Scholarly Communication and Information Behavior in Chinese Social Networks: A Sentiment Analysis of WeChat Academic Communities

Jiaqi Feng

Recommended citation:

Feng, Jiaqi (2024). "Scholarly Communication and Information Behavior in Chinese Social Networks: A Sentiment Analysis of WeChat Academic Communities". *Profesional de la información*, v. 33, n. 3, e330314.

https://doi.org/10.3145/epi.2024.0314

Manuscript received on 12th January 2024
Accepted on 30th March 2024

Jiaqi Feng
https://orcid.org/0009-0000-9325-6619
School for Marxism Studies
Shanxi University
Taiyuan 030006, Peoples R China
fjq@163.com

Abstract

WeChat has penetrated into academic communication circles; however, there is still a significant gap in understanding the way sentiment and information behaviors shape scholarly discourse on this platform. The present research aimed to explore the scholarly communication and information behavior in the Chinese social networks. For this purpose, a sentiment analysis of WeChat academic communities was performed. This study adopted a comprehensive methodology that involved collection of 800 WeChat articles on the basis of engagement metrics, followed by the data pre-processing. It involved cleaning, Chinese words segmentation using Jieba library and TF-IDF vectorization for text analysis. Results of SVM model demonstrated robust performance in sentiment analysis with an overall accuracy of 89% and consistent precision and recall rates across the sentiment categories. Comparison with existing studies also highlighted effectiveness of this model in classifying sentiments on WeChat. Utilization of SVM in sentiment analysis advances the theoretical understanding of text classification techniques in social media environments. The findings also provide valuable insights for researchers and practitioners so that they can leverage SVM for effective classification of SVM. Limitations and future research indications have also been explained in the study.

Keywords

Scholarly Communication, Chinese Social Networks, Support Vector Machine (SVM), Sentiment Analysis, WeChat.

1. Introduction

The rapid evolution of digital communication has transformed the environment of scholarly exchange and information behaviors. The digital communication application, WeChat, is considered as a multi-functional messaging and social media application across the globe. The academic communities have benefited from WeChat and have become central to the dissemination and discussion of scholarly content. In China WeChat has emerged as a pivotal platform not only for social interaction but also for academic engagement. Reports of **Statista** (2024) highlight that WeChat has become the king of apps in China (Figure 1). Within a decade, this Chinese mega app has now become one of the strongest global brands which have almost 1.3 billion monthly active users. It equates to almost 80 percent of the total population in China. Therefore, with the growing number and internationalization of users, WeChat has become an important platform for scientific communication. This is because it offers the potential of mapping and evaluating the communication of science within the context of China (**Becker *et al.***, 2023).

In a Springer Nature survey conducted in 2019, it has been reported that 94% of the 528 respondents in China indicated that they are using WeChat in a professional context (**Cong *et al.***, 2022). Scholarly communication traditionally relied on formal channels such as academic journals, conferences, or institutional repositories. WeChat also functions in a

CC BY

manner similar to other applications of social media. Therefore, this tool has the potential to facilitate education across a wide range of disciplines (**Luan *et al.***, 2020). WeChat with its wide-ranging influence and diverse functionalities also provides a unique environment for the academic discourse which also enable the researchers to share knowledge in real-time discussions. However, with its large and diverse user-base, a vast amount of user-generated content and increasing global reach, WeChat provides unique opportunities for researchers (**Zhang; Quan-Haase**, 2022).

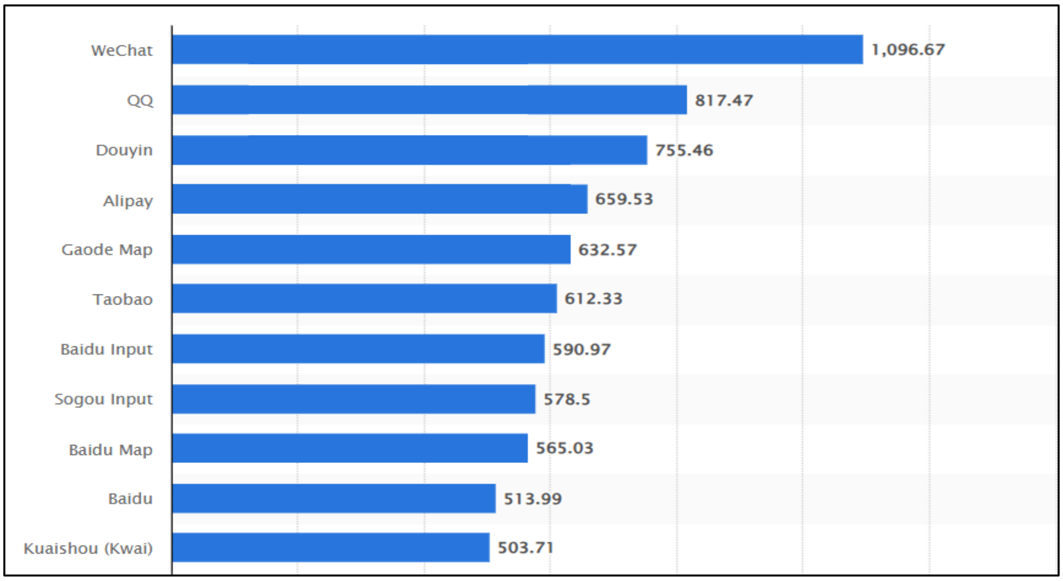

Figure 1: Usage of WeChat in China Source: Statista (2024).

There are ample studies that focus on the application of WeChat in distinct contexts (**Wang; Jiang**, 2024; **Zhang *et al.***, 2024a; **Chen *et al.***, 2020; **Zhou *et al.***, 2017), but to the best of researcher's knowledge, there is a scarcity of research that examines the scholarly communication and information behaviors in Chinese social network. Despite the growing usage of WeChat in China (**Cao *et al.***, 2024), there is a lack of comprehensive understanding regarding the sentiments of interaction and the nature of information exchange with these platforms. Furthermore, there is also a potential research gap regarding the sentiment analysis of WeChat academic communities.

To bridge this research gap, this study focuses on the sentiment analysis of WeChat academic communities to explore the way through which scholarly communication occurs within this platform. It also aims to understand the information behavior of its users; to understand the sentiments and information behavior in WeChat academic communities; and to grasp the way through which digital platforms shape the scholarly practices. Furthermore, the study also contributes towards the broader discourse regarding digital communication in academic. It offers practical implications for the researchers, educators and policymakers that seek to leverage social media for academic purposes.

