Quantitative Study of Reading Content in Social Reading Environments: A Case Study of Articles from Reading Promotion WeChat Official Accounts (Postprint)

Authors: Wang Lei, Jiang Zailing, Liu Yuting

Date: 2023-07-26T00:00:00+00:00

Abstract

[Purpose/Significance] With the continuous evolution of social media big data platforms, big data mining, analytical technologies, and database technologies, various types of data in social media have become accessible and utilizable. Accordingly, this study conducts a bibliometric research on social reading content based on social media data.

[Method/Process] First, big data platforms and text mining technologies were employed to collect basic information of reading promotion WeChat Official Accounts and the content of articles pushed by these accounts. Second, statistical analysis was conducted on the collected data to obtain the distribution of WeChat Official Accounts across categories, regions, and verification statuses, as well as metrics including the number of pushed articles, read counts, average read counts, like counts, and average like counts for each account. By applying relevant theories and techniques of informetrics and utilizing self-developed scripts, the Single Tweet Communication Index (STCI), co-occurrence matrix of high-frequency keywords in titles, and coupling matrix of WeChat Official Accounts based on article repetition frequency were derived. Additionally, an algorithm for the Publish Delay Index of Highest STCI WeChat Article (HSPDI) was proposed to investigate the influence of article republication frequency on communication effectiveness. Finally, manual examination of high-STCI tweets was performed to summarize their content characteristics.

[Results/Conclusion] There exists a certain positive correlation between the average read counts and average like counts of articles pushed by reading promotion WeChat Official Accounts; a severe phenomenon of pushing identical articles exists among many WeChat Official Accounts; whether an article is first pushed

by a WeChat Official Account has no significant impact on its communication effectiveness. Furthermore, content analysis of high-STCI tweets reveals that the characteristics of high-STCI tweets mainly include: satisfying readers' curiosity, fulfilling readers' needs for self-expression, and adhering to writing principles that align with the characteristics of the 'screen reading' era.

Full Text

Preamble

Authors: Wang Lei[1], Jiang Ziling[2], Liu Yuting[3]
[1]Hefei University of Technology Library, Hefei 230009
[2]School of Instrument Science and Opto-electronic Engineering, Hefei University of Technology, Hefei 230009
[3]School of Textile and Clothing Engineering, Soochow University, Suzhou 215021

Abstract

[**Purpose/Significance**] With the development of social media big data platforms and the continuous evolution of big data mining, analysis technologies, and database technologies, various types of data in social media have become accessible and utilizable. Accordingly, this study conducts an informetrics analysis of social reading content based on social media data. [**Method/Process**] First, we collected basic information about reading promotion WeChat public accounts and the content of their published articles using a big data platform and text mining technology. Second, statistical analysis of the collected data revealed the distribution of WeChat public accounts across categories, regions, and certifications, as well as the number of published articles, reading counts, average reading counts, like counts, and average like counts for each account. By applying theories and techniques from informetrics and using self-developed scripts, we obtained the Single Tweets Communication Index (STCI), a co-occurrence matrix of high-frequency keywords in titles, and a coupling matrix of WeChat public accounts based on article repetition frequency. Simultaneously, we proposed the Publish Delay Index of Highest STCI WeChat Article (HSPDI) algorithm to examine the impact of repeated article publication on communication effectiveness. Finally, we manually read and analyzed high-STCI articles to summarize their content characteristics. [**Result/Conclusion**] There exists a certain positive correlation between average reading counts and average like counts for articles published by reading promotion WeChat public accounts. Many public accounts exhibit severe duplication in article publication. Whether an article is published for the first time by a public account has no significant impact on its dissemination. Further analysis of high-STCI articles reveals that their main characteristics include: satisfying readers' curiosity, meeting readers' needs for self-expression, and complying with writing principles suited to the "screen reading" era.

