Special Issue: Medicine, Humanity and Media

# Tobacco control challenges in China: Big data analysis of online tobacco marketing information

Xiuli Wang [*], Ying Xiong, Wenwen Zhao

*School of New Media, Peking University, Director of Center for Social Media Research, Beijing, China*

## ARTICLE INFO

## ABSTRACT

*Objective:* To develop tobacco control strategies by analyzing online tobacco marketing information in China.

*Methods:* Using web-crawler software, this study acquired 106,485 pieces of online tobacco marketing information published on 11 different Internet platforms including Weibo, WeChat, Baidu, etc., from January—June 2018. The data were used to investigate the characteristics and social networks of online tobacco marketing via content and social network analysis.

*Results:* The total volume of online tobacco marketing during the study period was high, showing a positive trend. Of all the marketing subjects, those involving "flavor capsule", "Marlboro", and "Esse" were the most popular. The Weibo platform had the highest volume of online tobacco marketing information as well as the largest proportion of explicit marketing information. This was followed by other social media platforms such as Baidu Search, Baidu Tieba, and Xiaohongshu, where implicit marketing information predominated. The overall network structure of tobacco websites exhibited a significant centralization feature, where traditional and novel tobacco websites formed two clusters with almost no intersections. The China Tobacco Science and Education Website (http://www.tobaccoinfo.com.cn/) and E-Cigarette Home (http://ecigm.com/) were the two nodes of the highest degree centrality within the respective "circle", while the China Tobacco Monopoly Bureau Website (http://www.tobacco.gov.cn/) was the node with the highest closeness centrality. By contrast, Baidu Tieba's overall network structure was more decentralized, and the degree of correlation between different nodes was relatively low.

*Conclusion:* Online tobacco marketing demonstrated high volumes and wide coverage, and an intertwined network, thereby creating major obstacles for tobacco control. To address this issue, the government should strengthen network supervision of tobacco marketing and revise its current regulations. Meanwhile, Internet platforms should improve self-regulation by comprehensively removing and blocking tobacco-related information. Lastly, the media and public should advocate associated policies and support Internet platform supervision.

## What is known?

- Research on online tobacco marketing can be categorized into two opposing discourse systems: "Pro-tobacco," or "Anti-tobacco." The former is predominantly based on the perspective of management and marketing, where online tobacco marketing is regarded as the "main battlefield" for tobacco industry because of the profits it represents. On the contrary, the latter is based on the public health, preventive medicine, and health communication fields, where online tobacco marketing is considered a "heavy disaster area" that threatens people's health.
- Some studies have investigated online tobacco marketing including its background, motivations, and negative influences. Internet marketing currently sits in a legal grey area.

## What is new?

- This study used a systematic, empirical approach to comprehensively analyze volume characteristics, platform preferences, and social networks of online tobacco marketing in China.
- It shows that internet users under 29 years old are at a higher risk for viewing tobacco advertisements, and therefore more vulnerable to tobacco consumption.

\* Corresponding author. No.5 Yiheyuan Rd, Haidian District, Beijing, 100871, PR China.

*E-mail address:* Xiuli.wang@pku.edu.cn (X. Wang).

- Women have been increasingly targeted by tobacco companies via feminized cigarette brands and products.
- Governmental policies are urgently needed to enforce tobacco advertising regulations and promote public health effectively.

## 1. Introduction

China is the world's largest tobacco producer and consumer. Smoking's prevalence has brought significant health, economic, and social losses. Ineffective tobacco regulations will cause more than 200 million Chinese people to die from smoking-related diseases in this century, which will contribute to reduced productivity and tens of millions of people falling into poverty. In addition, pressure on social welfare, public health systems and the private sector will correspondingly increase, exacerbating growing inequality levels [1]. Due to the negative externalities of the tobacco industry, China joined the World Health Organization Framework Convention on Tobacco Control (FCTC) in 2003 and launched a series of tobacco control measures. Consequently, the result of the "2018 Chinese Adult Tobacco Consumption Survey" showed that the overall rate of smokers over 15 years old was 26.6% in China, a decline from previous years. However, this number was substantially higher than the goal set by the "Healthy China 2030" initiative, which set the goal that "The percentage of smokers over 15 years old should be less than 20%" [2].

It is worth mentioning that whilst various social entities represented by the National Health Commission of the People's Republic of China (referred to as "the National Health Commission") are actively promoting tobacco control compliance, the tobacco industry constantly seeks countermeasures, and the Internet has become the battlefront between both parties. As of June 2019, China's Internet penetration rate was 61.2%, with netizens and online shopping reaching 854 million and 639 million people, respectively [3]. In addition, recent developments in e-commerce and social media, have normalized the intersection between advertising and social media, "Owing to the constant practice and education of the online social shopping market, netizens have gradually accepted the dissemination of marketing information on social applications and started trusting this information" [4]. Through various Internet platforms, the tobacco industry has spread massive amounts of tobacco marketing information, which has been used to establish consumer groups, shape consumer behaviors, and consolidate a tobacco culture. Therefore, the Internet has become one of the focal points of tobacco control. To address this issue, this study adopted online tobacco marketing information as the research object, to recommend and develop tobacco regulations, and indicate directions for future research.

## 2. Literature review

Research on online tobacco marketing can be categorized into two opposing discourse systems: "Pro-tobacco," or "Anti-tobacco." The former is predominantly based on the perspective of management and marketing, where online tobacco marketing is regarded as the "main battlefield" for tobacco industry because of the profits it represents. On the contrary, the latter is based on the public health, preventive medicine, and health communication fields, where online tobacco marketing is considered a "heavy disaster area" that threatens people's health.

### 2.1. Literature on online tobacco marketing's background

Comprehensive analysis has been performed on the context and motives of online tobacco marketing by "Pro-tobacco" researchers [5—8]. The first was a reaction to the tightening and strengthening of tobacco control policies as well as the implementation of the "Advertising Law of PRC" (2015 edition), which has almost completely prohibited tobacco advertising through mass media. The second is based on online advertising's advantages such as communication speed, convenience, data processing efficiency, and cost, Internet marketing currently sits in a legal grey area. The last and most fundamental motivation is that the tobacco industry is both extremely profitable and legitimate, and is therefore regarded as an essential force that bolsters the country's finances and promotes local economies. Therefore, when the industry faces external pressure and development bottlenecks, there is an increased incentive to explore alternative methods to protect the industry and generate income.

