Nature and Diffusion of Gynecologic Cancer-Related Misinformation on Social Media

Abstract

**Background:** Over last two decades, the incidence and mortality rates of gynecologic cancers have increased in China and have become one of the most serious threats to women’s health there. With the widespread use of the Internet, an increasing number of individuals use social media to seek out and share a large amount of cancer-related information. However, health information on social media is not always accurate. Health, and especially cancer-related, misinformation has been widely spread on social media, which can affect individuals’ responses to cancer.

**Objective:** The aim of this paper was to examine the nature and diffusion of gynecologic cancer-related misinformation on Weibo, the Chinese equivalent of Twitter.

**Methods:** A total of 2,691 tweets related to gynecologic cancers posted on Weibo from June 2015 to June 2016 were extracted using the Python Web Crawler. Two medical school graduate students with expertise in gynecologic diseases were recruited to code the tweets in order to differentiate between true and false information as well as to identify the types of falsehoods. Next, a social network analysis was used to examine the diffusion cascades of misinformation through comparisons with true information.

**Results:** The findings revealed that while most of the gynecologic cancer-related tweets provided medically accurate information, approximately 30% of them contained misinformation. Furthermore, the results showed that tweets about cancer treatment contained
a higher percentage of misinformation than prevention-related tweets. The prevention-related misinformation diffused significantly more broadly and deeply than true information on social media. **Conclusions:** The findings of this study indicate that social media service providers and medical professionals should evaluate online resources regularly to control the production and diffusion of cancer misinformation on social media. More specifically, it is important to correct falsehoods related to the prevention of gynecologic cancers on social media and increase individuals’ abilities to assess the veracity of online information in order to limit the spread of prevention-related misinformation and to minimize the consequences of cancer-related misinformation. **Keywords:** social media; gynecologic cancers; misinformation; diffusion; China

**Introduction**

In recent years, cancer has become a major public health issue in China. According to data from the National Central Cancer Registry (2017), there were approximately 3.80 million new cancer cases and 2.30 million deaths due to cancer in China in 2014 [1]. For women, 2 out of the 10 most common cancers are gynecologic cancers, with breast cancer
(268.6 thousand new cases) and cervical cancer (98.9 thousand new cases) being the most prevalent\[2\]. The incidence and mortality rates of gynecologic cancers have increased substantially in China over last two decades, and it has become one of the most serious threats to women’s health there\[2\].

With the rapid development of the Internet, social media has become a popular means by which individuals can access a mind-blowing amount of health information. A growing number of individuals, especially women, turn to social media to seek out and share a variety of cancer-related information, such as information about cancer prevention and treatment and the experience of having it and to obtain social support in order to cope with the disease and manage emotions\[3,4\]. Medical professionals and traditional portals contribute to the health information available on social media, but a greater amount of information is generated and disseminated by ordinary users based on their subjective first-hand cancer experiences\[5\].

Online health information has been found to be effective in raising individuals’ awareness of diseases and fueling communication between lay persons and healthcare professionals\[6\]. Furthermore, online health information could help individuals increase their abilities to prevent certain diseases and enable them to effectively manage chronic health conditions\[7,8\]. Nevertheless, when using online resources, individuals might take great risks, as health information on social media is not always correct\[9,10\]. It has been shown that health-related, and cancer-related misinformation in particular, has been widely spread on social media, which affect individuals’ responses to cancer prevention and treatment\[11\]. Moreover, due to information overload, ordinary users may not have the knowledge and expertise needed to assess the veracity of cancer-related information online and to identify informative and trustworthy information on social media\[11\]. Although scholarly attention has been drawn to misinformation on social media, little
is known about the nature and diffusion of cancer-related misinformation there. In this study, we filled this gap in knowledge by examining the nature and diffusion of misinformation about breast cancer and cervical cancer on social media. Specifically, we used a content analysis not only to differentiate between true and false cancer information on social media but also to identify the types of falsehoods. Furthermore, the diffusion cascades of cancer-related misinformation were examined and compared with the diffusion cascades of true information using social network analysis.

