

Factors Influencing Public Attention to Media Convergence: Evidence from the Big Data Analysis of Baidu Index

Xiangkai Qiu^{1, 2, *}

¹School of Journalism and Communication, Minjiang University, Fuzhou 350108, China

² Research Institute of History for Science and Technology, Nanjing University of Information Science & Technology, Nanjing 210044, China

*Corresponding author. Email: daduing@qq.com

ABSTRACT

People currently spend many hours a day surfing the internet to obtain information they are interested in. Internet search behavior reflects the public's opinion on issues, namely, public attention. Its application in various fields has become extensive, and the analysis of its influencing factors has been highly valued by scholars. This paper aims to study the influencing factors of public attention by taking media convergence as an example. Daily media convergence public attention of China (excluding Hong Kong, Macao and Taiwan) from 2013 to 2020 was collected. This paper introduces the elastic coefficient, Gini coefficient and geographic concentration index to analyze the spatial-temporal characteristics of media convergence public attention. Multiple linear regression model is used to study the influencing factors of media convergence public attention. The relationships between the variables were tested using SPSS 22 software. The results show that there is a significant relationship between media convergence public attention and per capita GDP, government policy support as well as education level, which affect media convergence public attention greatly with coefficients of 2.849, 2.466 and 2.094. However, the influence intensity of internet development level on media convergence public attention is -0.325. The coefficient of determination was found to be 0.927, while the F test value was 6.335. The result shows that media convergence public attention is positively associated with economic development level, local government policy support and education level. On the contrary, the direction of the relationship between internet development level and media convergence public attention is negative. The spatial-temporal characteristics of media convergence public attention are greatly affected by macro policies. Therefore, we should strengthen macro policy guidance, promote the healthy development of media convergence industry, and increase the coverage of media convergence publicity in new media.

Keywords: Media convergence, Public attention, Baidu index.

1. INTRODUCTION

With the rise of big data applications, people think highly of public attention. Public attention reflects the public's views or attitudes towards an event or a phenomenon in a certain period of time. Fair use and appropriate analysis of public attention data is helpful for market operators to understand users' feedback and concerns, which is also conducive to promoting government departments to accurately identify public needs, to respond to the voice of the general public and online public opinion so as to improve the standard and quality of government services.

By applying internet search big data to predict and analyze the public's awareness and behavior of an event in various fields, domestic scholars have achieved a lot of encouraging research results. Taking keywords as statistical objects, Baidu index scientifically analyzes and calculates the weighted sum of search frequency of each keyword in Baidu web search based on the search volume of internet users, which can be used to predict stock trend, analyze the public attention of scenic spots, monitor online public opinion, predict novel coronavirus pneumonia and so on [1-4]. Corresponding to Baidu index, most foreign scholars conduct research based on Google Trends, a research field that bears similarity to

what domestic scholars have been doing on Baidu. By comparison the scopes of international researchers are broader, including forecasting unemployment rate, marketing decision-makings, communications, health etc. [5-10]. These studies show that the internet search data can reflect the public's views, attitudes, emotions and interests well.

In order to study the influencing factors of public attention, this paper takes the media convergence as an example, uses Baidu index to establish a Multiple Linear Regression model to explore the influencing factors of media convergence public attention in China (excluding Hong Kong, Macao and Taiwan), and analyzes its spatial-temporal evolution trend.

2. DATA SOURCES AND RESEARCH METHODS

2.1. Data Sources

2.1.1. Search Index

The search index data is from the official website of Baidu index (website: <http://index.baidu.com/v2/index.html/>) from January 1, 2013 to December 31, 2020. The geographical space of the study covers 31 provinces in China (excluding Hong Kong, Macao and Taiwan).

2.1.2. Data Pre-processing

The concept of media convergence originated in the West in the 1970s. Around 2005, it was first introduced into China by researcher Cai Wen of Renmin University of China. Since then, media convergence has been gradually attracting public attention. In China, the concepts of "Meijie" and "Meiti" (with subtle differences, which have a common meaning "media") often cause confusion in academic circles. Most scholars use the concept of "Meijie convergence", while a few scholars use "Meiti convergence". The other scholars use both of them, who believe that "Meijie convergence" is "Meiti convergence"^[11]. Due to historical reasons, it is difficult for ordinary netizens to distinguish them conceptually. Therefore, the Baidu index (hereinafter referred to as media convergence public attention) is studied in this paper taking "Meiti convergence" and "Meijie convergence" as keywords. The formula is as following: media convergence public attention = (Baidu index of "Meiti convergence" + Baidu index of "Meiti convergence")

2.1.3. Internet Penetration and Internet Population

The internet development level and internet population used in this paper is adopted from the

Statistical Reports on the Development of Internet in China issued by CNNIC in 2013-2020.