2. Literature Review

2.1. Social Networks in China

Social networks in China differ from those in western countries because of the unique digital and regulatory environment of the country, as well as cultural factors. WeChat was developed by Tencent (**Li**, 2024) as the most dominant platform in China. It combines messaging, social networking, services of payment and more. It serves as a one stop platform for communication, news, shopping and even government services. **Plantin and De Seta** (2019) demonstrate that WeChat has arguably become the most popular application of mobile in China today. Outside China, WeChat is often considered in business reports and in technology journalism as a classic digital platform in China that is poised to replace Google or Facebook as the leading model of novel products. In this regard, their research highlights that WeChat combines the properties of both infrastructure and platform. **Wang and Jiang** (2024) also explain that WeChat has garnered significant global attention as a versatile technology. With features similar to those of popular platforms such as Skype, Twitter, and Facebook, it facilitates both synchronous and asynchronous interactions which provides a seamless experience of learning which is unrestricted by location or time. Based on its advantages and widespread usage, WeChat is increasingly being integrated into language and cultural learning. **Ru and Nindum** (2024) have demonstrated its effective role in Chinese family communications.

Weibo is another important social network in China. It has been observed that during COVID-19, majority of researchers directed their focus towards studying trends and practices through Weibo such as sentiment analysis or psychological consequences (**Li *et al.***, 2020b; **Wang *et al.***, 2020; **Li *et al.***, 2020a; **Lyu *et al.***, 2020). Study by **Zhang *et al.*** (2023) indicated that the number of Weibo users increased significantly from 63 million in 2010 to 350 million in 2018. In this regard, new organizations publish news on their websites, using both applications, WeChat and Weibo accounts. Their research also indicated that although Weibo accounts in the state media of China serves as a tool to capture the public voice, and a channel for the attainment of social news; but it has also served as a propaganda tool for the facilitation of spreading positive news and management of crisis.

2.2. Scholarly Communication through Social Networks

Hailu and Wu (2021) found that, in the past two decades, scholarly communication environment has altered with the increasing popularization of information technologies. In this regard, the advancements in web technologies in particular has brought prominent variations in the formal and informal strands of scholarly communication. A few of these changes mentioned in their research involved shift from print to electronic publishing, emergence of institutional repositories and open access publishing, shift in libraries from buying individual journals to subscribing the electronic databases of publishers and the popularization of utilizing the academic social networking sites. **Shrivastava and Mahajan** (2021) also mentioned that with the advent of Web 2.0 tools and particularly social media, the researchers have become increasingly active on web. It has resulted in the transformation in the scholarly communication procedures through which the researchers share, and bookmark research works in online platforms.

Moreover, several other studies (**Lacka *et al.***, 2021; **Alamri *et al.***, 2020; **Sobaih *et al.***, 2020; **Ramzan *et al.***, 2023; **Alenezi *et al.***, 2023) have also been conducted to estimate the effectiveness of tools of social media for improving student integration in higher education. In this regard, social media networks have not remained confined to the scholarly communication but the study by **Sobaih *et al.*** (2020) also studied it as a significant contributor towards improving academic communication. Social networks such as ResearchGate, Academic.edu and even the mainstream platform such as Twitter or LinkedIn also enable the researchers to share their work, engagement with peers and attain feedback in real-time. **Adeniyi *et al.*** (2024) also discusses that ResearchGate, Facebook, Academic.edu, Mendeley, Twitter and Google+ are recognized as the most commonly known academic social networking tools. Among these, research gate emerged as the most popular academic social networking tool in this study (**Salami**, 2023; **Zhao; Li**, 2023). These results also align with previous reports which indicate that ResearchGate is the most widely used platform of social media among scholars.

2.3. Information Behaviors in Social Networks

There is a unanimity of opinion that social networks have reshaped the traditional information practices (**Blumenstock *et al.***, 2023; **Azzimonti; Fernandes**, 2023; **Ohara**, 2023). However, it enables the users to access the vast amounts of information interactively and quickly. In this regard, sharing information (whether through posts, direct messages, or posts) is a key behavior that facilitates rapid dissemination of ideas and knowledge. However, the informal and open nature of social networks can also lead to challenges. It involves spread of mis-information or the difficulty in evaluation of credibility of sources. Research by **Nuralievich** (2024) also highlighted that the modern means of communication are increasingly becoming source of rumor-based or unverified information. Despite of these challenges, social networks have become vital platforms for the exchange of information, collaboration and community building across different domains.

Bailey *et al.* (2024) also studied that the social connections influence beliefs and behaviors in the high-stake settings. Social networks empower the users (**Alemany *et al.***, 2020; **Senior *et al.***, 2022) so that they can create and share original content. However, this creation of content also facilitates the sharing of knowledge, skills, ideas, or opinions which often spark discussion and further dissemination. The users also play a critical role in the amplification of information by liking, commenting, and sharing posts. These behaviors can also result in the viral spread of information. Consequently, it reaches to a wide audience quickly. However, it also raises concerns regarding the dissemination of fake or unverified information (**Kayode-Adedeji; Nwakerendu**, 2022; **Ojha *et al.***, 2023; **Pröllochs; Feuerriegel**, 2023).

. Social networks are filled with user-generated content that widely varies in reliability. In social networks, therefore, users must often evaluate the credibility of information and its sources. This is because unlike traditional media, where information typically comes from the established outlets, the assessment of credibility of the content on social networks is based on the reputation of the content creator (**Pooja; Upadhyaya**, 2024). Such content is assessed in the context of existing of supporting evidence or the consensus within the community. Ease of sharing and viral nature of content on social media (**Atayeva**, 2024; **Saquete *et al.***, 2022) contributes towards the spreading of mis-information. Users may encounter misleading or false information (**Buchanan**, 2020) that can have significant consequences. As social networks continue to evolve, it also influences the ways in which users interact with information. It also underscores the need for ongoing research and critical engagement with these digital spaces.

3. Methodology

3.1. Data Collection

The data was collected through a systemic structure involving steps like data preprocessing, SVM classifier and predicted sentiments. Figure 2 shows this methodological roadmap.

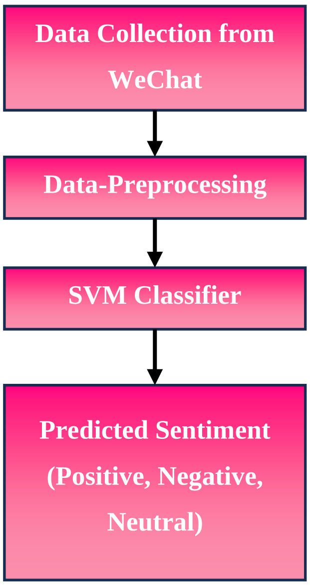

Figure 2: Methodological Roadmap.

Several engagement metrics were used to analyze overall interest of public by taking significant social interest receiving WeChat articles. The study found that articles were highly fragmented and just 7 percent of WeChat studies had more than 100,000 views. Similar results were obtained when evaluating social interest using extra engagement markers, but due to its impossibility in determining more attention receivers. the search outcomes were grouped according to reading volume. The selected WeChat articles with higher reading volume were identified to represent audience interest.