**Misinformation on Social Media**

Misinformation refers to false and inaccurate information that is spread intentionally or unintentionally; the term overlaps with many others, such as fabricated information, fake news, and distorted evidence [12]. With the widespread use of information and communication technologies (ICT), massive amounts of misinformation have become pervasive on the Internet. One study revealed that the average person in the U.S. encountered one to three fake news stories online in the month before the 2016 U.S. election [13]. About one-quarter of adults reported having shared fabricated political news online, sometimes by mistake and sometimes intentionally [14].

Many scholars have argued that social media is responsible for the high prevalence of online misinformation. Traditional media content is usually generated by professional journalists and editors who have the knowledge and resources to assess the veracity of information and to report news objectively, while social media content is user-generated, so each user is able to create and exchange a variety of information irrespective of its veracity [15,16]. Thus, it is not surprising that there is a large amount of misinformation on social media. The proliferation of online misinformation has caused negative consequences for
individuals and society. Specifically, misinformation such as fake news, rumors and inaccurate information not only causes the spread of unnecessary fears and conspiracies but also distorts individuals’ behavioral responses to certain issues, such as elections, natural disasters, and diseases \cite{14,17}. For example, misinformation about vaccinations makes many parents refuse immunizations for their children, which has led to a noticeable increase in vaccine-preventable diseases and has even caused deaths among children \cite{18}. Furthermore, misinformation exerts negative impacts on our society that may trigger financial panic and even strain diplomatic relations \cite{19}. The 2013 World Economic Forum listed misinformation as one of the main threats to human society \cite{15}.

To limit the amount of misinformation available and to minimize the negative effects of misinformation, many researchers have attempted to examine how misinformation spreads on social media and investigate the driving mechanisms that underlie the diffusion of online misinformation in various domains, such as natural disasters, science and politics. Specifically, Oh, Kwon, and Rao \cite{20} analyzed the working dynamics of rumors related to the Haiti Earthquake in 2010 based on data from Twitter and found that informational uncertainty and anxiety are key factors that determine the rapid spread of a rumor. Moreover, they indicated that reliable information with credible sources could reduce levels of anxiety on Twitter, which in turn limits the spread of rumors. Domenico, Mougel, & Musolesi \cite{21} explored the spread of a scientific rumor about the Higgs boson and proposed a model for its spread. They found that individuals were more likely to spread the rumor if most of their friends tweeted it repeatedly. In other words, after observing that many friends have shared a specific message, individuals tend to trust the credibility of the information and thus become more likely to accept and share it. Del Vicarioa et al. \cite{15} investigated and compared the
spreading process of misinformation about scientific and conspiracy stories on Facebook. They found that social homogeneity and polarization are the main driving forces underlying misinformation diffusion. In other words, the information is usually adopted by a friend with a similar profile. More recently, a study aimed to understand the diffusion structure of true and false news on Twitter, and found that false news spread farther, faster, deeper, and more broadly than the truth. Such differences in the diffusions of truth and falsehoods may be related to that false news tends to include more sentiments of fear, disgust, and surprise, which could lead to the virality of the information. Another explanation could be a novelty effect, this is, that people are more willing to exchange novel information.

Although much attention has been directed towards the spread of misinformation on social media, most studies have focused on political or scientific rumors or misinformation in general. Only a few studies have examined the nature and diffusion of health misinformation on social media. The unique differences in the contexts of health information require a specific investigation.