In addition to tobacco websites portals, direct-email marketing, and industry newspapers, Internet platforms with obvious social characteristics, such as forums, blogs, Weibo, WeChat, QQ, Wang-Wang groups, live broadcast platforms, as well as major e-commerce and second-hand goods trading platforms have entered researchers' purview. The "Pro-tobacco" party promotes to apply such strategies and models as P2P (peer to peer), O2O (online to offline), word-of-mouth marketing, precision marketing, micro marketing and integrated marketing to online tobacco marketing, and has proposed such concepts as "Internet + Tobacco", "Smart Tobacco", and "Retail Terminal Ecosystem" [9,10]. Furthermore, both the "Pro-tobacco" and "Anti-tobacco" parties have focused on "soft" marketing strategies such as establishing a culture, talking about feelings, holding activities, contributing to public welfare, advertising on hot spots, and embedding marketing information in films and televisions. The difference is that the former [11] regards this strategy as a good experience worth promoting, whereas the latter [12,13] criticizes it as a disguised advertisement. Moreover, several studies investigating "Illegal tobacco-related online activities" like selling illegally imported or fake, inferior cigarettes indicated some companies profited and escaped supervision by replacing keywords, changing IP addresses, using a virtual identity, communicating point-to-point, or using the "Pay on-line and deliver off-line" strategy [14—16].

### 2.2. Research on online tobacco marketing's influence and regulation

In addition to illegally selling imported or fake cigarettes, research on tobacco marketing's negative influences and corresponding governance measures were initiated primarily by the "Anti-tobacco" party. In 2001, Cohen et al. proposed four major tobacco control challenges brought by online tobacco marketing: selling tobacco to youth, tobacco tax evasion and smuggling, unrestricted tobacco advertising and marketing, and continued normalization of the tobacco industry and its products [17]. Of all issues, the negative influence of tobacco on adolescents deserves special attention. A survey conducted by Fei Qi et al. on the exposure of junior high school students in Qingdao to tobacco advertising and marketing suggested that in a 30-day period, 34% of students reported seeing tobacco advertising and marketing information, mostly over the Internet (12.8%) [18]. Another survey performed by Huifang Shi et al. found that in a 3-month period, college students accounted for 38% of all respondents exposed to tobacco marketing through new media [19]. These findings are similar to studies in the United States, Australia, and other countries, that found tobacco advertising across all mediums, Internet, retail outlets, film, and television, can reinforce a pro-tobacco media environment and increase the smoking susceptibility of adolescents [20—22]. In addition, Vishwanathan et al. emphasized the effectiveness of social media in convincing young women to smoke [23].

The literature on online tobacco advertising has proposed the following: (1) Policies and regulations: extend the regulatory scope

and strengthen the enforcement of relevant laws, clearly define "product placement", and regulate the permission criteria of online trading platforms [12,16,19,24]; (2) Technical monitoring: "Utilize the Internet to govern itself". With focused crawlers, in-depth information can be mined automatically to collect information, gather evidence, and aid offline operations [14,25]; and (3) Health communication: By combining new media with traditional media, targeted information can be shared with different groups. In addition, training can improve the communication skills of relevant personnel [19,26].

In summary, previous studies have investigated online tobacco marketing including its background, motivations, and negative influences. However, the majority focused either on the entire industry or a specific platform, but did not conduct a systematic, empirical study to develop targeted advocacy and national policies. This study intends to investigate the volume of online tobacco marketing information, the gender and age of communicators and recipients, the features of tobacco marketing, and the network structure between different primary communication subjects, so that to help Chinese government make better policies to regulate online tobacco marketing. In addition, governmental action items have been proposed based on current policies and media environment. And due to the distinct effects of tobacco on adolescents and women, age and gender were evaluated. More specifically, this paper aims to answer the following three research questions:

(1) What is the overall volume of online tobacco marketing information? What are the volume characteristics if categorized by gender, age, region and time periods?
(2) What are the characteristics of different online tobacco marketing platforms?
(3) What are the social network structures of representative online tobacco marketing platforms? Who are the key opinions leaders in these networks?

## 3. Research methods

This study focuses on online tobacco marketing information. Marketing is a behavior that identifies and satisfies human demands for profit. Individuals and organizations are both involved in marketing activities, either formally or informally, consciously or unconsciously [27]. According to FCTC and China's Advertisement Law, we define tobacco marketing information as any information that is to directly or indirectly promote tobacco consumption or the sales of tobacco products. The main participants of tobacco marketing consist of tobacco producers, sellers, managers, and consumers. In this study, online tobacco marketing information is operationalized as any information that deals with tobacco but not opposes to the production, sale and consumption of tobacco, including all kinds of tobacco advertisement, tobacco culture, tobacco news, online tobacco discussions, etc; while information related to Tobacco hazards, tobacco ban, tobacco control, and smoking cessation are not included. By acquiring information related to online tobacco marketing for analysis, this paper aims to explore the volume, platform, and network characteristics of online tobacco marketing information.

### 3.1. Data source

Based on the findings of previous research and our pilot survey of online tobacco marketing information, this study selected 11 Internet platforms commonly used by tobacco marketers to acquire data, including Baidu Search, the world's largest Chinese search engine; four mainstream tobacco industry portal websites that cover more tobacco industry news, advertisement and tobacco culture: Oriental Tobacco Website (https://www.eastobacco.com/), Tobacco Online (http://m.tobaccochina.net/), Yanyue Website (https://www.yanyue.cn/tobacco) (excluding the forum), and China Cigarette Website (https://www.cnxiangyan.com/); and six social media platforms that focus more on user generated content: Weibo, WeChat, Zhihu, Xiaohongshu, Baidu Tieba, and Yanyue Forum. These 11 internet platforms have a large number of users, are highly active and influential, and each has its own characteristics, which facilitate the researchers to collect enough samples while making comparisons among different platforms.

After comprehensively reviewing the sales, prevalence and origins of different brands of tobaccos sold in China, a total of 16 keywords from four categories (behavior, type, brand, company) were selected: smoking, smoke a cigarette, flavor capsule, low tar, slim cigarette, Guiyan, Changbaishan, Liqun, Huanghelou, Zhonghua, Shuangxi, Haomao, Marlboro, Esse, Sichuan Zhongyan (Abbreviation of China Tobacco Sichuan Industrial Co., Ltd.), and Yunnan Zhongyan (China Tobacco Yunnan Industrial Co., Ltd.). Among them, smoking and smoke a cigarette (Xiyan and Chouyan in Chinese) are two most common tobacco consumption behaviors; flavor capsule, low tar and slim cigarette are the three popular cigarette types that are discussed a lot online; Guiyan, Changbaishan, Liqun, Huanghelou, Zhonghua, Shuangxi, Haomao, Marlboro, Esse are popular domestic and foreign tobacco brands that are liked by male or female Chinese consumers; and Sichuan Zhongyan and Yunnan Zhongyan are two big and influential provincial tobacco company.