After getting the text interaction and content data from **Zhang *et al.*** (2024b) the WeChat, for responding data and filtering request the Fiddler were utilized, regular expression was created for data parsing, and a JSON structure of data was chosen for storing data. A total of 800 articles of WeChat were obtained with data set of text, title, account introduction, official account, and number of views, comments, and likes. Tencent, the parent company of WeChat, identified Xigua (西瓜) data site (http://data.xiguaji.com) as a primary participant, so we utilized their data to evaluate the active trades that has official WeChat accounts refer to the academics. A total of 400 IDs had labeled data; mostly accounts were technology-labeled, and information, finance, media, and education labeled. Out of the total IDs, 42% belonged to technology, 26 were about education, 17% were related to finance, 8% belonged to information while 7% were from media accounts. Figure 3 represents the words clouds of WeChat articles related to academics.

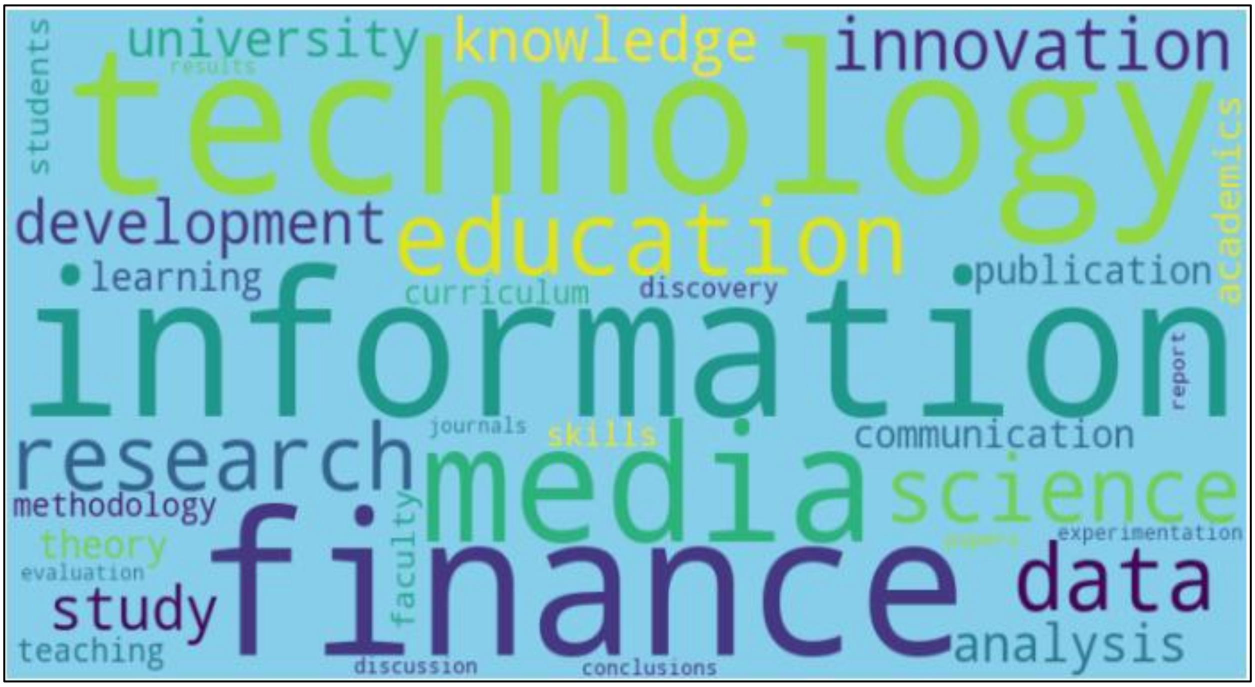

Figure 3: Words Clouds of WeChat Articles Related to Academics.

Out of the total extracted, a top 20 officially WeChat accounts were mainly established by Shanghai, Shenzhen, and Beijing companies and 12 accounts belonged to the technology. The details of the sample dataset are given in Table 1.

Table 1: Details of the Sample Dataset.

Sr. No	Accounts	Industry	Region	Read
1	36氪 (36 Krypton)	Technology	Beijing	88383
2	虎嗅APP (Tiger Snifng APP)	Technology	Beijing	36026
3	游戏陀螺 (Game Gyro)	Technology	Shenzhen	267675
4	VR陀螺 (VR Gyro)	Technology	Shenzhen	241407
5	量子学派 (Quantum School)	Technology	Shenzhen	116289
6	智东西 (Smart Stuff)	Technology	Beijing	103274
7	东西文娱 (East–West Entertainment)	Technology	Chengdu	52086
8	极客公园 (Geek Park)	Technology	Beijing	37473
9	钛媒体 (Titanium Media)	Technology	Beijing	164201
10	吉时通信 (Auspicious Time Communication)	Technology	Shanghai	170533
11	青亭网 (Green Pavilion Net)	Technology	Beijing	118340
12	维优 (VINEW)	Technology	Shanghai	300003

3.2. Data Preprocessing

Data preprocessing or text preprocessing prior to sentiment analysis mainly include phases such as data cleaning, normalization, word segmentation, tokenization, stemming/lemmatization, and vectorization. During the data cleaning phase, each unnecessary character is removed such as punctuation, numbers, URLs, and special characters. See the example below:

Example text

"**大学教育**为专业和个人的发展提供了非常宝贵的途径.https/:0"

Cleaned data

大学教育为专业和个人的发展提供了非常宝贵的途径

In the normalization phase, the lowercasing is applied and stop words are removed. The word segmentation phase also requires clear spaces in English text but in Chinese there is no clearance between words, so there is a need to divide Chinese sentences into various phrases (**Huang *et al.***, 2021). We utilize *jieba* library from python for Chinese word segmentation. See the example below:

Example cleaned text

大学教育为专业和个人的发展提供了非常宝贵的途径

Chinese word segmented data

"大学," "教育," "为," "专业," "和," "个人,", "的发展", "提供", "了非常", "宝贵", "的途径",

During the tokenization phase, the text is broken down into tokens or words individually. Stemming/lemmatization is the process of extracting or removing the last certain letters of a word, often resulting in incorrect spellings and meanings. Lemmatization takes context into account and converts words into their significant base forms, called lemmas. Vectorization refers to TF-IDF (term frequency- inverse document frequency) and it calculates the score of TF-IDF for each word in the corpus relative with that document. In vector, subsequently that information is added, so that each document in the corpus has its own vector, and the vector would have a TF-IDF score for every single term in the whole document set.

4. Results

4.1 Sentiment Analysis

Sentiment analysis in academics affects analyzing the opinions, emotions, thinking, writing and attitudes which are communicated by writers in a variety of educational contexts. This process allows faculty and institutions to measure overall views about learning, satisfaction, and engagement experience. Sentiment analysis offers an encouraging instrument to spontaneously analyze the emotions expressed through their views (**Belfin *et al.***, 2020). In this study, based on the number of reads articles was labeled a sentiment i.e., positive (+1), negative (-1), or neutral (0). Positive (+1) label will be given to favorable and encouraged view article on its topic, negative (-1) are those articles which have unsatisfactory number of views, neutral (0) in between range of satisfactory and unsatisfactory (**Azcárate**, 2023).