**Health Misinformation on Social Media**

With the use of social media in healthcare, a growing amount of health misinformation that poses a great threat to public health has spread on social media. Few studies have attempted to explore the nature of health misinformation on social media. For example, Syed-Abdul et al. investigated anorexia-related misinformation disseminated in YouTube videos. Three physicians reviewed 140 of the most-viewed videos about anorexia. The results revealed that 55.7% of the videos provided true information about the consequences of anorexia or advice on how to recover from it, while 29.3% of the videos were categorized as pro-anorexia, which could mislead viewers since anorexia can be
detrimental to one’s health. In addition, Qi et al. [22] conducted a content analysis of 47 popular psoriasis-related videos on YouTube. They found that up to 21% of the videos included misleading information, and only 17% of the videos provided useful advice. In addition, some scholars have proposed and tested intervention strategies to correct health misinformation. First, Pluviano, Watt, and Sala [23] evaluated the effectiveness of existing intervention strategies for vaccine misinformation corrections, such as sharing myths versus facts, employing facts and icon boxes, and showing images of unvaccinated sick children. The results of their study revealed that all three intervention strategies were ineffective and even led to unintended opposite effects, including reinforcing ill-founded beliefs about vaccination and decreasing intentions to vaccinate. Vraga and Bode [10] developed a corrective intervention using expert sources on Twitter. They found that corrective responses from ordinary users did not decrease individuals’ misperceptions about the Zika virus, while a corrective response posted by the Centers for Disease Control and Prevention (CDC) did significantly reduce individuals’ misperceptions. Moreover, two corrective responses from both ordinary users and the CDC are most effective for reducing misperceptions about the Zika virus. Bode and Vraga [24] compared the effectiveness of Facebook algorithmic and social corrections in limiting misperceptions about the Zika virus. The results suggested that the corrections produced by an algorithm and online friends were equally effective in reducing public misunderstandings. While a few recent studies have examined the veracity of health information online and interventions to address health misinformation, these studies only focused on diet, vaccinations, and infectious diseases. Cancer, one of the most serious public health issues, has received limited scholarly attention.
Cancer Misinformation on Social Media

The rise of news media has created an atmosphere of hype and hysteria about cancer in which individuals have been exposed to conflicting information. This leads to many misperceptions about cancer, including its causes, prevention, and treatment [25]. In the last decade, social media has exacerbated individuals’ uncertainty about cancer. Unlike traditional media, most health-related content on social media is generated and shared by patients and caregivers based on their own personal illness experiences. The content may include many false elements that can distort individuals’ attitudes and behaviors towards cancer prevention and treatment. Gage-Bouchard et al. [26] conducted a content analysis to assess the veracity of information related to lymphoblastic leukemia on 19 public Facebook pages and found that at least one-third of the information exchanged was not medically/scientifically accurate.

With its extremely large population, China contributes almost one-quarter of the global cancer burden: 22% of the world’s new cancer cases and 27% of the world’s cancer deaths occur there [27]. However, cancer is preventable if people are aware of its causes and science-based prevention strategies (WHO, 2017). A growing number of Chinese people use social media to exchange a variety of information related to cancer, such as their personal cancer experiences and cancer prevention and treatment strategies [28]. Research has indicated that compared with men, women are more likely to seek out health information online [4]. In other words, cancer misinformation on social media may be a larger threat to women than men. As such, it is important to investigate the veracity of online information about gynecologic cancers. According to data from the National Central Cancer Registry of China [11], breast cancer (268.6 thousand new cases) and cervical cancer (98.9 thousand new cases) are the most common gynecologic cancers in China. Thus, this study focused on information
related to those two types of cancer on Chinese social media.

Based on the aforementioned literature, misinformation, especially health misinformation, has been prevalent on social media. This has drawn much scholarly attention in the West \[11\]. Some recent studies have examined the veracity of health information on social media. However, these studies were exploratory in nature, only examining the frequency of misinformation; they did not take into account the types of misinformation and their diffusion processes. Therefore, this study was conducted to fill in the gaps in the literature by examining what misinformation about breast cancer and cervical cancer exists as well as spreads on one of most popular Chinese social media sites, Weibo, a Chinese version of Twitter. Corresponding to the research objectives, three research questions were proposed:

Research question 1: What is the distribution of true and false gynecologic cancer-related information on Weibo?

Research question 2: What kinds of gynecologic cancer-related misinformation exist on Weibo?

Research question 3: How do the diffusion structures of true and false cancer-related information differ?