Introduction

Throughout history, reading has been considered an important means of promoting social civilization [1]. Reading promotion plays a significant role in cultural inheritance, intellectual development, information source provision, innovation stimulation, and human spiritual enhancement [2]. Reviewing the history of human reading, reading activities were linear in the era of paper-only media [3]. During this period, although interaction and sharing existed among readers, such behaviors were often limited to small physical spaces, such as traditional reading clubs [4]. With the rise of the third technological revolution, Internet technology was vigorously developed, the relationship between humans and technology became inseparable, reading content carriers expanded from paper documents to electronic media, and reading extended from paper-based reading to digital reading [5]. Subsequently, the arrival of the Web 2.0 era broke the traditional one-way knowledge information transmission and mass communication characteristics, enabling two-way and even multi-way knowledge interactive communication and segmented communication [6]. Internet users have entered a more proactive era where they are not only information recipients but also hold the initiative in information production and dissemination, playing the roles of information creators and disseminators [7]. Against this backdrop, social media reading based on sharing and interaction has become a favored social reading method for the public, with representative platforms including Twitter and Facebook in the United States, and WeChat and Weibo in China [8]. Exploring readers' social transformation of reading behaviors and habits on social media to provide more convenient and humanized social reading services has become a hot issue of common concern in both industry and academia. Scholars at home and abroad have conducted extensive research on social reading from different perspectives, mainly including analysis of user reading behavior in social reading [9-10], reading information dissemination and utilization [11-14], the impact of social reading on adolescent reading [15-18], and library reading promotion strategies based on social reading [19-22].

However, this paper finds that current research lacks examination of social reading from the text content dimension. Therefore, this study explores the feasibility of social reading content informetrics based on social media data. In selecting social media data, this paper chooses reading promotion WeChat public accounts for the following reasons: (1) Compared with other professional social reading apps such as Zite, Flipboard, NetEase Cloud Reading, and Zaker, WeChat reading has a larger user base. For instance, the "2017 WeChat Data Report" indicated that in September of the same year, an average of 902 million people logged into WeChat daily. (2) Articles published by reading promotion WeChat public accounts are typically long texts of considerable length, making reading such WeChat articles a true "reading behavior." (3) When users read articles through WeChat public accounts (especially articles shared by friends), social behavior itself creates a reading entry point. Although subscribing to public accounts is a personal behavior, users frequently engage in sharing or

interactive activities during the reading process, which aligns well with the core characteristics of social reading [23]. Meanwhile, through investigation, this paper selects the Qingbo Index Big Data Platform, a research and formulation body for China's new media big data evaluation system and influence standards, as the data source [24].

2 Research Methods

The data source collection and analysis process is shown in Figure 1 [Figure 1: see original paper]. According to Figure 1, the design process can be divided into two parts: data collection and data analysis. First, we used the Qingbo Big Data Platform to export the basic data of the studied WeChat public accounts (WeChat ID, account name, category, certification information, region, function introduction, tags) and external feature data of WeChat articles (article title, reading count, like count, URL). Through the URL links in the external features, we used text mining technology to collect internal feature data of each public account' s articles (article content, word count, number of dynamic images, static images, total images, audio files, video files, title word count), and stored this data in a MongoDB database. Subsequently, using self-developed Python scripts, we obtained the Single Tweets Communication Index (STCI), a co-occurrence matrix of high-frequency keywords in titles, and a coupling matrix of WeChat public accounts based on article repetition frequency. Simultaneously, we proposed the Publish Delay Index of Highest STCI WeChat Article (HSPDI) algorithm to examine the impact of publication timing on communication effectiveness. Finally, we manually read and summarized the content characteristics of high-STCI articles. Although the volume of data collected and analyzed in this study is not particularly large (56,355 structured data entries), the program scripts written and the scalable MongoDB database can be fully applied to big data analysis.

2.1 Data Source Collection

First, we searched for "tag: reading promotion"in the Qingbo Big Data Platform, obtaining 395 relevant results. Based on the "Pareto Principle" (the vital few determine the outcome), we selected the top 20% by WCI index (79 accounts) for study and added them to a custom list. Table 1 shows the statistical results of WeChat IDs, account names, and WCI indices for these 79 reading promotion accounts.

Subsequently, based on the basic information of the 79 reading promotion WeChat public accounts, we used the custom list export function of the Qingbo Big Data Platform to export external features (article title, reading count, like count, URL) for 56,355 WeChat articles published between May 1, 2017, and October 31, 2017. Using text mining technology, we traversed all URL links of these 56,355 articles to obtain corresponding internal feature data, which was integrated and stored in a MongoDB database. Table 2 shows partial data stored in the MongoDB database.

2.2 Data Analysis

2.2.1 Basic Statistics of WeChat Public Accounts and Coupling Analysis Based on Article Repetition By running self-developed program scripts, we obtained the number of published articles, reading counts, average reading counts, like counts, and average like counts for each public account during the collection period, as well as statistical results for classification, certification information, and geographical distribution.

In the public account coupling analysis, we defined the coupling relationship strength between accounts based on the repetition frequency of articles. First, we used self-developed scripts to obtain basic information about repeatedly published articles (5,983 articles) during the collection period. Then, for each public account, we counted its article repetition frequency with every other account, using this repetition count as the coupling strength between the two accounts. Through self-developed algorithms, we obtained the coupling matrix and visualized the matrix data using NetDraw software to generate a coupling network diagram of public accounts based on repetition frequency.