In our study we used mean reads and median reads for positive and negative. The Positive (+1) Sentiment included articles which had noticeably read numbers above mean plus 1 SD (standard deviation). The Negative (-1) Sentiment were articles which had noticeably read numbers below median minus 1 SD (standard deviation). There were also Neutral (0) Sentiment articles which had noticeably read numbers between both positive and negative and were considered as neutral (0).

By using eq 1, 2, and 3 we calculated positive, negative, and neutral values.

$$M = \frac{\sum_{a=1}^{z} number\ of\ read}{z} \qquad (1)$$

$$median = \frac{z+1}{2} \qquad (2)$$

$$SD = \frac{\sqrt{\sum_{a=1}^{z}(number\ of\ read - \mathrm{M})2}}{\mathrm{z}} \qquad (3)$$

Thus, number of read greater than 170,000 were considered as positive (+1), less than 60,000 were considered as negative (-1), and the ranges from 600,000 to 170,000 were considered as neutral (0). (See Table 2)

Table 2: Sample Dataset for Training.

Accounts	Industry	Region	Title	Text	Read	SA
氪Krypton	Technology	Beijing	每小时时间分解霜的大气解三十五在线监测	通常用于激光和照明技术的种独特气体是他	88383	0
虎嗅 APP Tiger Sniffing APP	Technology	Beijing	一般用于吉羽的使用程序	少数应用程序测试结果显示应用程序无法有效改 善用户语言	36026	-1
吉时通信 Auspicious Time Communication	Technology	Shanghai	国設計中的使用研究	中国吉祥图底现代设计展现有效提升文化内涵	170533	+1

4.1. Model Selection

4.1.1. SVM (Support vector machine) Classifier

SVM is a supervised learning technique that is employed in ML to execute classification tasks. The goal of the SVM technique is to get the decision boundary or best achievable line that divides distinct data classes' data points. When dealing with feature spaces of high-dimensional, hyperplane is called this boundary. The concept is to increase the margin, i.e. the interval among nearest data point and the hyperplane for each class, constructing it easier to determine the data classes (**Zhang *et al.***, 2018). Our study validates the use of SVM in sentiment analysis due to their flexibility in handling large number of data and their utility in binary classification applications. Sentiment analysis greatly benefits from support vector machines because they optimize the separation between classes, thereby improving generalization to hidden data. Moreover, SVM is inferior in overfitting, particularly when imbalanced or small datasets are involved, making it a reliable choice for particular features of WeChat academic forum data.

This study used SVM in sentiment analysis to classify text into positive (+1), negative (-1), as well as neutral sentiments categories. For our classes (+1, 0, -1) SVM utilized (OvO) strategy. SVM attempts to maximize the margin to enhance performance for classification. The support vector is the data point nearest with hyperplane, determining its direction and position. Several types of SVM have been established, but linear SVM has high popularity and significant performance for text classification (**Tan; Zhang**, 2008). Figure 4 shows the performance of SVMs model.

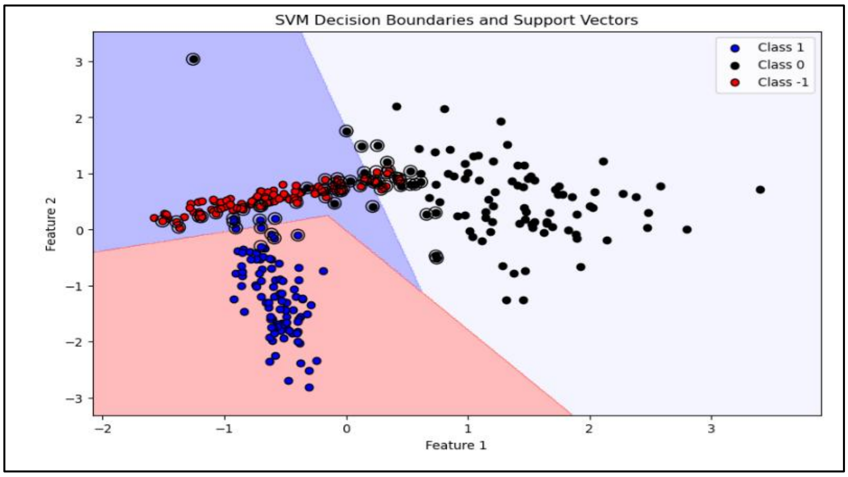

Figure 4: SVM Optimization Process.

Every case was labeled with one class $S_i \in (+1, 0, -1)$ classifying positive (+1), neutral (0), and negative (-1) sentiments. The goal of SVM is to maximize the gap among distinct sentiment categories. It involves training multiple binary classifiers for multi class tasks. $\underset{v,b}{min}\frac{1}{2}(w)^2$ determined by $S_j(v \times y_j + \beta) \geq 1$, where v is the weight-vector, S_j is label of sentiment, and β is a term bias. Classifier is trained for each couple of classes using OVO. A dual formulation is effective for enhancing computational ability is $\max_{\partial} \sum_{j=1}^{M} \partial_i - \frac{1}{2}\sum_{j=1}^{M}\sum_{k=1}^{M} \partial_j \partial_k S_j S_k (y_j + y_k)$ determined by $0 \leq \partial_j \leq C$ and $\sum_{j=1}^{M} \partial_j S_j = 0$. kernel function is used by SVM for mapping data in high-dimensions $kernel(y_{j,} y_K) = y_j \times y_K$. For classification of new sample y the decision function will be $\mathrm{D(y)} = \mathcal{A}(\sum_{j=1}^{M} \partial_j S_j) kernel(y_{j,} y_K) + \beta$.

4.2. Hyperparameter Tuning

4.2.1. Hyper-Parameters

SVM also has some hyper-parameters (**Sunkad**, 2016). Some of them were utilized in this study using Grid search algorithm, as shown in Table 3.

C: The parameter C in SVM permits in controlling the trade-off among misclassification and margin. Smaller C values promote wider margins at the expense of additional misclassifications, whereas larger C values promote accurate classifications at the expense of smaller margins.

Gamma: this parameter describes the reach of the effect of a unique training sample, where lower values mean “farther,” and higher values mean “closer.” This can be thought of as the radius’ inverse for influence of the samples that the model chooses as support vectors.

Kernel: these are the functions to calculate traffic paths. The linear option can be used if the data can be separated along a linear line, and Hyperlane is the easiest to calculate

Table 3: Tuning of Hyperparameters using Grid Search, CV (estimator=SVC).

Hyper Parameter	Values
C	[0.1, 10, 1000]
Gamma	[1, 0.01, 0.0001]
Kernel	['rbf', 'linear']

4.3. Data Splitting

Data are usually distributed to prevent overfitting. This is an example of a machine learning models that fit their training set but unable to adapt reliably to other data (**Song *et al.***, 2016). In machine learning models, raw data is usually divided into training, testing and validation groups. Whether a SVM model is trained and evaluated correctly is for ensuring the dataset is separated into a training set and a test set. In this study train test split method is employed in utilizing 80% of data for training SVM model, while the 20% is reserved for the testing of models’ performance.