To answer our research questions, a content analysis was first conducted to differentiate between true information and misinformation and to identify the types of falsehoods present. Next, the retweets of each piece of cancer-related information on the network were mapped using the Python Web Crawler and were analyzed using a social network analysis to understand how misinformation was spread and accepted by social media users.

**Methods**

**Data Collection**

We used the keywords “乳癌/乳腺癌” [breast cancer] and “子宫癌/宫颈癌”
[cervical cancer] to retrieve tweets about breast cancer and cervical cancer on Weibo, a Chinese version of Twitter, that were written from June 2015 to June 2016, and data for seven weeks were randomly selected from the results. A total of 2,691 tweets from these seven weeks were extracted using the Python Web Crawler in July 2016. For each tweet, we crawled the content, post time, sender’s username, sender’s profile information, and diffusion path.

In terms of ethical issues, Weibo is considered a public domain in which data are freely accessible to the public. To minimize the potential harm to Weibo users, all of the data collected from Weibo were de-individualized in order to maintain the sender’s anonymity. Moreover, all of the messages presented in this manuscript were paraphrased or written in aggregate to prevent identification of the senders.

**Coding Procedure**

Two medical school graduate students with expertise in gynecologic diseases were recruited to complete the coding. Initially, two coders who were asked to pilot the project coded 10.41% (n = 280) of the total tweets in order to develop and refine the following coding categories: (1) background knowledge, which refers to basic information about breast cancer and cervical cancer, including each one’s prevalence, causes and symptoms; (2) prevention, which refers to methods and actions that can lower the risk of getting cancer, include maintaining a healthy lifestyle, avoiding exposure to known cancer-causing substances, and taking medicines or vaccines; (3) diagnosis, which refers to the act of identifying a disease from its signs and symptoms; and (4) treatment, which refers to drugs or methods that can attack specific types of cancer cells to help the patient fight the disease, in accordance with Lee’s (2011) guidelines. Next, two coders independently coded the full set of
2,691 messages. Specifically, the coders categorized each message as 0 (not applicable), 1 (true), and 2 (false) for the four initial categories. Krippendorff’s alpha tests (2007) revealed an acceptable level of inter-coder reliability for all of the variables: 0.91 for background knowledge, 0.88 for prevention, 0.93 for diagnosis, and 0.89 for treatment (Krippendorff, 2011). Finally, when a message was categorized into “false” in any category, two coders indicated the types of falsehoods using a conventional content analysis.

First, all of the false tweets were read repeatedly to achieve immersion and obtain a sense of the entire situation. Second, while reading the tweets, the two coders highlighted exact words from the text as codes or created new codes to capture key concepts. Third, these codes were sorted into categories based on their relationships. Thereafter, the second and third procedures were repeated to keep the acceptability and reliability of the designated categories high. Lastly, each category was defined. The validity of the coding was checked using a deviant case analysis.

**Information Diffusion**

For each tweet, we mapped the diffusion path of how the original tweet became retweeted by others. Each retweet cascade represents a falsehood or truth propagating on Weibo that was verified by the coders. We then quantified the cascades with multiple indices including the retweet amount, numbers of comments and likes, range, and structural virality, which measure the breadth and depth of the diffusion. More specifically, the retweet amount counts the number of retweets the original tweet received. The number of comments and number of likes indicates the level of engagement of the audience in the information diffusion process. Range is the depth of the information diffusion network as indicated by the number of hops in a diffusion chain. Structural virality measures the intuitive
distinction between broadcast and viral diffusion and allows for interpolation between them. These indices cover the breadth and depth of the diffusion process, which present a multiple-dimensional measurement of the prevalence of a piece of information on social media. It provides a more comprehensive understanding of the diffusion structures of information online.