2.2.2 WeChat Communication Index (WCI) and Single Tweets Communication Index (STCI) Algorithms The WCI (WeChat Communication Index V13.0) [25] in the Qingbo Index Big Data Platform has richer indicators than previous versions, providing more comprehensive evaluation dimensions and optimizing the weight of each indicator to make the ranking more accurate. Its calculation is shown in Formula (1):

$$WCI = \{30\%\cdot[0.85\cdot\ln(R/d+1)+0.15\cdot\ln(10\cdot Z/d+1)]+30\%\cdot[0.85\cdot\ln(R/n+1)+0.15\cdot\ln(10\cdot Z/n+1)]+30\%\cdot[0.85\cdot\ln($$

Where: - n = number of articles published by the public account during the data collection period - R = total reading count of all articles (n) during the collection period - Z = total like count of all articles (n) during the collection period - d = number of days in the collection period - R_t and Z_t = total reading and like counts for top articles during the collection period - R_{max} and Z_{max} = maximum reading and like counts for articles published by the account during the collection period

However, this index primarily evaluates the overall popularity and development trend of a WeChat account. To identify high-communication-index individual articles, this paper utilizes our previous research result, the Single Tweets Communication Index (STCI) [26], to evaluate single article communication effectiveness. Its calculation is shown in Formula (2):

$$STCI = \{60\%\cdot(20\%\cdot\ln(R/d+1)+80\%\cdot\ln(R+1))+40\%\cdot(20\%\cdot\ln(10\cdot Z/d+1)+80\%\cdot\ln(10\cdot Z+1))\}^2\cdot 10 \quad \text{(Formula}$$

Where: - R = reading count of the article from publication to data collection cutoff date - Z = like count of the article from publication to data collection cutoff date - d = number of days from article publication to data collection cutoff date

Data collection for this paper was conducted on November 7, 2017, covering the period from May 1, 2017, to October 31, 2017 (if an article was published on May 1, 2017, then $d = 190$ days).

2.2.3 Highest STCI Article Delay Index (HSPDI) Algorithm To determine the relationship between publication timing and STCI index for repeatedly published articles, we propose the Publish Delay Index of Highest STCI WeChat Article (HSPDI). The specific calculation steps are: (1) Extract all publication dates $\{date_2, date_5, date_1, date_4, date_3 \dots date_n \dots date_i\}$ and corresponding STCI indices $\{STCI_2, STCI_5, STCI_1, STCI_4, STCI_3 \dots STCI_n \dots STCI_i\}$ for the same article from external feature data; (2) Sort each date set and corresponding STCI index set in chronological order to obtain the sorted publication date set $\{date_1, date_2, date_3, date_4, date_5 \dots date_i \dots date_n\}$ and corresponding STCI index set $\{STCI_1, STCI_2, STCI_3, STCI_4, STCI_5 \dots STCI_i \dots STCI_n\}$; (3) Identify the highest STCI index $STCI_i$ for the same article and record its ordinal position i; (4) Divide the ordinal position i by the total number of elements n in the set to obtain the article' s HSPDI. The calculation model is as follows:

$$
\begin{aligned}
DATE &= \{date_2, date_5, date_1, date_4, date_3 \dots date_n \dots date_i\} \\
STCI &= \{STCI_2, STCI_5, STCI_1, STCI_4, STCI_3 \dots STCI_n \dots STCI_i\} \\
DATE' &= \{date_1, date_2, date_3, date_4, date_5 \dots date_i \dots date_n\} \\
STCI' &= \{STCI_1, STCI_2, STCI_3, STCI_4, STCI_5 \dots STCI_i \dots STCI_n\} \\
Max(STCI') &= STCI_i \rightarrow \text{take } i \rightarrow HSPDI = i/n
\end{aligned}
$$

Where $DATE$ is the set of different publication dates for the same article, $STCI$ is the set of corresponding article STCI indices, $DATE'$ is the sorted publication date set, $STCI'$ is the sorted STCI index set, and n is the number of elements in the set.