4.4. Model Evaluation

Examining the SVM models’ performance includes various metrics for evaluation of model prediction in sentiment analysis (+1, 0, -1). These are important for analyzing the overall model performance and its effectiveness in a given domain, as shown in Table 4.

Table 4: Evaluation Metrics Results.

Metric	Positive (+1)	Negative (-1)	Neutral (0)
Accuracy	89%	89%	89%
Precision	90%	87%	70%
Recall	89%	70%	70%

Accuracy: Greater accuracy shows model provides accurate predictions for most events.

$$A = \frac{correct\ predictions}{total\ predictions}$$

Precision: it shows the number of instances that were predicted to be a specific sentiment actually turned out to be that sentiment. When model has low false positive rate, it shows the high precision.

$$P = \frac{TP}{TP + FP}$$

Recall: Notes the model's ability to identify all instances of a given sentiment. When model effectively found the most relevant cases it will show high recall rate

$$R = \frac{TP}{TP + FN}$$

4.5. Confusion Matrix

The confusion matrix provides a tabular structure for distinct prediction results and results of the classification problem and helps in visualizing the results, as exhibited in Table 5.

Table 5: Confusion Matrix.

	Predicted Positive	Predicted Negative	Predicted Neutral
Actual Positive	60	20	20
Actual Negative	10	15	25
Actual Neutral	15	30	22

4.6. Performance Analysis

Tan and Zhang (2008) used SVM for sentiment in Chinese and got 86% accuracy. Another study **Zhao and Wei** (2017) aimed to find response of users by clicking ‘‘thumbs up’’ in social media response, and they found there was an impact on results using WeChat social media. They suggested transforming and evolving the impact in scholarly measures. **Wang *et al.*** (2023) aimed to find the collaboration among depression and WeChat usage in Chinese elderly and middle aged for social participation. Their logistic regression model significantly improved on WeChat usage (**Huang; Sun**, 2014). On network of Weibo, implications, and information diffusion for collaborative experience in China and got effective results but their limitations were found in methodology they suggested another model.

Xue *et al.* (2021) explore how Chinese teachers of higher education adopt WeChat by creating OCoP (online communities of practice) in technology of mobile enhancement teaching for the purpose of professional learning. While OCoP had a positive impact on knowledge and practices of teachers, difficulties like lack of time were also identified. **Wu** (2014) utilized WeChat for enhancement of user wellbeing through literature review of Positive Psychology and its impact on social media. They suggested positive interventions implementation on WeChat. This study **Song *et al.*** (2016) examines online expressive behavior on China's Weibo, They found that expressive discussions, especially those including fear, sadness and anger, influence and dominate users beyond cognitive discussions, where agreeing users come together and the status of the sticker plays an important role.

5. Discussion

The present research aimed to explore the Scholarly Communication and Information Behavior in Chinese Social Networks. For this purpose, a Sentiment Analysis of WeChat (**Yang *et al.***, 2022) Academic Communities was performed. The results demonstrated that the use of SVM (Supper Vector Machine) for the sentiment analysis of WeChat articles, which were related with academics, yielded a strong predictive performance. In this regard, an overall accuracy of 89% was estimated across positive, negative, and neutral sentiments. However, the precision and recall matrix also showed that the model was particularly effective in the prediction of positive sentiments with a precision rate of 90% and a recall rate of 89%. However, the performance of model slightly diminishes for the neutral and negative sentiments. It reflects a lower recall for the negative predictions and a more moderate precision and recall for neutral predictions.

These findings are adequately aligned with the existing literature where SVM model has been resulted to be effective in handling the sentiment classification tasks (**Tripathi**, 2021; **Gopi *et al.***, 2023; **Raghunathan; Saravanakumar**, 2023), although the challenges in certain categories of sentiments still exist. Comparatively, the results of this study are also consistent with the previous studies where SVM models also achieve higher accuracy (**Chaganti *et al.***, 2020) along with possessing a few limitations in dealing with nuanced sentiment expressions. It highlights the robustness of SVM in sentiment analysis but also underscores the need for further refinement across all the categories of sentiments, particularly in the neutral classifications.

In addition, the confusion matrix further emphasizes these discrepancies by indicating those areas where the model misclassified the sentiments specially while distinguishing between negative and neutral sentiments. The results also underscore the strong accuracy of SVM model. It aligns with the research conducted by **Isnan *et al.*** (2023). They also mention that SVM is one of the best approaches that is often utilized for the sentiment analysis due to the level of accuracy that produces better results. SVM has been widely used and resulted to be impressive for sentiment analysis and tasks of text mining successfully (**Rahat *et al.***, 2019; **Fikri; Sarno**, 2019). SVM has also been considered effective for dealing with the large datasets such as those from WeChat articles. It aligns with the known strength of SVM in handling high-dimensional information and its ability to create clear boundaries of decision among distinct classes. However, the performance of this model in predicting neutral statements is comparatively weaker which is a common challenge in sentiment analysis. This is because neutral statements usually contain more nuanced or less distinctive patterns of language (**Sutoyo**, 2023). In a nutshell, the findings of this study highlight that while WeChat is a valuable platform for the dissemination of academic content, there is a need for strategies that can better target and engage specific audiences to enhance the influence of such content.

6. Research Implications

6.1. Theoretical Implications

The present study holds various theoretical implications. The findings of this study contribute to the existing literature regarding digital engagement and sentiment analysis within the contexts of social media. By demonstrating the fragmented nature of user’s engagement with the academic content on platforms such as WeChat. This study also highlights the complexities of dissemination of online content and the factors that drive user interaction. This study also supports the theories related with diffusion of information and digital divide. It suggests that even within academic circles, the content engagement and visibility are highly uneven. Lastly, this research also underscores the relevance of sentiment analysis as a tool for understanding public perceptions within educational contexts. In this way, it reinforces its value in assessing large-scale digital interactions.

6.2. Practical Implications

From a practical perspective, this study offers valuable insights for the content creators and educational institutions that aims to optimize their strategies of outreach on the platforms such as WeChat. The identification of key factors that influence article visibility and reader engagement can guide the development of more targeted content strategies particularly in the fields of education and technology where the interest is higher. Moreover, the use of sentiment analysis to gauge audience responses can also aid institutions and educators to tailor their messaging so that it can better align with the preferences of audience. Consequently, the effectiveness of academic communications on social media also improves. This approach can also inform the design of more engaging content that resonates with broader audiences thereby improving the influence of academic discourse in digital spaces.