Subsequently, tools such as Python, Requests, Scrapy, and Selenium were used to develop a web-crawler that automatically mined data and recognized keywords in the relevant networks from 1st January 2018 to 30th June 2018. It should be noted that owing to technical limitations on data display and restricted use of web-crawlers by each platform (eg.: when a keyword was searched in Baidu, only 76 result pages were returned), the acquired data were taken from the Chinese-language information displayed on the web pages, rather than the platforms' entire content. Moreover, the authenticity of the user information could not be verified. Therefore, the above data could only be used to interpret within the observation's scope.

Due to the complexity of the online tobacco information, three students went through all crawled data manually in more than three weeks, and removed non-tobacco related information, duplicate information, and information related to tobacco hazards, tobacco ban, tobacco control and smoking cessation. After the data cleaning, a total of 106,485 pieces of effective online tobacco marketing information were obtained. Each data point included such information as URL, source, time, nickname, title, and specific content. In addition, another 105,294 pieces of user gender information as well as 83,996 pieces of user location data related to tobacco marketing were collected via the Application Program Interface of Weibo. These data were supplemented by the Baidu index. Lastly, 539 pieces of friendly link information were obtained by logging into relevant websites.

### 3.2. Data analysis

To understand the presentation of tobacco marketing on the Internet, a content analysis was performed on captured data to distinguish between explicit and implicit tobacco marketing information. Explicit marketing information refers to the information that directly promotes tobacco transactions such as the brand, type, characteristics, price, and purchase channel of a certain tobacco. Alternatively, implicit information refers to the content that indirectly supports tobacco consumption, such as tobacco stories, cultural talks, corporate news, and other information not involving specific transactions. If a piece of tobacco marketing contained both
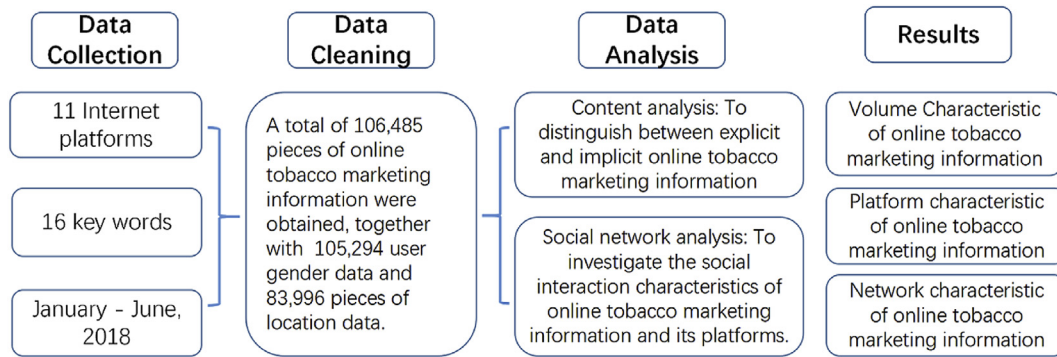
**Fig. 1.** Technical roadmap of the study.

explicit and implicit information, it was categorized by the length of both. Subsequently, after three coders went through unified training, 1000 pieces of information were randomly selected and encoded to test their reliability. The Krippendorff's α value of the reliability test was 0.9, indicating a high consistency among the three coders. Once the reliability was verified, they then completed the coding analysis of the remaining data.

To better understand the social network relationship between tobacco marketing information and Internet platforms, four mainstream tobacco industry portal websites from Baidu Tieba were selected as samples. Subsequently, a social network analysis was performed on the friendly link relational data and the post reply relational data of these websites by Networkx in Python, followed by data visualization using Echarts. By calculating the neighboring relationships between different nodes, we were able to explore the social interaction characteristics of various online tobacco marketing entities such as their overall and individual network structures on different platforms, and identify opinion leaders and activists with special significance. Fig. 1 shows the technical roadmap of the study.

## 4. Results

Through data analysis, it was found that online tobacco marketing information exhibited the characteristics described below in terms of volume, platform, and network.

### 4.1. Volume characteristics of tobacco marketing information

Volume refers to the number of online discussions about a topic. Statistical analysis on the volume of a topic can reveal the degree of attention, trend of changes, and other information, making it an important indicator for public opinions. According to the statistics of the collected data, the total volume of online tobacco marketing in the first half of 2018 was 106,485. More specifically, volume on behaviors such as "smoking" and "smoke a cigarette" were 12,165 and 8,299, respectively. Regarding the type of tobacco, the volume of "flavor capsule" was the highest (21,926), followed by "slim cigarette" (7275) and "low tar" (1569). For brand, the two foreign brands with the highest volumes were "Marlboro" (20,018) and "Esse" (11,869). When it came to domestic brands, the highest volume was for "Zhonghua" (6321), whereas those with the lowest volumes were "Haomao" and "Changbaishan" (388 and 310, respectively). In terms of company, the volumes of "Yunnan

Zhongyan" and "Sichuan Zhongyan" were also low.

#### 4.1.1. Period characteristics

During this study, an average of 17,748 pieces of tobacco marketing information were released on the Internet each month. The respective volume of online tobacco marketing information from January to June was 12,426; 8868; 21,701; 19,902; 24,051; and 19,537, respectively. The data displayed an overall increasing trend, with February's volume having the lowest ebb, March having a sharp rise in advertisements, and the volume peaking in May. The low ebb in February may be associated with the holiday period of industries with sales and express delivery decreased during the Spring Festival. Alternatively, the peak in May may be related to the escalating popularity of tobacco-related topics when approaching the "World No Tobacco Day". Nevertheless, "World No Tobacco Day" did not inhibit the release of online tobacco marketing information.

#### 4.1.2. Regional characteristics

Based on information from Weibo, among those publishing marketing information containing the monitored keywords, regions of relevant activities were classified into 4 levels: (1) Guangdong, Beijing, and overseas, the activities of which were >6000 people; (2) Seven provinces and cities, including Shandong, Anhui, Jiangsu, Zhejiang, Fujian, Shanghai, and Henan, the activities of which were between 3000 and 4500; (3) Nine provinces and cities, including Hunan, Sichuan, Hubei, etc., the activities of which were between 1500 and 3000; and (4) Sixteen provinces and cities, including Gansu, Jiangxi, Guizhou, etc., the activities of which were below 1500. These results strongly correlated with the netizen population, Internet penetration, level of e-commerce development, and overall degree of economic development of provinces and cities in China [28].