Specific calculation example: The article "To All Who Don' t Read: You Will 'Die' at 35" was repeated $n = 4$ times, with publication dates $DATE = [2017-09-02, 2017-09-28, 2017-08-23, 2017-09-03]$ and corresponding STCI indices $STCI = [658.54, 425.8, 416.09, 586.47]$. After sorting by date, we get $DATE' = [2017-08-23, 2017-09-02, 2017-09-03, 2017-09-28]$ and $STCI' = [586.47, 658.54, 416.09, 425.8]$. The highest STCI index is $STCI_2 = 658.54$ with ordinal position $i = 2$, so the HSPDI is $i/n = 2/4 = 0.5$.

2.2.4 WeChat Article Content Analysis and Co-occurrence Analysis Based on Title High-Frequency Words In the co-occurrence analysis based on title word segmentation, we extracted article titles from each structured data entry in the MongoDB database and obtained meaningful nouns (n), verbs (v), and gerunds (vn) from the titles using the "jieba" Chinese word segmentation module. Through self-developed scripts, we counted high-frequency words in titles and generated a high-frequency word co-occurrence matrix. Finally, we imported the data into Ucinet software and used NetDraw visualization software to draw a co-occurrence network map of high-frequency words in article titles. To further analyze the content of WeChat public account articles, we manually read high-STCI articles (top 30 by STCI ranking) and low-STCI articles (bottom 30 by STCI ranking) to summarize the content characteristics of articles published by WeChat public accounts.

3 Results and Discussion

3.1 Overall Analysis of Reading Promotion Public Account Articles

We statistically analyzed 56,355 structured data entries collected between May 1, 2017, and October 31, 2017, obtaining the number of published articles, reading counts, average reading counts, like counts, and average like counts for 79 reading promotion public accounts. Figure 2 [Figure 2: see original paper] shows the visualized statistical chart.

Where: - Article count = total number of articles published by each account during the collection period - Reading count = sum of reading counts for all articles during the collection period - Average reading count = reading count divided by article count - Like count = total number of likes received during the collection period - Average like count = like count divided by article count

The top five accounts by article count are: Meiwen Yuedu (1,472 articles), Shuimu Wenzhai (1,471 articles), Baicaoyuan Shudian (1,470 articles), Cihuai Dushuhui (1,470 articles), and Dushu 369 (1,467 articles). The top five by reading count are: Shidian Dushu (146 million), You Shu (144 million), Baicaoyuan Shudian (113 million), Yixingqi Yiben Shu (109 million), and Yuedu (101 million). The top five by average reading count are: Shidian Dushu (100,000+), You Shu (99,900), Yixingqi Yiben Shu (99,100), Yidu (88,200), and Shudan Lailai (85,600). The top five by like count are: Shidian Dushu (9.4978 million), You Shu (5.7629 million), Baicaoyuan Shudian (2.0357 million), Guoxue Wenhua (1.9619 million), and Yixingqi Yiben Shu (1.7552 million). The top five by average like count are: Shidian Dushu (6,487), You Shu (4,007), Yixingqi Yiben Shu (1,589), Shudan Lailai (1,480), and Guoxue Wenhua (1,470).

Figure 2 shows no obvious correlation between article count and reading count, average reading count, like count, or average like count. However, there is a strong positive correlation between average reading count and average like count. In other words, high article count does not necessarily mean high average reading or like counts, but high average reading count often corresponds to high

average like count. This suggests that whether readers click to read after seeing an article title is a subjective decision, but whether they like it after reading depends on personal habits, demonstrating the importance of article titles for dissemination.

3.2 Statistics by Category, Certification, and Geographical Distribution

We extracted category, certification information, and geographical distribution data from the basic information of the 79 public accounts, as shown in Table 3 (partial data).

Table 3 shows that most high-WCI-index accounts have obtained certification, leading us to conclude that whether a WeChat public account is certified is related to its social reading promotion effectiveness—that is, certification status affects a public account' s social reading promotion performance.

We also analyzed the geographical origins of the 79 reading promotion accounts, finding that Beijing has the most with 15 accounts, followed by Guangdong (9), Jiangsu (6), Shanghai (4), Shandong (3), and Henan (3). Geographically, WeChat public accounts are mainly concentrated in economically developed coastal and central China regions. Combined with Figure 2, we can see that among the top 5 by reading count (Shidian Dushu, You Shu, Baicaoyuan Shudian, Yixingqi Yiben Shu, Yuedu) and top 5 by like count (Shidian Dushu, You Shu, Baicaoyuan Shudian, Guoxue Wenhua, Yixingqi Yiben Shu), most are certified accounts, indicating that certified accounts have greater credibility and readers prefer them.