7. Limitations and Future Research Indications

The limitation of this study is on the reliance upon WeChat data which may limit capturing the diversity of user engagement fully across different platforms of social media. It potentially limits the generalizability of the findings. Furthermore, as the study focuses on specific subsets of articles with higher reading volumes. It can also overlook the comprehensiveness of engagement with less popular content. In this regard, future researchers can explore cross-platform comparisons so that they can better understand the way academic content is consumed on different networks of social media. Investigating the role of different content formats such as interactive elements or videos can also be helpful in providing deeper insights concerning the optimization of digital academic communication strategies in driving engagement.

References

Adeniyi, Adedayo; Tiamiyu, Mutawakilu; Mbada, Kikelomo A; Adeniyi, Wuraola; Olakorede, David; Ademoyegun, Adekola B; Mbada, Chidozie E (2024). "Academic Social Networking Tools Usage Among Nigerian University Academics". *Mediterranean Journal of Education,* v. 4, n. 1, pp. 39-54. *https://doi.org/10.26220/mje.4638*

Alamri, Mahdi M; Almaiah, Mohammed Amin; Al-Rahmi, Waleed Mugahed (2020). "Social Media Applications Affecting Students' Academic Performance: A Model Developed for Sustainability in Higher Education". *Sustainability,* v. 12, n. 16, pp. 6471. *https://doi.org/10.3390/su12166471*

Alemany, José; Del Val, Elena; García-Fornes, Ana (2020). "Empowering Users Regarding the Sensitivity of their Data in Social Networks through Nudge Mechanisms." In: *Proceedings of the 53rd Hawaii International Conference on System Sciences.* pp. 2539-2548. HICSS. *http://hdl.handle.net/10125/64052*

Alenezi, Mamdouh; Wardat, Saja; Akour, Mohammed (2023). "The Need of Integrating Digital Education in Higher Education: Challenges and Opportunities". *Sustainability,* v. 15, n. 6, pp. 4782. *https://doi.org/10.3390/su15064782*

Atayeva, Nurana (2024). "Understanding Concept of Virality in Social Media from Users' Perspective." Master's thesis, University of Jyväskylä. *http://urn.fi/URN:NBN:fi:jyu-202406134593*

Azcárate, Asun López-Varela (2023). "Narratives of Sustainability in the Anthropocene. Interdisciplinary Dimensions". *Cultura,* v. 20, n. 1, pp. 7-13. *https://doi.org/10.3726/CUL012023.0001*

Azzimonti, Marina; Fernandes, Marcos (2023). "Social media networks, fake news, and polarization". *European Journal of Political Economy,* v. 76, pp. 102256. *https://doi.org/10.1016/j.ejpoleco.2022.102256*

Bailey, Michael; Johnston, Drew; Koenen, Martin; Kuchler, Theresa; Russel, Dominic; Stroebel, Johannes (2024). "Social Networks Shape Beliefs and Behavior: Evidence from Social Distancing during the COVID-19 Pandemic". *Journal of Political Economy Microeconomics,* v. 2, n. 3. *https://doi.org/10.1086/729533*

Becker, Marco; Eigenfeld, Roland; Kerpes, Thomas (2023). "Understanding the commercialization of intellectual property patents in Europe: Highlighting Implications and Regulations for the biotechnology sector". *Journal of Commercial Biotechnology,* v. 28, n. 1, pp. 252-264. *https://doi.org/10.5912/jcb1902*

Belfin, R V; Grace Mary Kanaga, E; Kundu, Suman (2020). "Application of Machine Learning in the Social Network." In: *Recent Advances in Hybrid Metaheuristics for Data Clustering.* De, Sourav; Dey, Sandip; Bhattacharyya, Siddhartha (Eds.), pp. 61-83. John Wiley & Sons Ltd. *https://doi.org/10.1002/9781119551621.ch4*

Blumenstock, Joshua E; Chi, Guanghua; Tan, Xu (2023). "Migration and the Value of Social Networks". *Review of Economic Studies,* pp. rdad113. *https://doi.org/10.1093/restud/rdad113*

Buchanan, Tom (2020). "Why do people spread false information online? The effects of message and viewer characteristics on self-reported likelihood of sharing social media disinformation". *PloS One,* v. 15, n. 10, pp. e0239666. *https://doi.org/10.1371/journal.pone.0239666*

Cao, Chun; Meng, Qian; Zhang, Huijuan (2024). "A longitudinal examination of WeChat usage intensity, behavioral engagement, and cross-cultural adjustment among international students in China". *Higher Education,* v. 87, n. 3, pp. 661-683. *https://doi.org/10.1007/s10734-023-01029-5*

Chaganti, Sai Yeshwanth; Nanda, Ipseeta; Pandi, Koteswara Rao; Prudhvith, Tavva GNRSN; Kumar, Niraj (2020). "Image Classification using SVM and CNN." In: *2020 International conference on computer science, engineering and applications (ICCSEA).* pp. 1-5. IEEE. *https://doi.org/10.1109/ICCSEA49143.2020.9132851*

Chen, Xin; Zhou, Xi; Li, Huan; Li, Jinlan; Jiang, Hua (2020). "The value of WeChat application in chronic diseases management in China". *Computer Methods and Programs in Biomedicine,* v. 196, pp. 105710. *https://doi.org/10.1016/j.cmpb.2020.105710*

Cong, Ting; Fang, Zhichao; Costas, Rodrigo (2022). "WeChat uptake of chinese scholarly journals: an analysis of CSSCI-indexed journals". *Scientometrics,* v. 127, n. 12, pp. 7091-7110. *https://doi.org/10.1007/s11192-022-04347-0*

Fikri, Mohammad; Sarno, Riyanarto (2019). "A comparative study of sentiment analysis using SVM and SentiWordNet". *Indonesian Journal of Electrical Engineering and Computer Science,* v. 13, n. 3, pp. 902-909. *https://doi.org/10.11591/ijeecs.v13.i3.pp902-909*

Gopi, Arepalli Peda; Jyothi, R Naga Sravana; Narayana, V Lakshman; Sandeep, K Satya (2023). "Classification of tweets data based on polarity using improved RBF kernel of SVM". *International Journal of Information Technology,* v. 15, n. 2, pp. 965-980. *https://doi.org/10.1007/s41870-019-00409-4*

Hailu, Milkyas; Wu, Jianhua (2021). "The use of academic social networking sites in scholarly communication: Scoping review". *Data and Information Management,* v. 5, n. 2, pp. 277-298. *https://doi.org/10.2478/dim-2020-0050*

Huang, Kaiyu; Xiao, Keli; Mo, Fengran; Jin, Bo; Liu, Zhuang; Huang, Degen (2021). "Domain-Aware Word Segmentation for Chinese Language: A Document-Level Context-Aware Model". *Transactions on Asian and Low-Resource Language Information Processing,* v. 21, n. 2, pp. 1-16. *https://doi.org/10.1145/3481298*

Huang, Ronggui; Sun, Xiaoyi (2014). "Weibo network, information diffusion and implications for collective action in China". *Information, Communication & Society,* v. 17, n. 1, pp. 86-104. *https://doi.org/10.1080/1369118X.2013.853817*