#### 4.1.3. Gender characteristics

According to the media report, of the 308 million Chinese smokers over 15 years of age, approximately 296 million were male, and the remaining 12 million were females [29]. Although the number of smoking women is low, there is a potential, significant upward trend among young women, making them the target marketing group of tobacco companies [1]. Based on the collected data, women were more active in posting tobacco marketing information on Weibo than men (61,366:43,928), especially those information containing such keywords as "smoking", "smoke a cigarette", "flavor capsule", and "slim cigarette." Compared to

domestic brands, they preferred two foreign brands of "Marlboro" and "Esse", especially the latter which targets women. The Baidu indices of relevant keywords suggested that although the search proportion[1] and target group index[2] of women were slightly lower than men, their attention on "Esse" was substantially higher. These results indicate that women, although not the main consumers of tobacco, have published or searched for a lot of tobacco related information and become an important group involved in online tobacco marketing, either as tobacco sellers (many women are in WeChat businesses or work as purchasing agents) or (potential) users.

### 4.1.4. Age characteristics

While young women showed a growing interest in tobacco, up to two-thirds of Chinese men started smoking during their adolescence [1]. Since both utilization and acceptance levels of the Internet are high among adolescents, once lured by tobacco marketing, they are more susceptible to developing a smoking habit. Therefore, this age group has become the target of both the "Pro-tobacco" and "Anti-tobacco" parties. Due to technical and platform limitations, we couldn't acquire age data for online tobacco marketers, and consequently the Baidu index was adopted to illustrate the participation characteristics of netizens of different age groups. The "Age distribution" from Baidu indices on related tobacco keyword searches suggested that the proportion of searches conducted by netizens aged between 20 and 29 was the highest, and the proportion of searches conducted by netizens below 19 years old only ranked second, their target group index was the highest. In addition, both age groups demonstrated an obvious preference for the two foreign brands, which indicates that netizens who search for tobacco information are predominantly young people.

### 4.2. Platform characteristics of tobacco marketing information

As a digital infrastructure, Internet platforms allow users to build their own products, services, and markets as well as interact with other groups [30,31]. In this study, platforms disseminated tobacco marketing, connected various tobacco marketing entities, promoted consumption behaviors, and established and strengthened tobacco culture. In this study, 11 Internet platforms targeted at different users were comprehensively analyzed.

### 4.2.1. Volume of each platform

The user scale, activity level and interactivity of a platform were all important factors affecting volume. In the first half of 2018, online tobacco marketing showed significant platform preferences, which can be classified into three levels according to volumes. The first level is Weibo with a volume of (66,350), which was substantially higher than the other 10 platforms combined (40,135). The second level contained five social media platforms, including Baidu Search (14,526) and Baidu Tieba (8,518), Xiaohongshu (4,728), WeChat (3,102), Yanyue Forum (2,628), and Zhihu (1,768). The third level contained the four mainstream tobacco industry portal websites, the total volume was 4865. These results reflect the "Network effects," i.e., "The more users a platform has, the more valuable it is to other users." On this basis, larger volume can be created by introducing more users [31].

---

[1] Search proportion = number of people with a certain search feature in the target group/total number of people with the same search feature.

[2] Target group index ("TGI") = (the proportion of people with a certain feature in the target group/the proportion of people with the same feature in the overall population) × standard number 100. A TGI over 100 indicates that the target group pays more attention to a certain type of issues than the overall population.

### 4.2.2. Content characteristics of each platform

The content analysis results suggested that there were 43,443 examples of explicit marketing and 63,024 examples of implicit marketing information in the analyzed data. Different platforms showed different ratios between explicit and implicit information. The proportion of explicit marketing information was the highest on Weibo, reaching 59.7%. This was more obvious for information related to "Marlboro," "Esse," "Zhonghua," and "flavor capsule," where sellers often disclose purchase channels, communicate with consumers, and complete transactions in private chats, or leave contact information such as QQ numbers, WeChat accounts, QR codes, etc. on the website. Alternatively, of the four mainstream tobacco industry websites, other than the China Cigarette Website (https://www.cnxiangyan.com/), the explicit marketing information was about half (48.0%) of the total, and the other websites only showed an explicit marketing content between 10% and 20%. However, although these websites do not offer direct purchase channels, they provide comprehensive and authoritative introductions to the brands, types, characteristics and prices of various domestic cigarette brands. Similarly, on Baidu Search and other social media platforms, except for Xiaohongshu, explicit marketing information was over 10% (10.1%), and the remaining websites were all below 10%. Zhihu had the lowest proportion (2.8%). Of these platforms, the proportion of explicit marketing information was relatively high for the geo-oriented Tieba owing to its convenience for transactions in the same city. On the contrary, Zhihu usually releases this kind of information through Q&As, whereas Xiaohongshu has now become the "Marketing front for flavor capsule cigarettes for ladies." It is worth noting that some explicit marketing information hides behind implicit marketing information and even non-marketing information, making it difficult to be captured by general crawlers. For example, several tobacco marketing WeChat official accounts posted "Read the original article," "Mini-program," "QR code," and "Link" as the springboard to transfer users, where entry to shopping malls or selling and gifting activities are embedded in irrelevant content, making them invisible to general data mining. Therefore, in addition to the user scale and interactivity, a platform's infrastructure and user rules, such as the restrictions and supports for different types of marketing information, also affects the communication behavior.

### 4.3. Social network characteristics of tobacco marketing information

A social network is a set of relationships between different nodes. Through social network analysis, it is possible to quantitatively characterize various relationships, thereby identifying their structures and interpreting a specific social phenomenon [32]. Investigating the social networks of online tobacco marketing information and platforms can help understand the status of various communication entities as well as the direction and intensity of relevant information flows.