3.3 Coupling Analysis Based on Repeated Articles

Through self-developed algorithms, we obtained a coupling matrix of 79 WeChat public accounts based on repeated article frequency, with partial data shown in Table 4 . This is a fully connected network with a density of 3.41. Due to the dense connections, it is difficult to describe network details after visualization. To more clearly represent the relationships in the coupling network, we extracted data with coupling strength values greater than or equal to 30, removed isolated nodes, and obtained a deeply aggregated public account relationship network, shown in Figure 3 [Figure 3: see original paper].

In Figure 3, circles represent public accounts, with size indicating the number of repeated articles published (larger circles = more repetitions). The thickness of lines between circles represents coupling strength between two accounts.

Although reading promotion accounts do not jointly publish articles, using repetition frequency to define coupling relationships reveals potential connections. In other words, these accounts may not have organized joint publications but do pay attention to the same content within their own article systems. Figure 3 shows that the top five coupling strengths are between Dushu and Yuedu (429),

Yixingqi Yiben Shu and Shuiqian Bandu (396), Yuedu and Shuimu Wenzhai (374), Guoxue Wenhua and Duwenzhai Jingxuan (359), and Duwenzhai Jingxuan and Shidian Wenzhai (341). Combining with Table 1, Dushu's WCI index is 908.03 while Yuedu' s is 1217.2—a large gap—showing that accounts with strong coupling may have significantly different WCI indices. However, Yixingqi Yiben Shu' s WCI is 1359.9 and Shuiqian Bandu' s is 1285.22; Guoxue Wenhua' s WCI is 1278.73 and Duwenzhai Jingxuan' s is 1286.22—showing that strongly coupled accounts can also have similar WCI indices. Therefore, we conclude that mutual article borrowing is common among reading promotion WeChat public accounts, and accounts with many identical articles (stronger relationships) do not necessarily have the same communication capabilities.

3.4 Highest STCI Article Delay Index (HSPDI) Analysis

Using self-developed scripts, we extracted external feature data for all repeatedly published articles (5,983 articles) from each public account. By running the HSPDI algorithm script, we obtained the HSPDI for these 5,983 articles.

Based on the common notion of news exclusivity and timeliness, we hypothesized that earlier-published articles would have higher STCI indices—that is, the HSPDI should be the reciprocal of repetition frequency. To test this hypothesis, we ran self-developed scripts to obtain the HSPDI for each article (5,983 articles), used repetition frequency as the x-axis and HSPDI as the y-axis, and visualized the distribution using the Matplotlib library, resulting in Figure 4 [Figure 4: see original paper]. Each crosshair corresponds to one article, with brightness representing the number of articles with the same repetition frequency and HSPDI index (brighter = more articles). Gray dots represent the minimum HSPDI index for articles with the same repetition frequency (i.e., the reciprocal of repetition frequency).

Clearly, the crosshairs and brightness levels are randomly distributed, and the HSPDI index is not the minimum value we hypothesized. Additionally, when analyzing the relationship between repetition frequency and STCI index, we found that the top three most repeated articles— "Spending Money Best Reveals a Person' s Basic Quality and Mentality" (repeated 35 times), "The Best Living State of a Woman" (30 times), and "Whether a Woman Lives Well Can Be Seen from One Point" (23 times)—have highest STCI indices of 777.38, 1164.91, and 758.98, respectively.

From this, we infer that article publication timing is unrelated to its STCI index, indicating that users do not care whether articles are original when reading WeChat articles. Meanwhile, comparing the STCI indices of the most repeated articles with those of high-STCI articles shows that repetition frequency is unrelated to article communication capability.

3.5 High-STCI Article Content Analysis

We first removed punctuation from 56,355 WeChat article titles, then used the "jieba"Chinese word segmentation library to extract meaningful nouns (n), verbs (v), and gerunds (vn) from titles, obtaining 653,779 words total, with 35,036 unique words after deduplication. We selected words with frequency exceeding 200 and removed non-meaningful words to obtain 80 high-frequency words, shown in Table 5 . We imported this 80$\times$80 co-word matrix into Ucinet software and used NetDraw to visualize the high-frequency word co-occurrence network, shown in Figure 5 [Figure 5: see original paper].

This is a fully connected network with density 3.01. In Figure 5, circles represent high-frequency words, with size indicating word frequency in titles (larger circles = higher frequency). Line thickness between circles represents co-occurrence frequency between words. Through cluster analysis, we categorized the main content areas covered by WeChat public account articles: (a) love and relationships between men and women, including topics about excellent, characterful, emotionally intelligent, hardworking, and humorous men, and topics about beautiful, photogenic, happy, and mature women; (b) family issues of infidelity and divorce involving husbands or wives; (c) children' s education issues related to parents and teachers, including college entrance examinations and future development; (d) collection, reading, and recitation of classic literature and poetry; (e) reflections on life and living, relationships with friends, and principles of being a good person; (f) Chinese and world news and historical stories.