Isnan, Mahmud; Elwirehardja, Gregorius Natanael; Pardamean, Bens (2023). "Sentiment analysis for TikTok review using VADER sentiment and SVM model". *Procedia Computer Science,* v. 227, pp. 168-175. *https://doi.org/10.1016/j.procs.2023.10.514*

Kayode-Adedeji, Tolulope; Nwakerendu, Ike (2022). "The Dissemination of fake news on social media: A demographic analysis of audience involvement". *European Conference on Social Media,* v. 9, n. 1, pp. 289-297. *https://doi.org/10.34190/ecsm.9.1.166*

Lacka, Ewelina; Wong, T C; Haddoud, Mohamed Yacine (2021). "Can digital technologies improve students' efficiency? Exploring the role of Virtual Learning Environment and Social Media use in Higher Education". *Computers & Education,* v. 163, pp. 104099. *https://doi.org/10.1016/j.compedu.2020.104099*

Li, Jiawei; Xu, Qing; Cuomo, Raphael; Purushothaman, Vidya; Mackey, Tim (2020a). "Data mining and content analysis of the Chinese social media platform Weibo during the early COVID-19 outbreak: retrospective observational infoveillance study". *JMIR Public Health and Surveillance,* v. 6, n. 2, pp. e18700. *https://doi.org/10.2196/18700*

Li, Sijia; Wang, Yilin; Xue, Jia; Zhao, Nan; Zhu, Tingshao (2020b). "The Impact of COVID-19 Epidemic Declaration on Psychological Consequences: A Study on Active Weibo Users". *International Journal of Environmental Research and Public Health,* v. 17, n. 6, pp. 2032. *https://doi.org/10.3390/ijerph17062032*

Li, Wan (2024). "Data Privacy and China's “Super App” WeChat". *Penn State Journal of Law & International Affairs,* v. 12, n. 1, pp. 6. *https://elibrary.law.psu.edu/jlia/vol12/iss1/6*

Luan, Hui; Wang, Miao; Sokol, Rebeccah L; Wu, Shiyou; Victor, Bryan G; Perron, Brian E (2020). "A scoping review of WeChat to facilitate professional healthcare education in Mainland China". *Medical Education Online,* v. 25, n. 1, pp. 1782594. *https://doi.org/10.1080/10872981.2020.1782594*

Lyu, Xiaoting; Chen, Zhe; Wu, Di; Wang, Wei (2020). "Sentiment analysis on Chinese Weibo regarding COVID-19." In: *Natural Language Processing and Chinese Computing: 9th CCF International Conference, NLPCC 2020, Zhengzhou, China, October 14–18, 2020, Proceedings, Part I 9.* pp. 710-721. Springer. *https://doi.org/10.1007/978-3-030-60450-9_56*

Nuralievich, Nurmatov Azizbek (2024). "Studying the Psychological Characteristics of Teens With a Tendency to Adoption False Information in Social Networks". *International Journal of Advance Scientific Research,* v. 4, n. 03, pp. 59-63. *https://doi.org/10.37547/ijasr-04-03-13*

Ohara, Muammar Revnu (2023). "The Role of Social Media in Educational Communication Management". *Journal of Contemporary Administration and Management (ADMAN),* v. 1, n. 2, pp. 70-76. *https://doi.org/10.61100/adman.v1i2.25*

Ojha, Rudra Pratap; Srivastava, Pramod Kumar; Awasthi, Shashank; Srivastava, Vineet; Pandey, Purnendu Shekhar; Dwivedi, Radhe Shyam; Singh, Rajesh; Galletta, Antonino (2023). "Controlling of Fake Information Dissemination in Online Social Networks: An Epidemiological Approach". *IEEE Access,* v. 11, pp. 32229-32240. *https://doi.org/10.1109/ACCESS.2023.3262737*

Plantin, Jean-Christophe; De Seta, Gabriele (2019). "WeChat as infrastructure: The techno-nationalist shaping of Chinese digital platforms". *Chinese Journal of Communication,* v. 12, n. 3, pp. 257-273. *https://doi.org/10.1080/17544750.2019.1572633*

Pooja, K; Upadhyaya, Pallavi (2024). "What makes an online review credible? A systematic review of the literature and future research directions". *Management Review Quarterly,* v. 74, n. 2, pp. 627-659. *https://doi.org/10.1007/s11301-022-00312-6*

Pröllochs, Nicolas; Feuerriegel, Stefan (2023). "Mechanisms of true and false rumor sharing in social media: collective intelligence or herd behavior?". *Proceedings of the ACM on Human-Computer Interaction,* v. 7, n. CSCW2, pp. 1-38. *https://doi.org/10.1145/3610078*

Raghunathan, Nilaa; Saravanakumar, Kandasamy (2023). "Challenges and Issues in Sentiment Analysis: A Comprehensive Survey". *IEEE Access,* v. 11, pp. 69626-69642. *https://doi.org/10.1109/ACCESS.2023.3293041*

Rahat, Abdul Mohaimin; Kahir, Abdul; Masum, Abu Kaisar Mohammad (2019). "Comparison of Naive Bayes and SVM Algorithm based on sentiment analysis using review dataset." In: *2019 8th International Conference System Modeling and Advancement in Research Trends (SMART).* pp. 266-270. IEEE. *https://doi.org/10.1109/SMART46866.2019.9117512*

Ramzan, M; Javaid, Z K; Fatima, M (2023). "Empowering ESL students: Harnessing the potential of social media to enhance academic motivation in higher education". *Global Digital & Print Media Review,* v. 6, n. 2, pp. 224-237. *https://doi.org/10.31703/gdpmr.2023(VI-II).15*

Ru, Xu; Nindum, Kornkanok (2024). "The Role of Wechat Application in Chinese Family Communication: A Case Study of Chengdu City, Sichuan, China". *Asian Crime and Society Review,* v. 11, n. 2, pp. 44-54. *https://doi.org/10.14456/acsr.2024.11*

Salami, Rita Otibhor (2023). "Influence of Use of Academic Social Media on Research Activities of Library and Information Science Educators in Universities in Nigeria." Doctoral Dissertation, University in Minna. *http://repository.futminna.edu.ng:8080/jspui/handle/123456789/28272*

Saquete, Estela; Zubcoff, Jose; Gutiérrez, Yoan; Martínez-Barco, Patricio; Fernández, Javi (2022). "Why are some social-media contents more popular than others? Opinion and association rules mining applied to virality patterns discovery". *Expert Systems with Applications,* v. 197, pp. 116676. *https://doi.org/10.1016/j.eswa.2022.116676*