2.1.4. Demographic Data and National Economic Data

Both demographic data (including regional population, per capita GDP, financial expenditure on culture, sports and media, population with junior college and above) and national economic data are from the statistical yearbook of 2013-2020 issued by the National Bureau of Statistics (website: <http://www.stats.gov.cn>).

2.2. Research Methods

According to StatCounter, Baidu has been the undisputed leader in China's search engine industry for many years, whose average market share made up 69.46% of the online search market in 2019. In China, the data from Baidu search engine can roughly reflect the overall public attention behavior of the domestic public.

Baidu index released by Baidu Inc. is a big data visualization product for capturing users' search behavior based on Baidu's massive internet search data. By analyzing the intensity of attention to the keywords searched online in the past, it can directly reflect keyword search trend and the spatial-temporal distribution of public attention. It can objectively reflect and offer insight into the needs and concerns of netizens for a certain event. In this paper, "search index" in Baidu index is used as the behavior index of media convergence public attention. The "search index" reflects the intensity of the internet users' attention to keyword search and its continuous changes.

2.2.1 Elastic Coefficient (K)

The elastic coefficient (K) is mainly used to test the annual change of internet users' overall attention to media convergence. The public attention of media convergence has a certain relationship with the number of internet users in the current period. In order to identify the change of internet users' attention to media convergence, the elasticity coefficient is used, which is calculated as follows:

$$K = \frac{\Delta S/S}{\Delta G/G} \quad (1)$$

where K is the elasticity coefficient, S and ΔS are Baidu index and its changes respectively. G and ΔG are the number of internet users in the current period and its changes separately. According to the formula, K demonstrates that the growth rate of internet users' attention to media convergence is greater than that of the number of internet users. ($k > 1$).

2.2.2. Gini Coefficient (I_G)

Gini coefficient reflects the difference degree of media convergence public attention between geographical regions. It is expressed by the following formula:

$$I_G = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}} / \bar{x} \tag{2}$$

where I_G is Gini coefficient, \bar{x} is the average value of the sample area, x_i is Baidu search value of a certain geographical area's media convergence public attention, n is the number of sample geographical areas. The larger the value of I_G is, the more significant the difference of media convergence public attention is in geographical differences.

2.2.3. Geographical Concentration Index (E_G)

Geographic concentration index is used to test the spatial concentration degree of media convergence public attention. It can be written as:

$$E_G = \sqrt{\sum_{i=1}^n p_i^2} \tag{3}$$

where E_G is the geographic concentration index, p_i is the proportion of Baidu index of a certain region in the total sample. The larger the value of E_G is, the higher the degree of geographic concentration of media convergence public attention is.

3. INFLUENCING FACTORS OF MEDIA CONVERGENCE PUBLIC ATTENTION

3.1. Hypothesis and Analysis of Influencing Factors

After reviewing previous studies and conducting an extensive literature analysis, as well as considering the data integrity and accessibility, the current author assumes that the main factors influencing the public's media convergence public attention in geographical areas are local internet development level, economic development level, local government policy support and education level.

(1) Internet development level. The internet is a necessary prerequisite for the public to pay attention to media convergence, and it is also the most convenient tool. With higher internet penetration it is more likely and convenient for the public to search for information related to media convergence through the internet so that the media convergence public attention becomes intensive. Therefore, this paper selects the internet penetration rate as the representation of internet development level. This paper hypothesizes that internet development level is associated with media convergence public attention and the direction of the relationship is positive.

(2) Economic development level. The economic development level is directly related to the development of media convergence. Generally speaking, economic development can enhance the level of media convergence, and in turn increases media convergence public attention. Therefore, this paper selects per capita GDP as the representation of economic development level. This paper hypothesizes that economic development level is related to the media convergence public attention and the direction of the relationship is positive.

(3) Government policy support. Media convergence needs a lot of government investment and supporting policies. Generally speaking, the better economic development can help increase the proportion of investment in media convergence in GDP and the investment in publicity on the internet. It is increasingly likely for the public to pay attention to media convergence through the internet so that the media convergence public attention will be more intensive. Therefore, this paper selects the ratio of financial expenditure on culture, sports and media to local GDP as the representation of government policy support, hypothesizing that there is a critical relationship between government policy support and media convergence public attention and the direction of the relationship is positive.