3.6 Detailed Content Analysis of High-STCI Articles

To further understand WeChat article content characteristics, we manually read high-STCI articles (top 30 by STCI ranking) and low-STCI articles (bottom 30 by STCI ranking). Table 6 shows statistics for the top 10 articles by STCI ranking, while Table 7 shows statistics for the bottom 10 articles.

Manual reading revealed that public accounts publishing more high-STCI articles are mainly Shidian Dushu (19 articles) and You Shu (11 articles), while those publishing more low-STCI articles are mainly Meiwen Gongxiang (18 articles) and Zhihui Guanli (7 articles). Combined with the cluster analysis results, these 30 high-STCI articles do not cover classic literature, poetry collection, reading, or recitation, indicating that society currently faces a "deep reading crisis." In the digital age, the Internet precisely releases various types of sensory and cognitive stimuli (repetitive, instantaneous, high-intensity, intelligent, interactive). Moreover, much information is easily accessible with minimal effort. Meanwhile, people are often highly stressed and multitasking when facing complex information data. Under the combined influence of external and internal factors, readers' reading speed tends to increase, content switches more rapidly, attention becomes more easily distracted, at the cost of reduced comprehension. Time has become a scarce resource, and even brief reflection may be considered a waste of time, causing "deep reading" to drift further away from people' s

lives.

Further analysis of 30 low-STCI articles revealed that most involve advertising and promotion, such as "Annual sales of 1 billion boxes, popular for 40 years, what does this 'national delicacy' from a European legend rely on?" and "G20 designated towels, changing a lifestyle, 4 for 99 yuan!" Such content consumes readers' trust in the public account, causing them to selectively ignore articles based on titles alone.

Further analysis of high-STCI articles revealed three main characteristics:

(1) **Satisfying readers' curiosity.**

(a) Using suspenseful and controversial titles. For example, the article "Celebrity romance crashes the Internet, yet this Chinese man collectively kneeled for by foreigners is ignored···" uses the strong contrast between the high popularity of celebrity romance and the protagonist' s obscurity to create suspense (who is this Chinese man with such great charm?), prompting readers to click and read. The article "Good women must learn to talk about money" creates controversy in the title (the role of money in love and marriage), generating discussion that makes readers want to know how the article addresses the topic.

(b) Creating dense, arrayed stimulus points. The article "First prize work: 'My Mother is a Madwoman,' makes countless people cry···" opens with suspense (why is the mother mad), then connects the madwoman' s tragic experiences and conflicts with "me," finally resolving the conflict when the "madwoman" accidentally dies falling off a cliff. The plot is tightly linked with reversals, creating a stimulus point almost every two screens, making it irresistible.

(2) **Meeting readers' self-expression needs.**

(a) Using titles that resonate with people. The article "Young people who trade their lives for money, trade until they die" strikes directly at the heart, causing strong resonance and reading interest. "Never underestimate a 30-year-old woman" directly voices the feelings of 30-year-old women—though no longer young girls, they have become warriors who can face life' s challenges through age and experience.

(b) Starting from familiar life scenarios to solve problems for readers and help them express inner thoughts. The article "Why are Chinese mothers so resentful? Mother-in-law overstepping, husband absent, self misplaced" begins with classic family conflicts, analyzes the reasons as mother-in-law overstepping, husband absent, and self misplaced, gains high reader recognition, and proposes practical solutions—everyone staying in their proper place—meeting reader needs. The article "Never 纠缠 with low-level peo-

ple" selects familiar life scenarios: everyone has encountered unreasonable people, and reasoning with them often brings trouble. The familiar situation immerses readers while providing solutions—the best approach is to ignore them.

(3) **Complying with "screen reading" era writing principles.**

(a) Simple, colloquial language—use short sentences instead of long ones. The article "A certain island country claims a coral island built by China in the South China Sea over three years mysteriously disappeared, saying it was eaten by Chinese starfish; China: don' t talk nonsense" uses many short sentences, creating rhythm and musicality that reduces readers' comprehension costs.

(b) Creating imagery with words. The article "First prize work: 'My Mother is a Madwoman,' makes countless people cry···" uses numerous detailed descriptions to vividly portray the "madwoman" —a mentally disturbed mother who deeply loves her child—shocking readers' hearts.

