$ in $\boldsymbol{G}$ at time $s$ is described as (10), where $C(L)$ defines the set of nodes $v_j$

$$I(v_i) = -\sum_{v_j \in C(L)} TE(v_i \to v_j) \quad (10)$$

The social influence of node $v_i$ changes over period as the social network evolves. As a consequence, utilizing a time series of user establishment of relationships, we can apply (10) to compute the evolution of social impact. In this article, TE is calculated using the release time of posting and the timestamps of responses. Based on the timestamps of posts given page $L$, we calculate the random procedure of page $L$. For every individual $v_j$ who reacts to the posts on page $L$, we analyze an overall history bunch of posts as a stochastic process. Then, using (10), we determine the development of page $L$'s effect.

## References


[1] World Health Organization Media Center, *Tobacco*. (2014) Available at: http://www.who.int/mediacentre/factsheets/fs339/en/. (Accessed: 15 September 2014).

[2] Ellis, L. D., Soo, E. C., Achenbach, J. C., Morash, M. G. & Soanes, K. H. Use of the Zebrafish Larvae as a Model to Study Cigarette Smoke Condensate Toxicity. *PLoS ONE* **9,** e115305 (2014).

[3] Giovino, G. A. The tobacco epidemic in the United States. *Am. J. Prev. Med.* **33,** s318–s326 (2007)

[4] Hecht, S. S. Tobacco carcinogens, their biomarkers and tobacco-induced cancer. *Nat. Rev. Cancer* **3,** 733–744 (2003).

[5] Yanbaeva, D. G., Dentener, M. A., Creutzberg, E. C., Wesseling, G. & Wouters, E. M. Systemic Effects of Smoking. *Chest* **131,1557**–1566 (2007)

[6] World Heart Federation, *Tobacco: totally avoidable risk factor of CVD*. (2012) Available at: http://www.world-heart-federation. org/press/fact-sheets/tobacco-totally-avoidable-risk-factor-of-cvd/. (Accessed: 3 February 2015).

[7] Flensborg-Madsen, T. *et al.* Tobacco smoking as a risk factor for depression. A 26-year population-based follow-up study. *J. Psychiatr. Res.* **45,** 143–149 (2011).

[8] Grucza, R. A. *et al.* Probing the smoking-suicide association: do smoking policy interventions affect suicide risk? *Nicotine Tob. Res.* **16,** 1487–1494 (2014).

[9] Christakis, N. A. & Fowler, J. H. The collective dynamics of smoking in a large social network. *New Engl. J. Med.* **358,** 2249–2258 (2008).

[10] Taki, F. A., Pan, X., Lee, M. H. & Zhang, B. Nicotine exposure and transgenerational impact: a prospective study on small regulatory microRNAs. *Sci. Rep.* **4,** 7513 (2014).

[11] Freeman, B. & Chapman, S. British American Tobacco on Facebook: undermining article 13 of the global World Health Organization Framework Convention on Tobacco Control. *Tob. Control* **19,** e1–e9 (2010).

[12] Liang, Y. *et al.* An Integrated Approach of Sensing Tobacco-Oriented Activities in Online Participatory Media. *IEEE Syst J* (In press)

[13] Liang, Y., Zheng, X., Zeng, D. D., Zhou, X. & Leischow, S. J. An Empirical Analysis of Social Interaction on Tobacco-Oriented Social Networks. *Paper presented at International Conference Smart Health*, Beijing, China. Berlin Heidelberg: Springer. 2013, August 3-4.

[14] Wang, F. *et al.* Chinese Tobacco Industry Promotional Activity on the Microblog Weibo. *PLoS ONE* **9,** e99336 (2014).

[15] Wang, F., Zheng, P., Freeman, B. & Chapman, S. Chinese tobacco companies' social media marketing strategies. *Tob. Control* (In press)

[16] Savell, E., Gilmore, B. A. & Fooks, G. How does the tobacco industry attempt to influence marketing regulations? A systematic review. *PLoS ONE* **9,** e87389 (2014).