**Results**

**Content Analysis**

Of the 2,691 total tweets on Weibo, 1,144 (42.51%) only expressed personal emotions and experiences about breast cancer and cervical cancer; these are neither true nor false. However, 1,547 (57.49%) tweets contained medically oriented information for which the veracity was evaluated. Specifically, out of the 1,547 medically oriented tweets, 1,023 (66.13%) tweets provided medically accurate information and 524 (33.87%) contained misinformation. A chi-squared test indicated that true information was significantly more prevalent than misinformation, \(X^2 (1, N = 1547) = 160.96, p < .001.\)

The 1,547 tweets were categorized into four different types: background knowledge, prevention, diagnosis and treatment. The most commonly exchanged type of cancer information was background knowledge (48.42%, \(n = 749\)), followed by prevention (30.19%, \(n = 467\)), treatment (12.21%, \(n = 189\)), and diagnosis (9.18%, \(n = 142\)) (see Table 1).

<table>
<thead>
<tr>
<th>Category</th>
<th>Frequency N (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Background Knowledge</td>
<td>749 (48.42%)</td>
</tr>
<tr>
<td>Prevention</td>
<td>467 (30.19%)</td>
</tr>
<tr>
<td>Treatment</td>
<td>189 (12.21%)</td>
</tr>
<tr>
<td>Diagnosis</td>
<td>142 (9.18%)</td>
</tr>
</tbody>
</table>
Table 2. 
*Truth and Falsehoods of Medically Oriented Cancer Information on Weibo.*

<table>
<thead>
<tr>
<th>Category</th>
<th>Truth N (%)</th>
<th>Falsehoods N (%)</th>
<th>Total N (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Background Knowledge</td>
<td>462 (61.68%)</td>
<td>287 (38.32%)</td>
<td>749 (100%)</td>
</tr>
<tr>
<td>Prevention</td>
<td>400 (85.65%)</td>
<td>67 (14.35%)</td>
<td>467(100%)</td>
</tr>
<tr>
<td>Treatment</td>
<td>33 (21.15%)</td>
<td>156 (82.53%)</td>
<td>189(100%)</td>
</tr>
<tr>
<td>Diagnosis</td>
<td>128 (90.14%)</td>
<td>14 (9.86%)</td>
<td>142(100%)</td>
</tr>
</tbody>
</table>

Note: Chi-squared tests showed that the frequencies are significantly different between true information and misinformation for each category. Tweets about treatment contained a higher percentage of misinformation than true information; specifically, 156 (82.53%) tweets related to cancer treatment included misinformation. Followed by tweets about background knowledge, 287 (38.32%) contained misinformation. Only 67 (14.35%) prevention-related tweets and 14 (9.86%) diagnosis-related tweets were not medically accurate (see Table 2).