(c) Using "golden sentences" to highlight key points. The article "Good women must learn to talk about money" intersperses stories with incisive statements like "Only when you have enough money can you have the confidence and qualification to pursue what you truly think is important," creating strong impact.

(d) Using concrete instead of abstract. The article "A letter to a cheating husband" describes the protagonist discovering her husband' s infidelity: "These lines of text, each as big as a meteorite, crashed into me without warning and with thunderous force, smashing my internal organs to pieces." Comparing text messages to meteorites crashing into the heart instantly brings to life an image of a heartbroken woman.

(e) Restrained expression. The article "She lost her husband in middle age, was imprisoned at 51, experienced half a lifetime of hardships, yet at 90 became China' s most exquisite woman!" tells of protagonist Zheng Nian' s tumultuous life, each period worthy of extensive description, but the author uses precise, appropriate language, ending the tumultuous experience concisely and decisively, leaving readers wanting more with fragmented descriptions.

(f) Comfortable layout. Reading promotion articles are typically lengthy, and many high-STCI articles use multi-paragraph, short-sentence layouts with bold fonts, underlining, and colored fonts to enhance readability.

4 Conclusions and Recommendations

4.1 Conclusions

This paper collected internal and external feature data of reading promotion WeChat public accounts and their articles using social media big data platforms and text mining technology, then conducted empirical research on the collected data using statistical and informetrics theories and techniques. The main conclusions are:

(1) No significant correlation exists between the number of articles published by WeChat public accounts and their reading counts, average reading counts, like counts, or average like counts. However, a strong positive correlation exists between average reading count and average like count—accounts with higher average reading counts tend to have higher average like counts.

(2) Statistics based on category, certification, and geographical distribution show that WeChat public accounts in economically developed coastal and central China regions have relatively higher WCI indices, indicating that regional economic development level affects residents' reading methods—people in economically developed areas are more accustomed to reading on the WeChat platform. We also found that certified accounts have greater credibility, and users prefer to read articles from certified WeChat public accounts.

(3) Coupling analysis based on repeated articles reveals that mutual article borrowing is common among reading promotion WeChat public accounts, yet accounts with many identical articles (stronger relationships) do not necessarily have the same communication capabilities.

(4) Analysis of the Highest STCI Article Delay Index shows that article timeliness and repetition are unrelated to communication index—users do not care whether articles are original when reading WeChat articles.

(5) Content analysis of high-STCI articles reveals that articles with strong communication capabilities mainly involve lifestyle content such as male-female relationships, children' s education, and success, with less involvement in literary content like recitation and poetry, indicating that society faces a "deep reading crisis" in the new media environment.

(6) Study of low-STCI articles found most involve advertising and promotion. Analysis of high-STCI articles identified main characteristics: using suspenseful/controversial titles; creating dense, arrayed stimulus points; using resonant titles; starting from familiar life scenarios to solve problems; using simple, colloquial language; creating imagery with words; using "golden sentences" ; using concrete instead of abstract; restrained expression; and comfortable layout.

4.2 Recommendations

Based on these conclusions, we propose suggestions for improving reading promotion WeChat public account operations:

(1) Reasonably utilize WeChat backend data (subscription numbers, daily unique IP visits, etc.) combined with indicators from WeChat big data platforms (like counts, reading counts, WCI index) or custom indicators (such as our STCI or HSPDI) to identify user interest points for precise 推送.

(2) Increase publicity efforts for WeChat public accounts, penetrate less economically developed areas to broaden the audience, actively pursue certification, and optimize WeChat menu bars to improve user experience.

(3) Addressing the serious phenomenon of article borrowing among reading promotion accounts, we recommend strengthening protection of original article intellectual property to establish account brands. Instead of reposting, focus on creating high-communication-index original articles.

(4) After summarizing high-STCI article characteristics, we believe creating high-communication-index WeChat articles requires compliance with "screen reading" era writing principles while satisfying readers' curiosity or self-expression needs.

Acknowledgments

We thank Sun Jing and Yi Yongmei for their assistance with data description, proofreading, and Chinese abstract translation during paper writing.