### 4.3.1. Tobacco website

In this study, due to the significant centralization feature of mainstream tobacco industry portal websites such as Oriental Tobacco Website (https://www.eastobacco.com/), Tobacco Online (http://m.tobaccochina.net/), Yanyue Website (https://www.yanyue.cn/tobacco) (excluding the forum), and China Cigarette Website (https://www.cnxiangyan.com/), their tobacco marketing information was controlled by the website editor and released unidirectionally. Other than a few comments from netizens, interactions within the platform were not active. Therefore, it is more meaningful to investigate the social network between different websites than that of a specific website. We used these 4
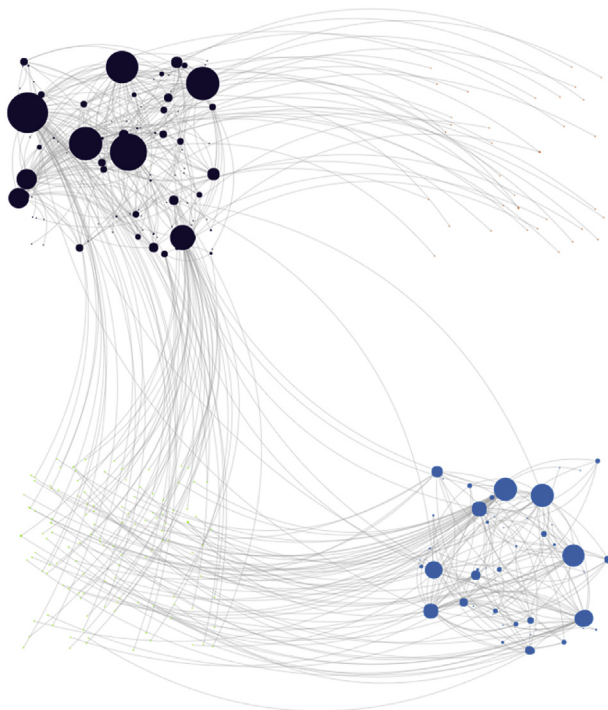
Fig. 2. Overall network structure of tobacco websites
(*Note*: black denotes traditional tobacco websites; blue denotes novel tobacco websites; red denotes relevant government websites; and green denotes relevant non-government websites.)

mainstream portal websites of the tobacco industry as the starting points, followed by exploring websites which have friendly links with them, and so on so forth. Owing to relevance and workload, the "Nominalist" strategy was adopted, so that the collection of this part of relational data would terminate at local or foreign tobacco and non-tobacco websites to define the network boundary. On this basis, a total of 84 websites were identified.

4.3.1.1. *Overall network structure.* During the social network analysis of tobacco websites, each website was considered a node. An edge was then created when there was a friendly link between two websites. From the ratio of the actual number of edges to the number of lines of a complete graph (that is, all points are connected), the density of this social network was calculated to be 0.00686, indicating a low degree of overall correlation between different nodes. In view of restrictions on the collection and selection of relational data as well as the effect of website type on its social interactions, the nodes of traditional tobacco websites, novel tobacco websites represented by e-cigarettes and heat-not-burn cigarettes, government websites, and non-government websites were manually labelled as black, blue, red, and green, respectively, which are then visualized in Fig. 2. As shown in Fig. 2, the subgraphs formed around the traditional and the novel websites display much larger densities than the overall network. This is because traditional tobacco websites were predominantly linked to government websites such as the Ministry of Finance, Ministry of Industry and Information Technology, whereas novel tobacco websites were connected with non-government websites that provide services such as wedding products and advertisements. Therefore, the two types of websites formed two separate "circles" with minimal interactions. The only connections were a few shared friendly links with some non-governmental, but influential websites such as Yanyue Website (https://www.yanyue.cn/tobacco) and China Cigarette Website (https://www.cnxiangyan.com/).

Therefore, traditional tobacco websites have more intimate relationship with related government agencies, while novel tobacco websites have more social interaction with the common public. Accordingly, different strategies and measures should be used to prevent the marketing behaviors of the two types of tobacco websites.

4.3.1.2. *Individual network structure.* An individual network structure refers to the network relationship between a certain node and multiple other associated nodes [33]. Measures in this study were based on degree centrality supplemented by closeness centrality. The degree centrality of the friendly link relational data was calculated by categorizing the data into either "linked" or "linking," which corresponded to indegree and outdegree, respectively. As shown in Fig. 1, the higher the degree centrality of a node, the larger its area, which indicates that this node had attracted more attention or is more active, and is closer to the center of the overall network structure. On the contrary, if the degree centrality of a node is low, it is more peripheral and located towards the edge.

As shown in Table 1, the China Tobacco Science and Education Website had the highest degree centrality among the traditional tobacco websites, whereas for novel tobacco websites, it was E-Cigarette Home. Except for the Yanyue Website, the other websites selected as platform samples, such as Tobacco Online (China Cigarette Website), and Oriental Tobacco Website were all on the top ten list. These tobacco websites are all very active in their daily operation and are consciously expanding their impact among the public, which should be paid more attention by tobacco control agencies. It is worth mentioning that the China Tobacco Monopoly Bureau Website demonstrates the highest relative indegree, suggesting that this node exhibits the largest influence and acts as an opinion leader. The result accords with China Tobacco Monopoly Bureau's unified, vertical and monopolistic tobacco management system, and also points out its unique position in regulating online tobacco marketing. Alternatively, closeness centrality refers to the average of the shortest distance between a specific node and all other accessible nodes within the network structure. The top three websites with the highest closeness centrality are: China Tobacco Monopoly Bureau Website (0.04638), China Tobacco Science and Education Website (0.04252), and Oriental Tobacco Website (0.03925). The larger the closeness centrality, the shorter the distance between this node and other accessible nodes, and therefore the faster the information communication. As the information dissemination hub, the three websites played a big role in the competition between the pro-tobacco and anti-tobacco discourse.

4.3.2. *Baidu Tieba*

Baidu Tieba was launched in 2003. It was selected as a representative sample of social media in this study because it comprises the world's largest online Chinese community, and one of the platforms where online tobacco marketing entities interact frequently. It should be noted that only the relational data of Baidu Tieba associated with the relevant keywords were used, and therefore the calculated results could be smaller than the reality. However, these data were still of great significance for analysis.