[17] Elkin, L., Thomson, G. & Wilson, N. Connecting world youth with tobacco brands: YouTube and the internet policy vacuum on Web 2.0. *Tob. Control* **19,** 361–366 (2010).

[18] Richardson, A. & Vallone, M. A. YouTube: a promotional vehicle for little cigars and cigarillos? *Tob. Control* **23,** 21–26 (2014).

[19] Seidenberg, B. A., Rees, W. V. & Connolly, N. G. Swedish Match marketing on YouTube. *Tob. Control* **19,** 512–513 (2010).

[20] Carroll, V. M., Shensa, A. & Primack, A. B. A comparison of cigarette- and hookah-related videos on YouTube. *Tob. Control* **22,** 319–323 (2013).

[21] Freeman, B. & Chapman, S. Is YouTube telling or selling you something? Tobacco content on the YouTube video-sharing website. *Tob. Control* **16,** 207–210 (2007).

[22] Hua, M., Yip, H. & Talbot, P. Mining data on usage of electronic nicotine delivery systems (ENDS) from YouTube videos. *Tob. Control* **22,** 103–106 (2013).

[23] Blue, L. *Five smartphone apps that promote smoking.* (2012) Available at: http://healthland.time.com/2012/10/24/five-smartphone-apps-that-promote-smoking/. (Accessed: 15 September 2014)

[24] BinDhim, N. F., Freeman, B. & Trevena, L. Pro-smoking apps for smartphones: the latest vehicle for the tobacco industry? *Tob. Control* **23,** e4 (2014).

[25] Cavazos-Rehg, P. A., Krauss, M. J., Spitznagel, E. L., Grucza, R. A. & Bierut, L. J. The Hazards of new Media: Youth's exposure to tobacco ads/Promotions. *Nicotine Tob. Res.* **16,** 437–444 (2014).

[26] Backinger, C. L. *et al.* YouTube as a source of quitting smoking information. *Tob. Control* **20,** 119–122 (2011).

[27] Duke, J. C., Hansen, H., Kim, A. E., Curry, L. & Allen, J. The Use of Social Media by State Tobacco Control Programs to Promote Smoking Cessation: A Cross-Sectional Study. *J. Med. Internet Res.* **16,** e169 (2014).

[28] Hefler, M., Freeman, B. & Chapman, S. Tobacco control advocacy in the age of social media: using Facebook, Twitter and Change. *Tob. Control* **22,** 210–214 (2013).

[29] Freeman, B. New media and tobacco control. *Tob. Control* **21,** 139–144 (2012).

[30] Ribisl, K. M. & Jo, C. Tobacco control is losing ground in the web 2.0 era: invited commentary. *Tob. Control* **21,** 145–146 (2012).

[31] Center for Tobacco Products, *Food and Drug Administration Research Priorities*. (2012) Available at: http://www.fda.gov/ downloads/tobaccoproducts/newsevents/ucm293998.pdf. (Accessed 15 September 2014)

[32] Centers for Disease Control and Prevention, *Best Practices for Comprehensive Tobacco Control Programs*. (2014) Available: http:// www.cdc.gov/tobacco/stateandcommunity/best_practices/pdfs/2014/comprehensive.pdf (Accessed: 15 September 2014)

[33] Liang, Y. *et al.* Exploring How the Tobacco Industry Presents and Promotes Itself in social media. *J. Med. Internet Res.* **17,** e24 (2015).

[34] Liang, Y. et al. Exploring How the Tobacco Industry Presents and Promotes Itself in social media. J. Med. Internet Res. 17, e24 (2015).

[35] Luo, C., Zheng, X., Zeng, D. D. & Leischow, S. J. Portrayal of Electronic Cigarettes on YouTube. BMC Public Health 14, 1028 (2014).

[36] Konstantinos, E. F. et al. Impact of flavour variability on electronic cigarette use experience: An internet survey. Int. J. Environ. *Res. Public Health* **10,** 7272–7282 (2013).