References

[1] Zhao Wenjun, Chen Huanzhi, Jiang Weijin. Review of digital reading research at home and abroad in recent 10 years[J]. Library and Information Service, 2017, 61(18): 128-136.
[2] Zhang Xiaomei. On service innovation in reading promotion[J]. Library and Information, 2017, 61(1): 29-30.
[3] He Yun, He Lanman. On national reading promotion strategies from the dual relationship between traditional reading and digital reading—taking Japan as an example[J]. Library, 2015(7): 34-38.
[4] Antonio J, Alonso J, Gomez R, et al. Social reading: platforms, applications, clouds and tags[M]. Oxford: Chandos Publishing Ltd, 2013.
[5] Sun Yigang, Zhang Wei, Xie Qiang, et al. Digital reading service and promotion of National Digital Library[J]. Library Journal, 2016(10): 4-8.
[6] Zhang Xiaoqing, Zhang Zhihe, Xiang Chunyan. Research on knowledge dissemination based on Web 2.0[J]. Modern Communication (Journal of Communication University of China), 2010(4): 123-126.
[7] Jiang Zhifang. Review of research on information user needs and information

user behavior[J]. Library Tribune, 2006(4): 284-286, 278.
[8] Yang Jing. Research on library reading promotion based on social media[J]. Library Research, 2017(5): 57-60.
[9] Li W, Wu Y. Adolescents' social reading: motivation, behaviour, and their relationship[J]. Electronic Library, 2017, 35(2): 246-262.
[10] Zhang Yun, Mao Yihong. Investigation and analysis of college students' social reading behavior[J]. Journal of National Library of China, 2016, 25(2): 77-87.
[11] Liu Jing. Investigation and analysis of certified user microblogs of university libraries in China—taking Sina Weibo as the platform[J]. Library Science Research, 2012(1): 90-95.
[12] Kibugi S. The use of social media in the dissemination of information in selected public and private university libraries in Kenya[J]. Innovation: journal of appropriate librarianship and information work in Southern Africa, 2013(47): 100-120.
[13] Zinoviev D, Duong V, Zhang H. A game theoretical approach to modeling information dissemination in social networks[EB/OL]. [2018-10-10]. https://arxiv.org/pdf/1006.5493v1.pdf.
[14] Yang D, Chow T W S, Zhong L, et al. True and fake information spreading over the Facebook[J]. Physica a: statistical mechanics & its applications, 2018, 505: 984-994.
[15] Chance R, Lesesne T. Rethinking reading promotion[J]. Teacher Librarian, 2012, 39(5): 26-32.
[16] Dikev W, Eke H N, Babarinde E T. Social media and reading among secondary school students in Enugu State[J]. Mousaion, 2013(31): 60-81.
[17] Sampe E A B, Aforo A A. Impact of social media on academic reading: a study at Kwame Nkrumah University of Science and Technology, Kumasi, Ghana[J]. Asian Journal of Humanities and Social Studies, 2014, 2(1): 92-99.
[18] Li W, Wu Y, Liu Y. Study on relationship between adolescents' social reading motivation and their behaviors: a case study of middle school students' reading via WeChat in Shanghai[J]. Library & Information Service, 2014, 40(6): 115-128.
[19] Sun L, Xie J. An exploration of reading promotion and readers' advisory at Chinese public libraries[J]. Reference & User Services Quarterly, 2016, 56(2): 96.
[20] Yuan Shifen. Research on new media reading promotion strategies in university libraries[J]. Modern Information, 2013, 33(1): 74-77, 81.
[21] Zhang Qiu, Yang Ling, Wang Man. Development status and countermeasures of university library WeChat public platform services[J]. Library Construction, 2014(2): 61-65, 69.
[22] Wan Muchen, Ou Liang. Empirical research on reading promotion in university libraries based on WeChat public platform[J]. Library and Information Service, 2015, 59(22): 72-78.
[23] Li Wu, Wu Yuehua, Liu Yu. Study on relationship between adolescents' social reading motivation and behavior—taking Shanghai middle and high school students' WeChat reading as an example[J]. Library and Information

Service, 2014, 58(23): 61-68.
[24] Qingbo Big Data. Qingbo Big Data Platform[EB/OL]. [2017-12-19]. http://www.gsdata.cn/.
[25] Qingbo Index. WeChat Communication Index WCI (V13.0)[EB/OL]. [2017-12-19]. http://www.gsdata.cn/site/usage.
[26] Wang Lei, Lü Penghui, Zhang Renqiong. Informetrics analysis of university library WeChat information[J]. Library and Information Knowledge, 2017(5): 49-58, 88.

Author Contributions

Wang Lei: Responsible for topic and outline formulation, data processing and analysis, data summarization, and paper revision.
Jiang Ziling: Responsible for paper writing and revision, data processing and analysis.
Liu Yuting: Responsible for data processing and analysis, data summarization.

Note: Figure translations are in progress. See original paper for figures.