4.3.2.1. *Overall network structure.* As shown in Fig. 3, compared to the tobacco websites, the network structure of tobacco marketing information on Baidu Tieba was more decentralized, displaying an arrangement similar to the "Multi-centric pattern of difference sequence", where most of the users were on the edge and periphery of the network [34]. User interactions on relevant keywords happened in tobacco-themed forums, and multiple other geo-oriented and interest-oriented forums. Despite a large total number of nodes and edges in the social network, its density was low

**Table 1**
Top ten Chinese tobacco websites with the highest degree centrality.

| Ranking | Tobacco Website | Absolute Indegree | Relative Indegree | Absolute Outdegree | Relative Outdegree | Degree Centrality |
|---|---|---|---|---|---|---|
| 1 | China Tobacco Science and Education Website (http://www.tobaccoinfo.com.cn/) | 8 | 0.04082 | 33 | 0.16837 | 0.20919 |
| 2 | Shangpin Cigar Website (http://www.xuejia.mobi/) | 1 | 0.00510 | 27 | 0.13776 | 0.14286 |
| 3 | China Tobacco Information Website (http://www.echinatobacco.com/) | 4 | 0.02041 | 22 | 0.11224 | 0.13265 |
| 4 | Tobacco Online (http://m.tobaccochina.net/) | 0 | 0.00000 | 26 | 0.13265 | 0.13265 |
| 5 | China Cigarette Website (https://www.cnxiangyan.com/) | 3 | 0.01531 | 22 | 0.11224 | 0.12755 |
| 6 | Chinese Tobacco Association Website (http://www.tobacco.org.cn/) | 6 | 0.03061 | 15 | 0.07653 | 0.10714 |
| 7 | Oriental Tobacco Website (https://www.eastobacco.com/) | 7 | 0.03571 | 12 | 0.06122 | 0.09693 |
| 8 | E-Cigarette Home (http://ecigm.com/) | 3 | 0.01530 | 16 | 0.08163 | 0.09693 |
| 9 | China Tobacco Monopoly Bureau Website (http://www.tobacco.gov.cn/) | 9 | 0.04592 | 8 | 0.04081 | 0.08673 |
| 10 | China E-Cigarette Information Website (http://www.ecig100.com/) | 2 | 0.01020 | 15 | 0.07653 | 0.08673 |

*Note*: The number of "linked" website connections is defined as the absolute indegree. The relative indegree is defined as the ratio of the absolute degree to the maximum accessible degree. The absolute and relative outdegrees are defined with the same approach. The degree centrality is defined as the sum of the relative indegree and the relative outdegree.
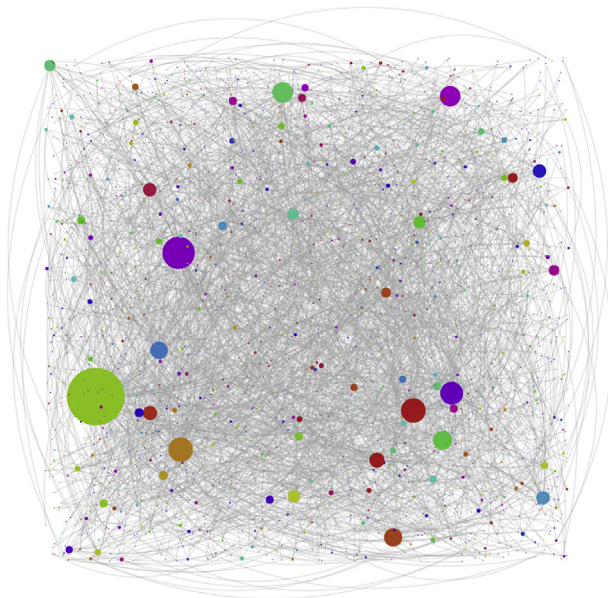


Fig. 3. Overall network structure of Baidu Tieba (tobacco-related interactions).

(0.00031), indicating that the overall degree of correlation between different nodes was extremely small. Given the low influence of these nodes on the overall network, more effective measures will be needed to block and filter the tobacco marketing information in Baidu Tieba.

*4.3.2.2. Individual network structure.* Relational data of Baidu Tieba were acquired through post reply behaviors, where "replied" was defined as the indegree and "replying" as the outdegree. The top user who had the highest degree centrality was "Chaobendezhu", whose activity was completely composed of indegree, indicating that discussions raised by his threads attracted a large attention. Alternatively, despite his relatively low degree centrality, user "Zhuaizhuaizhuaizhuiamoaxian6," ranked first in absolute outdegree, suggesting that he was enthusiastic about discussing these topics. In addition, the top three users with the highest closeness centrality at different nodes of Baidu Tieba were: "Chaobendezhu" (0.02888), "Xiaojiongrita" (0.02087), and "Yifanfengshung0" (0.01779), the interactions between these nodes and other accessible nodes were smoother and faster (refer to Appendix A for detailed data). Furthermore, the local social network formed by user "Chaobendezhu" and other associated users was located

towards the center within the overall network structure (refer to the local network structure diagram in Appendix B). These top users were likely to be the leaders or members of some tobacco marketing groups, and should be paid more attention when regulating the online tobacco marketing information.

The above social network analyses of tobacco websites and Baidu Tieba were refined to static investigations within the period from January to June 2018, and therefore the results cannot reflect social interactions beyond the specific period and keywords.

## 5. Discussion and conclusion

Long before the Internet, various marketing entities had already launched one-to-one, one-to-many, and many-to-many tobacco marketing strategies and disseminated implicit or explicit marketing information. The emergence of the Internet incorporated these activities, and facilitated their unification, integration, and development, creating an unprecedented, complex situation. Through big data and social network analysis, this study comprehensively analyzed the volume characteristics, platform preferences, and social networks of online tobacco marketing in China. Whether from the perspective of research orientation, object, or method, this study has supplemented the research related to online tobacco marketing, and provided support for tobacco control. The main conclusions of the study are as follows.

First, the total volume of internet tobacco marketing information between January and June 2018 was relatively high and displayed an increasing trend within the period. More specifically, information associated with "flavor capsule," "Marlboro" and "Esse" demonstrated a higher volume. In the east and central regions where the number of netizens, Internet penetration rate and e-commerce development were high, tobacco marketing entities were more active. In terms of gender, women's enthusiasm for posting tobacco marketing information or searching for relevant information online was comparable to that of men, although their attention on "female cigarette" brands were substantially higher. In terms of age, netizens under the age of 29 demonstrated the highest interest in searching for tobacco-related information, which could include a significant number of youths.

Second, the Weibo platform had the highest volume of online tobacco marketing, and the largest proportion of explicit marketing information. This was followed by other social media platforms, such as Baidu Search and Baidu Tieba, Xiaohongshu, WeChat, Yanyue Forum, and Zhihu, where implicit marketing information was dominant. On the contrary, the volumes of the mainstream portal websites of the tobacco industry, such as the Oriental

Tobacco Website, Tobacco Online, the Yanyue Website (excluding the forum), and the China Cigarette Website were the lowest, with an intermediate ratio between explicit and implicit marketing information.

Third, within this study, the overall network structure of tobacco websites exhibited a prominent centralization feature, where traditional tobacco websites and novel tobacco websites formed two separate clusters with minimal interactions. The China Tobacco Science and Education Website and E-Cigarette Home were the two nodes with the highest degree centrality in each respective "Circle", while the China Tobacco Monopoly Bureau Website showed the highest closeness centrality. In comparison, the overall network structure of Baidu Tieba was more decentralized, and the overall degree of correlation between different nodes was low. User "Chaobendezhu" ranked first in both degree centrality and closeness centrality and was located at the center of the social network, indicating that he was an opinion leader on tobacco marketing topics in Tieba.