[37] Cover, T. M. & Thomas, J. A. *Elements of information theory* 2nd Edition (John Wiley & Sons, 2006)

[38] Schreiber, T. Measuring information transfer. *Phys. Rev. Lett.* **85,** 461–464 (2000).

[39] Xia, S. & Liu, J. A computational approach to characterizing the impact of social influence on individuals' vaccination decision making. *PLoS ONE* **8,** e60373 (2013).

[40] Moussaid, M., Kammer, J. E., Analytis, P. P. & Neth H. Social influence and the collective dynamics of opinion formation. *PLoS ONE* **8,** e78433 (2013).

[41] Edwards, J. *Facebook Advertisers Complain of A Wave Of Fake Likes Rendering Their Pages Useless*. (2014) Available at: http:// www.businessinsider.com/facebook-advertising-fake-likes-2014-2#ixzz3CDUl0Xa0. (Accessed: 15 September 2014)

[42] Fire, M., Kagan, D., Elyashar, A. & Elovici, Y. Friend or foe? Fake profile identification in online social networks. *Soc. Netw. Anal. Min.* **4,** 1–26 (2014)

[43] Christakis, N. A. & Fowler, J. H. The spread of obesity in large social network over 32 years. *New Engl. J. Med.* **357,** 370–379 (2007).

[44] Gallos, L. K., Barttfeld, P., Havlin, S., Sigman, M. & Makse, H. A. Collective behavior in the spatial spreading of obesity. *Sci. Rep.* **2,** 454 (2012).

[45] Demongeot, J. & Tatamasco, C. Evolution of social networks: the example of obesity. *Biogerontology* **15,** 611–626 (2014)

[46] RestFB. Available: http://restfb.com/. Accessed 3 February 2015.

[47] Facebook. Available: https://developers.facebook.com/. Accessed 3 February 2015.

[48] Carletta, J. Assessing agreement on classification tasks: The kappa statistic. *Comput. Linguist.* **22,** 249–254 (1996).

[49] Boccaletti, S., Latora, V., Moreno, Y., Chavez, M. & Hwang, D. U. Complex networks: Structure and dynamics. *Phys. Rep.* **424,** 175–308 (2006).

[50] Newman, M. E. J. Modularity and community structure in networks. *Pro. Natl. Acad. Sci.* **103,** 8577–8582 (2006).

[51] Bollobas, B. & Riordan, O. *Percolation* (Cambridge University Press, 2006).

[52] Stauffer, D. & Aharony, A. *Introduction to percolation theory: Revised Second Edition* (Taylor & Francis Ltd., 1994)

[53] Bethe, H. A. Statistical theory of superlattices. *P Roy Soc. Lond. A. Mat.* **150,** 552–575 (1935).

[54] Baek, S. K., Minnhagen, P. & Kim, B. J. Percolation on hyperbolic lattices. *Phys. Rev. E* **79,** 011124–011131 (2009).

[55] Bolourian, A. H. A., Moshfeghi, Y. & Van Rijsbergen, C. J. Quantification of Topic Propagation Using Percolation Theory: A Study of the ICWSM Network. *Paper presented at of the third International AAAI Conference on Weblogs and social media*, California, USA. AAAI Press. 2009, May 17-20.

[56] Kaiser, A. & Schreiber, T. Information transfer in continuous process. *Physica. D* **166,** 43–62 (2002).

[57] Sun, J. & Bollt, E. M. Causation entropy identifies indirect influences, dominance of neighbors and anticipatory couplings. *Physica. D* **267,** 49–57 (2014).

[58] Lizier, J. T. & Prokopenko, M. Transfer entropy and transient limits of computation. *Sci. Rep.* **4,** 5394 (2014).

[59] Steeg, V. G. & Galstyan, A. Information transfer in social media. Paper presented at the 21st international conference on World Wide Web, Lyon, France. New York: ACM. 2012, April 16-20.