Finally, as shown in Table 3, the types of falsehoods were identified and listed to categorize the true information and misinformation into four different categories. Specifically, the falsehoods in background knowledge mainly included epidemiology, risk factors, prognosis, and pathology. Prevention-related tweets had a relatively small amount of misinformation that involved three types of falsehoods: lifestyle and vaccinations. Diagnosis-related misinformation was divided into two types: clinical manifestations and diagnostic techniques and procedures. Cancer treatment-related misinformation mainly included surgery, radiation therapy, drug therapy, and other therapies (See Table 3).
<table>
<thead>
<tr>
<th>Category</th>
<th>Types of Falsehoods</th>
<th>Definition</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Background Knowledge</td>
<td>Epidemiology</td>
<td>The distribution and determinants of health and disease conditions in specified populations.</td>
<td>The cancer prevalence rate is 10% higher in China than the world average.</td>
</tr>
<tr>
<td></td>
<td>Risk Factors</td>
<td>An aspect of personal behavior or lifestyle, environmental exposure, inborn or inherited characteristic, which, on the basis of epidemiological evidence, is known to be associated with a health-related condition.</td>
<td>Using preservative-containing cosmetics is one of the main causes of breast cancer.</td>
</tr>
<tr>
<td></td>
<td>Pathology</td>
<td>A specialty concerned with the nature and cause of disease as expressed by changes in cellular or tissue structure and function caused by the disease process.</td>
<td>Breast hyperplasia is the beginning of breast cancer.</td>
</tr>
<tr>
<td></td>
<td>Prognosis</td>
<td>A prediction of the probable outcome of a disease based on an individual’s condition and the usual course of the disease as seen in similar situations.</td>
<td>Triple-negative breast cancer has a better prognosis than the normal type, and the five-year survival rate is high.</td>
</tr>
<tr>
<td>Prevention</td>
<td>Lifestyle</td>
<td>Typical way of life or manner of living characteristic of an individual or group.</td>
<td>Drinking five cups of coffee a day or regular exercise such as cycling could reduce the risk of developing breast cancer by at least 20%.</td>
</tr>
<tr>
<td>Vaccinations</td>
<td>Administration of vaccines to stimulate individuals’ immune responses.</td>
<td>The HPV vaccine can reduce the risk of cervical cancer by 100%.</td>
<td></td>
</tr>
<tr>
<td>---------------</td>
<td>-----------------------------------------------------------------------</td>
<td>---------------------------------------------------------------</td>
<td></td>
</tr>
<tr>
<td>Diagnosis</td>
<td>Clinical Manifestations: A symptom is observed by the patient subjectively but cannot be measured directly, whereas a <strong>sign</strong> is objectively observable by others.</td>
<td>Any abnormality of the breast is an early symptom of breast cancer.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Diagnostic Techniques and Procedures: Methods, procedures and tests performed to diagnose a disease, disordered function, or disability.</td>
<td>A compound derived from urinary thiol is the only reagent that can detect early cervical cancer.</td>
<td></td>
</tr>
<tr>
<td>Treatment</td>
<td>Surgery: Operations carried out for the correction of deformities and defects, repair of injuries, and diagnosis and cure of certain diseases.</td>
<td>Precancerous lesions in the endometrium indicate the need for surgery to remove the uterus.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Radiotherapy: Ionizing radiation carried out to treat malignant neoplasms and some benign conditions.</td>
<td>Up to 60% of cancer patients need radiotherapy in various stages of treatment.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Drug Therapy: Drugs and chemicals, including chemotherapy, targeted therapy, endocrine therapy, etc.</td>
<td>The new drug Pertuzumab (Perjeta) has been used together with Herceptin and chemotherapy to shrink tumors completely, so some patients do not need surgery.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Other Therapies: Other therapies, including traditional Chinese medicine, biotherapy, interventional therapy, etc.</td>
<td>With the application of traditional Chinese medicine, most patients with breast cancer will not need surgery.</td>
<td></td>
</tr>
</tbody>
</table>

**Diffusion Networks**

Out of all of the 524 tweets categorized as misinformation, only 64 received at least one retweet ($M = 5.09$, $SD = 70.40$), 72 received comment(s) ($M = 0.53$, $SD = 2.62$), and 132 received like(s) ($M = 2.68$, $SD = 31.80$). The popularity of the tweets was unevenly distributed, with several tweets receiving a large number of retweets, while the majority...
received no retweets or likes. For instance, the two most popular false messages were about gynecologic cancer prevention and treatment methods that involved eating specific foods, such as garlic, mushrooms, and red wine; these received 1,143 and 1,131 retweets, respectively. The senders of this popular misinformation all had verified accounts on Weibo and large numbers of followers, including grassroots users, medical professionals, and news media.

Out of the 1,023 tweets categorized as true information, only 143 received retweet(s) (M = 4.06, SD = 32.58). A total of 120 tweets received comment(s) (M = 1.78, SD = 14.41), and 167 tweets received like(s) (M = 2.54, SD = 28.08). Similarly, most of the tweets vanished into obscurity after being published, and only several tweets reached a high degree of popularity. Moreover, most of these popular tweets were about cancer prevention methods such as lifestyle choices and vaccinations.