In summary, online tobacco marketing demonstrated high volumes and wide coverages, and intertwined in a network, thereby creating major obstacles for tobacco control. Based on the above findings, the following three strategies were proposed to facilitate the implementation of targeted government regulation of online tobacco marketing:

(1) Regarding government regulation, we should strengthen supervision, enforce implementation, and improve regulation. First, within the existing framework of laws and regulations, technical means should be adopted to supervise online tobacco marketing behaviors, and in particular monitor Weibo and other social media, to identify active accounts on the Internet and accurately target illegal activities. This will increase contraband costs. In April 2019, the Beijing Center for Diseases Prevention and Control stated it would "Deeply analyze surveillance data, collect representative cases, and routinely announce results" [35]. In November 2017, Beijing Chaoyang District Trade and Industry Bureau issued a fine of 1 million RMB to the WeChat official account "Wanmijulebu" [36] for tobacco marketing, which is worth highlighting. Second, the application scope and legal hierarchy of relevant regulations should be improved based on the FCTC. For example, the Advertising Law should explicitly prohibit "Publishing any sort of tobacco advertisements" on the Internet; the definition of implicit (or disguised) marketing information should be clarified and penalties against it strengthened; and "Interim Measures for the Administration of Online Advertising," should be formulated, and relevant clauses prohibiting tobacco advertising should be refined (such as the respective responsibilities of the account holder and the platform) and their legal effectiveness elevated. Third, the prohibition of online tobacco marketing should be incorporated into local tobacco control regulations, so that issues involving regional tobacco administration and territorial jurisdiction of relevant departments can be addressed. In addition, eastern and central regions, where tobacco marketing is more active, should be listed as the key regulatory areas. Lastly, associated prohibitions should be included in policies and regulations concerning the health of women and youths, and priority support should be provided to corresponding tobacco control activities.

(2) From the perspective of platform self-discipline, we should comprehensively remove tobacco marketing information, establish screening criteria, and improve regulations. First, the platform's self-discipline should be supported, and existing tobacco marketing information should be completely cleaned up, so that the online volume of Weibo and other social media can be effectively reduced. In April 2019, after the release of a large amount of tobacco marketing on Xiaohongshu, the website stated it "Opposes any sort of tobacco-related communication" and has thereafter removed any tobacco-related posts [37]. Second, in light of the unpredictable and negative impact of tobacco marketing once published, platforms should follow the "Red Flag Principle" to pre-alert and prevent any tobacco marketing rather than deleting it after the fact. Third, in order to maximize the interests of youths, technical means should be adopted to restrict adolescents from viewing any tobacco advertisements. For example, some traditional tobacco websites display "No access to people under 18 years old" in the corner of the page or in a pop-up window. However, despite these warnings, there is no actual measure preventing minors from accessing the website. To protect youth, we recommend implementing screening technologies, or restricting the dissemination of relevant information to within the industry or the company. Furthermore, as "Opinion leaders", websites such as the China Tobacco Science and Education Website, China Tobacco Monopoly Bureau Website, and Oriental Tobacco Website should set an example by following regulations. Lastly, relevant actions should make up the platform rules. For example, the "Administration Rules of Prohibited Information on Taobao Platform" issued in 2019 listed "Tobacco products and devices" as prohibitions.

(3) From the perspective of news media, public welfare organizations, and public participation, we should advocate policies, strengthen platform supervision, improve health communication strategies, and establish a health culture against tobacco. Although government regulation and platform self-discipline can curb online tobacco marketing in the short term, over the long term, it is necessary to engage news media, public welfare organizations, and the public. This affects policy enforcement and platform rules, and facilitates the long-term establishment of a healthy culture. In addition, to advertise against online tobacco marketing, various communication strategies should be adopted flexibly according to the characteristics of the platform. By utilizing multiple anti-tobacco marketing approaches such as news reports, public welfare communications and daily advertisements, a sustainable, healthier culture can be cultivated.

This study conducted a comprehensive investigation of online tobacco marketing information in terms of volume, platform and network. This study adds to our understanding of tobacco marketing through an innovate look at how tobacco marketing and sales occur on websites and social media, despite policies that ban tobacco advertising. Although the results of this study outlined strategies for online tobacco control, broader success depends on weakening tobacco marketing campaigns through regulatory methods including pricing, taxation, and industrial transformation. Future studies should verify the above conclusions by investigating other periods, platforms, and keywords. Also, further research could focus on a certain platform or brand and conduct a more in-depth and exhaustive case study on the daily interactions between smokers online and offline, further exploring the processes and mechanisms of their discourse and social practices.

## Funding

collection, analysis, interpretation of data, writing of the report, or decision to submit the article for publication.

## CRediT authorship contribution statement

**Xiuli Wang:** Conceptualization, Methodology, Writing - review & editing. **Ying Xiong:** Writing - Original Draft, Visualization. **Wenwen Zhao**: Data Curation.

## Declaration of competing interest

All authors declare no other financial relationships with any organizations that might have an interest in the submitted work in the previous three years; no other relationships or activities that could appear to have influenced the submitted work.