Senior, Isaac Ampofo Atta; Junior, Isaac Ampofo Atta; Ampofo, Beatrice (2022). "Social Media Use for Information Management in Organization and How It Can Engage and Empower Employees in The Era of COVID-19". *Archives of Infect Diseases & Therapy,* v. 6, n. 1, pp. 105-113. *https://www.opastpublishers.com/open-access-articles/social-media-use-for-information-management-in-organization-and-how-it-can-engage-and-empower-employees-in-the-era-of-co.pdf*

Shrivastava, Rishabh; Mahajan, Preeti (2021). "Influence of social networking sites on scholarly communication: A study using literature in Artificial Intelligence". *Journal of Librarianship and Information Science,* v. 53, n. 3, pp. 522-529. *https://doi.org/10.1177/0961000616678309*

Sobaih, Abu Elnasr E; Hasanein, Ahmed M; Abu Elnasr, Ahmed E (2020). "Responses to COVID-19 in Higher Education: Social Media Usage for Sustaining Formal Academic Communication in Developing Countries". *Sustainability,* v. 12, n. 16, pp. 6520. *https://doi.org/10.3390/su12166520*

Song, Yunya; Dai, Xin-Yu; Wang, Jia (2016). "Not all emotions are created equal: Expressive behavior of the networked public on China's social media site". *Computers in Human Behavior,* v. 60, pp. 525-533. *https://doi.org/10.1016/j.chb.2016.02.086*

Statista (2024). "WeChat - Statistics & Facts." *https://www.statista.com/topics/9085/wechat*

Sunkad, Zubin A (2016). "Feature selection and hyperparameter optimization of SVM for human activity recognition." In: *2016 3rd International Conference on Soft Computing & Machine Intelligence (ISCMI).* pp. 104-109. IEEE. *https://doi.org/10.1109/ISCMI.2016.30*

Sutoyo, Muhammad (2023). "Practicing Values of Philosophical Sufism in the Pencak Silat of Brotherhood Faithful Heart of Terate". *European Journal for Philosophy of Religion,* v. 15, n. 1, pp. 1-18. *https://doi.org/10.24204/ejpr.2023.4110*

Tan, Songbo; Zhang, Jin (2008). "An Empirical Study of Sentiment Analysis for Chinese Documents". *Expert Systems with Applications,* v. 34, n. 4, pp. 2622-2629. *https://doi.org/10.1016/j.eswa.2007.05.028*

Tripathi, Milan (2021). "Sentiment Analysis of Nepali COVID19 Tweets Using NB, SVM AND LSTM". *Journal of Artificial Intelligence,* v. 3, n. 03, pp. 151-168. *https://doi.org/10.36548/jaicn.2021.3.001*

Wang, Gaoling; Duan, Jing; Kan, Qianqian; Zhou, Yuqin; Cheng, Zhaopeng; Tang, Shaoliang (2023). "The correlation analysis of WeChat usage and depression among the middle-aged and elderly in China: the mediating role of social participation". *BMC Public Health,* v. 23, n. 1, pp. 462. *https://doi.org/10.1186/s12889-023-15349-9*

Wang, Junze; Zhou, Ying; Zhang, Wei; Evans, Richard; Zhu, Chengyan (2020). "Concerns Expressed by Chinese Social Media Users During the COVID-19 Pandemic: Content Analysis of Sina Weibo Microblogging Data". *Journal of Medical Internet Research,* v. 22, n. 11, pp. e22152. *https://doi.org/10.2196/22152*

Wang, Xiaoji; Jiang, Wenying (2024). "Exploring the affordances of WeChat for Chinese cultural knowledge learning among learners of Chinese in an international exchange program". *Computer Assisted Language Learning,* v. 37, n. 3, pp. 558-584. *https://doi.org/10.1080/09588221.2022.2046612*

Wu, Jiaqi (2014). "How WeChat, the Most Popular Social Network in China, Cultivates Wellbeing". *Master of Applied Positive Psychology (MAPP) Capstone Projects*. *http://repository.upenn.edu/mapp_capstone/65*

Xue, Sijia; Hu, Xiao; Chi, Xinli; Zhang, Jiangzhi (2021). "Building an online community of practice through WeChat for teacher professional learning". *Professional Development in Education,* v. 47, n. 4, pp. 613-637. *https://doi.org/10.1080/19415257.2019.1647273*

Yang, Zaoli; Wu, Qingyang; Venkatachalam, K; Li, Yuchen; Xu, Bing; Trojovský, Pavel (2022). "Topic identification and sentiment trends in Weibo and WeChat content related to intellectual property in China". *Technological Forecasting and Social Change,* v. 184, pp. 121980. *https://doi.org/10.1016/j.techfore.2022.121980*

Zhang, Chang; Zhang, Dechun; Shao, Hsuan Lei (2023). "The softening of Chinese digital propaganda: Evidence from the People's Daily Weibo account during the pandemic". *Frontiers in Psychology,* v. 14, pp. 1049671. *https://doi.org/10.3389/fpsyg.2023.1049671*

Zhang, Heng; Li, Hongxiu; Li, Chenglong; Lu, Xinyuan (2024a). "Disentangling user fatigue in WeChat use: the configurational interplay of fear of missing out and overload". *Internet Research,* v. 34, n. 7, pp. 160-186. *https://doi.org/10.1108/INTR-07-2023-0565*

Zhang, Jinman; Quan-Haase, Anabel (2022). "WeChat." In: *The SAGE Handbook of Social Media Research Methods.* Quan-Haase, A.; Sloan, L. (Eds.), pp. 598-613. London: Sage. *https://ir.lib.uwo.ca/fimspub/371*

Zhang, Sheng Tai; Wang, Fei Fei; Duo, Fan; Zhang, Ju Liang (2018). "Research on the Majority Decision Algorithm based on WeChat sentiment classification". *Journal of Intelligent & Fuzzy Systems,* v. 35, n. 3, pp. 2975-2984. *https://doi.org/10.3233/JIFS-169653*

Zhang, Yang; Xie, Yinghua; Li, Longfei; Liang, Yian; Yu, Houqiang (2024b). "How is public discussion as reflected in WeChat articles different from scholarly research in China? An empirical study of metaverse". *Scientometrics,* v. 129, n. 1, pp. 473-495. *https://doi.org/10.1007/s11192-023-04892-2*

Zhao, Rongying; Wei, Mingkun (2017). "Academic impact evaluation of Wechat in view of social media perspective". *Scientometrics,* v. 112, n. 3, pp. 1777-1791. *https://doi.org/10.1007/s11192-017-2439-2*

Zhao, Xi; Li, Na (2023). "Biological Genomics-Driven Pathway Analysis in Modern Nursing". *Journal of Commercial Biotechnology,* v. 28, n. 3, pp. 212-219. *https://doi.org/10.5912/jcb1592*

Zhou, Rui; Hentschel, Jasmine; Kumar, Neha (2017). "Goodbye Text, Hello Emoji: Mobile Communication on WeChat in China." In: *Proceedings of the 2017 CHI conference on human factors in computing systems.* pp. 748-759. ACM Digital Library. *https://doi.org/10.1145/3025453.3025800*