By comparing the diffusion structures of the true information and misinformation, we found that true information was generally well diffused and accepted by social media users. Figure 1 displays the complementary cumulative distribution functions of the true information and misinformation cascades. It revealed that true information had better diffusion performance than the misinformation. While several false messages had been extreme popular and received a large number of retweets, most of the false messages received fewer retweets than the true messages. Additionally, true messages had better diffusion performance in terms of commenting, diffusion ranges and structural virality. All of the indices showed that true information spread more deeply and broadly than misinformation, reaching a larger audience on social media.
Figure 1. Complementary cumulative distribution functions (CCDFs) of true information and misinformation cascades.

**Group Comparisons between True Information and Misinformation**

Multivariate analyses of variance (MANOVA) were conducted to test the significant differences in the diffusion networks between true information and misinformation for the sub-categories of background knowledge, treatment, and prevention. The diagnosis category was excluded since there were not enough messages in the misinformation group to include it (N = 14). In the MANOVA, the independent variable was the veracity of information (true information versus misinformation), while the dependent variables were diffusion indices, including the number of retweets, number of comments, number of likes, range, and structural virality. Table 4 displays the means of all of the comparison groups. As seen in the table, true information for the categories of background knowledge and treatment had higher
diffusion indices compared to misinformation. However, for the category of prevention, misinformation diffused better than true information. The results of the MANOVA analyses are listed below.

Specifically, in terms of the category of background knowledge, the MANOVA results showed that there was at least one significant difference between the diffusion indices of the true information and misinformation groups, as Pillai’s Trace = .035, $F(2, 747) = 5.39, p < .001$. A post-hoc analysis showed that there were significant differences between the two groups in terms of the number of retweets, number of comments, range, and structural virality. The difference in the number of likes was insignificant. Similarly, in the category of cancer treatment-related information, the MANOVA results showed that there was at least one significant difference between the diffusion indices of the true and false information groups, as indicated by Pillai’s Trace = .049, $F(2, 481) = 4.91, p < .001$. Significant differences between true information and misinformation were also found in terms of the number of retweets, number of comments, range, and structural virality. The difference in the number of likes was not significant.

The MANOVA for the category of prevention-related information was also significant (Pillai’s Trace = .032, $F[2, 362] = 1.97, p = .07$). Significant differences in the number of retweets, range, and structural virality were found between the true information and misinformation groups. No significant differences were found for the number of comments and number of likes. This indicates that prevention-related misinformation spreads better than true information on social media.
Figure 2. Diffusion networks: (A) the full retweet network for true information (red) and misinformation (green), (B) the largest retweet network for truth, (C) the largest retweet network for misinformation.
Table 4.

*Estimated Marginal Means and MANOVA Results*

<table>
<thead>
<tr>
<th></th>
<th>Retweets</th>
<th>Range</th>
<th>Structural Virality</th>
<th>Comments</th>
<th>Likes</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean (s.e.)</td>
<td>F-value</td>
<td>Mean (s.e.)</td>
<td>F-value</td>
<td>Mean (s.e.)</td>
</tr>
<tr>
<td><strong>Background Information</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Truth (n = 462)</td>
<td>3.80 (0.92)</td>
<td>5.30*</td>
<td>0.21 (0.02)</td>
<td>20.18**</td>
<td>0.25 (0.02)</td>
</tr>
<tr>
<td>Falsehood (n = 287)</td>
<td>0.37 (1.17)</td>
<td>0.05 (0.03)</td>
<td>0.06 (0.03)</td>
<td>0.35 (0.48)</td>
<td>1.38 (0.62)</td>
</tr>
<tr>
<td><strong>Treatment</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Truth (n = 33)</td>
<td>2.00 (0.47)</td>
<td>10.87**</td>
<td>0.27 (0.07)</td>
<td>4.43*</td>
<td>0.34 (0.08)</td>
</tr>
<tr>
<td>Falsehood (n = 156)</td>
<td>0.29 (0.22)</td>
<td>0.12 (0.03)</td>
<td>0.12 (0.04)</td>
<td>0.35 (0.22)</td>
<td>0.50 (0.19)</td>
</tr>
<tr>
<td><strong>Prevention</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Truth (n = 400)</td>
<td>5.62 (4.21)</td>
<td>7.78**</td>
<td>0.14 (0.03)</td>
<td>6.26*</td>
<td>0.15 (0.03)</td>
</tr>
<tr>
<td>Falsehood (n = 67)</td>
<td>36.64 (10.29)</td>
<td>0.31 (0.07)</td>
<td>0.31 (0.07)</td>
<td>1.21 (2.21)</td>
<td>12.90 (6.31)</td>
</tr>
</tbody>
</table>