## Acknowledgement

## Appendices. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ijnss.2020.07.002.

## References

[1] World Health Organization Regional Office for the Western Pacific. China's unbearable costs: health, economic, and social losses brought by the prevalence of tobacco in China. 2017. License No. CC BY-NC-SA 3.0 IGO.
[2] Chinese Center for Disease Control and Prevention. Chinese Adult tobacco consumption survey. 2018. 31st May 2019. [in Chinese].
[3] China Internet Network Information Center. reportThe 44th statistical report on internet development in China. 30th August 2019. [in Chinese]
[4] China Internet Network Information Center. Chinese social application user behavior research report. 2016. 27th December 2017. [in Chinese].
[5] Cheng C. Analysis of the outlet of tobacco advertising after the revision of the advertising law. In: Proceedings of the Shanghai Tobacco Association Conference; August 2016. p. 70—3 [in Chinese].
[6] Jiang W. Research on the "Internet +" business strategy for commercial tobacco companies. Mod Bus 2016;11:132—4. https://doi.org/10.3969/j.issn.1673-5889.2016.11.075 [in Chinese].
[7] Lyu C. Novel marketing models promote tobacco development — research on the informatization of cigarette marketing in the "Internet +" environment. Thematic Compilation of the Chinese Tobacco Association; 2016 [in Chinese].
[8] Zhou Y. Research on the marketing strategy of X tobacco company. Master Thesis of. Heilongjiang University; 2017 [in Chinese].
[9] Zhao J. A preliminary study on Internet + smart tobacco marketing. Sci Technol Innov Herald 2015;25:249—50. https://doi.org/10.3969/j.issn.1674-098X.2015.25.119 [in Chinese].
[10] Song N, Han F, Qu X. Research on the implementation and application of company micro-marketing strategies: taking the tobacco industry as an example. China Collective Econ 2016;3:68—71. https://doi.org/10.3969/j.issn.1008-1283.2016.03.034 [in Chinese].
[11] Kong J. Research on the Weibo marketing strategy of China's top companies in the tobacco industry — taking China Tobacco Fujian Industrial Co., Ltd.. as an example. Master Thesis of. Fuzhou University; 2013 [in Chinese].
[12] Cao X, Jin Xianglan. Tobacco advertisement challenges legal blind spots by "disguise". Leg Econ 2013;10:20—1 [in Chinese].
[13] Yu H, Ma X. Return of the "life" logic: big data analysis of the China tobacco
[14] report and its inspiration. Mod Commun 2019;8:41—9. https://doi.org/10.3969/j.issn.1007-8770.2019.08.009 [in Chinese].
[14] Lyu Y. Research on the investigation of illegal online tobacco sale cases. People's Public Security University of China; 2017 [in Chinese].
[15] Ma Y, Wang Yongzhong, Wei Ping. Strengthen top-level design and improve regulation quality: countermeasures for tobacco monopoly regulation on the we-media platform. 2015 compilation of outstanding papers of the Chinese Tobacco Association. 2015. p. 2465—75 [in Chinese].
[16] Jiang R. Research on the regulation problems and countermeasures of tobacco-related illegal acts on the internet: taking fuzhou chenbei tobacco monopoly Bureau as an example. Master Thesis of. Fujian Normal University; 2018 [in Chinese].
[17] Cohen JE, Sarabia V, Ashley MJ. Tobacco commerce on the internet: a threat to comprehensive tobacco control. Tobac Contr 2001;10(4):364—7.
[18] Qi F, Li S, Jia X, Wang Y, Liu H, Lin P, Geng M, Yang J. Investigating the tobacco consumption of junior high school students in Qingdao and their exposure to tobacco advertising and promotion. Chin J Public Health 2016;10:1314—8. https://doi.org/10.11847/zgggws2016-32-10-05 [in Chinese].
[19] Shi F, He Y, Xu G, Zhu J. Research on the status and application progress of new media in the field of tobacco control for adolescents. J Shanghai Jiaot Univ (Med Sci) 2019;2:213—6. https://doi.org/10.3969/j.issn.1674-8115.2019.02.020 [in Chinese].
[20] Cavazos-Rehg PA, Krauss MJ, Spitznagel EL, et al. Hazards of new media: youth's exposure to tobacco Ads/promotions. Nicotine Tob Res 2014;16(4):437—44.
[21] Dunlop S, Freeman B, Perez D. Exposure to internet-based tobacco advertising and branding: results from population surveys of Australian youth 2010-2013. J Med Internet Res 2016;18(6):e104.
[22] Soneji S, Pierce JP, Choi K, et al. Engagement with online tobacco marketing and Associations with tobacco product use among U. S. Youth. J Adolesc Health 2017;61(No. 1):61—9.
[23] Vishwanathan HK, James L. Gender and tobacco marketing in social media: how celebrity smoking culture is conveyed and perceived by young women. Can J Cardiol 2018;34(4):e1.
[24] Qian H, Wang F, Sun S. From the perspective of media literacy: a study on the knowledge, belief and behavior of smoking among middle school students in Shanghai. In: Conference proceedings of the tsinghua university international center for communication; 2019. p. 146—58 [in Chinese].
[25] Qu H, Liu Z, Long J, Xu Y. Design and implementation of a prototype system for the automatic monitoring of tobacco-related online illegal activities. In: Conference proceedings of the Chinese association of automation; 2013. p. 110—4 [in Chinese].
[26] Wu S, Zhang X, Wang L, Yin H, Chen L, Fu W. Positive impacts of new media on people's attitudes towards tobacco prevention/control. China J Health Psychol 2017;1:88—92. https://doi.org/10.13342/j.cnki.cjhp.2017.01.022 [in Chinese].
[27] Kotler P, Keller KL. Marketing management. fifteenth ed. Edinburgh Gate: Pearson Education Limited; 2016.
[28] Tsinghua University National. Engineering laboratory for E-commerce technologies, China center for internet economy research of central university of finance and economics, Chinese academy of social sciences evaluation, China international electronic commerce center of Ministry of commerce, yibang power research institute. China e-commerce development index report (2018). 2019 [in Chinese].
[29] Wu B. Whilst decreased by 8 million, the total number of smokers in China is over 300 million. Expert: must follow up on tobacco tax and price increases. Southern Metropolis Daily; 31st May 2019 [in Chinese].
[30] Gawer A. Platform dynamics and strategies: from products to services. In: Annabelle Gawer A, editor. Platforms, markets and innovation. Cheltenham: Edward Elgar; 2009. p. 45—76.
[31] Srnicek N. Platform capitalism. Cambridge: Polity; 2017.
[32] Scott J. Social network analysis. fourth ed. Thousand Oaks, CA: Sage; 2017.
[33] Wu J, Shi L. Research on the user interaction in online medical communities based on social network analysis. Inf Sci 2017;7:120—5. https://doi.org/10.13833/j.cnki.is.2017.07.021 [in Chinese].
[34] Wang Y, Liu Y. Research on the WeChat information communication network of public welfare organizations based on social network analysis. J Mass Commun Monthly 2019;8:21—6. https://doi.org/10.15897/j.cnki.cn51-1046/g2.2019.08.004 [in Chinese].
[35] Beijing Center for Disease Prevention and Control. Press release of 2018 online tobacco marketing data surveillance results in China [in Chinese], http://www.bjcdc.org/article/56545/2019/4/1555568179240.html; 18th April 2019.
[36] Xie J. Beijing Chaoyang District Trade and Industry Bureau issued a fine of 1 million RMB to illegal tobacco advertising. sixth ed. China Business News; 2017-11-28 [in Chinese].
[37] Zhang X. Xiaohongshu completely removes all tobacco-related soft articles. A04th edition. Beijing Youth Daily; 2019-04-17 [in Chinese].