Notes: Background: Pillai’s Trace = .035, $F(2,747) = 5.39$, $p < .001$; Treatment: Pillai’s Trace = .080, $F(2,187) = 3.16$, $p < .01$; Prevention: Pillai’s Trace = .034, $F(2,465) = 3.25$, $p < .01$; *$p < .05$. **$p < .01$. ***$p < .001$. 
Discussion

Individuals have increasingly used social media to exchange cancer-related information. However, such information may include many false elements, and these could distort individuals’ attitudes and behaviors toward cancer prevention and treatment. This is the first study to examine the nature and diffusion of cancer-related misinformation on the Internet.

First, the findings revealed that of the 2,691 total tweets examined, more than half included medically oriented information about cancer. Moreover, we found that while the most of the medically oriented tweets provided accurate information, more than 30% contained misinformation. The findings suggest that patients should confirm the veracity of tweets on social media before accepting and following their advice. Moreover, social media service providers and medical professionals should evaluate online resources regularly to minimize cancer-related misinformation on social media.

Additionally, the results indicated that social media tweets related to cancer treatment contained a substantially greater percentage of misinformation than true information. However, the network analysis of the information cascade showed that for cancer treatment, background knowledge, and diagnosis, true information spread to a wider range of audiences than misinformation. This finding is inconsistent with that of Vosoughi et al.’s study of the diffusion of fake news. This could be due to the unique context of gynecologic cancers. Unlike Vosoughi et al. (2018), who examined all news stories on Twitter, we focused only on misinformation related to gynecologic cancers. Due to wide media coverage of female celebrities’ gynecologic cancers, many Chinese people are well aware of the topic, thus limiting the spread of misinformation related to it on social media. This indicates that the
spread of misinformation online could be specific to certain topics. While there was a relatively small amount of prevention-related misinformation about gynecologic cancers on social media, this misinformation diffused significantly more broadly and deeply than true information. This could be because prevention-related misinformation usually discussed cancer-preventing superfoods, exaggerating their efficacy, which is appealing to the general public due to its novelty and the general public’s uncertainty towards cancer. As a result, it spread more broadly and deeply than true information. Furthermore, a large amount of prevention-related misinformation provided ways or actions to prevent breast cancer and cervical cancer that individuals could perform by themselves. In other words, the prevention-related misinformation contained self-efficacy and response efficacy, which could help individuals reduce anxiety and fear as well as reduce their perceived threat from cancer. Thus, individuals are more willing to retweet these prevention-related messages. These findings suggest that medical professionals should make efforts to correct misinformation regarding methods of preventing gynecologic cancers on social media and increase individuals’ abilities to assess the veracity of online information in order to minimize the consequences of cancer misinformation.

**Conclusions**

This study is the first to examine the nature and diffusion of cancer-related misinformation on Chinese social media. First, a total of 2,691 gynecologic cancer-related messages were content analyzed to differentiate between true and false information on social media as well as to identify the types of falsehoods present. Additionally, a social network analysis was used to examine the diffusion cascades of misinformation through comparisons
with true information. The results indicated that while most of the gynecologic cancer-related messages provided medically accurate information, approximately 30% contained misinformation. More importantly, although cancer treatment tweets included a greater percentage of misinformation than true information, the prevention-related misinformation diffused significantly more broadly and deeply than true information. Future research should investigate the mechanisms behind the broad and deep diffusion of prevention-related messages on social media and elucidate effective strategies or campaigns that could limit the spread of this misinformation.
References


10.1371/journal.pone.0181640.