(4) Education level. The level of education affects the public choice of media to some extent. Generally speaking, the higher the education level is, the more they are willing to choose the new media. L. S. Ge et al. found that there was a significant correlation between educational background and news audience type. There are more high-tech media users with higher educational background, and more traditional media users with lower educational background [12]. Specifically, education level is represented by the proportion of the population with junior college and above in the population over 6 years old. This paper hypothesizes that there is a significant relationship between the education level of the public and the media convergence public attention, and the direction of the relationship is positive.

3.2. Model

In order to test the impact of the above four hypothesized factors (internet development level, economic development level, government policy support and education level) on media convergence public attention, Multiple Linear Regression model is used. The formula can be written as:

$$\ln y_{it} = c + a_1 \ln x_{1it} + a_2 \ln x_{2it} + a_3 \ln x_{3it} + a_4 \ln x_{4it} + \delta_{it} \tag{4}$$

where y_{it} is Baidu index in a region at a certain time. $x_{1it} \sim x_{4it}$ is internet development level, per capita

GDP, government policy support and educational level. $a_1 \sim a_4$ is the influence coefficient of the four factors, C is constant term, δ_{it} is random error, i.e. residual error. Some dependent variables contain inconsistent data dimensions. In order to decrease the error effect, natural logarithm is taken for all dependent variables and independent variables.

4. ANALYSIS OF DATA FEATURES

4.1. Annual Difference of Total Public Attention Media Convergence

The data shows the annual evolution trend of search index, PC search index and mobile search index in Baidu index with "Meijie convergence" and "Meiti convergence" as keywords. It illustrates the evolution trend of the public attention media convergence in China (excluding Hong Kong, Macao and Taiwan) from 2013 to 2020.

(1) From 2013 to 2020, the public's media convergence public attention is on the rise with great changes, which indicates that China's related industries of media convergence may be in a rapid rising period, especially in 2013-2014 and 2018-2019. We can see clearly that the search index curves are particularly steep during these two periods, with an increase of 54.6% and 35.3% respectively. The most important reason for this phenomenon is that, on August 18, 2014, the Guidance on Promoting the Integration of Traditional and Emerging Media was issued, marking the media convergence up to the national strategic level. On August 21, 2018, the national publicity and ideological work conference proposed that "We should do a solid job in the construction of county-level media center to better guide and serve the masses." All counties and districts across the country have started the construction of Media Convergence Center. These two milestone events make media convergence receive more and more attention from industries and netizens. And the public's media convergence public attention shows an obvious upward trend.

(2) The mobile search index of media convergence continues to rise, accounting for 36.0% of the total search index in 2013 and more than 50% in 2018, indicating that with the popularity of mobile internet and intelligent devices, more and more people search, obtain and pay attention to the relevant information of media convergence through mobile phones and other mobile devices.

(3) In 2018, the total number of searches on PC was overtaken by the total number of searches on mobile, indicating that the public's access to and attention to the information about media convergence on the mobile

devices has gradually become the mainstream.

4.2. Annual Variation of Media Convergence Public Attention

Figure 1 shows the trend of media convergence public attention from 2013 to 2020.

(1) On the whole, the elastic coefficient of media convergence public attention presents an obvious fluctuation trend. In 2014, 2015, 2017 and 2019, the increase of elasticity coefficient was more than 1, which indicated that the public's media convergence public attention was on the rise. Especially in 2014, the elasticity coefficient was 7.35, which reached a small peak, and then fluctuated slowly until 2019 when a large increase took place again. This is consistent with the development trend of public attention presented by the overall search index in the previous analysis, indicating that in the year between 2014 and 2018, driven by two milestone events, more and more people pay attention to and learn the relevant information about media convergence through Baidu search engine.

(2) From 2014 to 2018, the elasticity coefficient decreased overall. Especially in 2016, the elasticity coefficient was less than 1, and in 2018, the elasticity coefficient was negative, which indicated that after the milestone event of 2014, the rapid development of media convergence over a period of time may have a marginal effect, and the growth rate of the public's media convergence public attention has slowed down.

(3) In 2020, the elasticity index dropped rapidly to -1.97. One reason was probably because the epidemic-related news and information occupied most of the media resources after the outbreak of novel coronavirus pneumonia (COVID-19). Thus, the public's media convergence public attention has dropped rapidly. As can be seen in the figure 1, the elastic coefficient of media convergence public attention falls dramatically. The other plausible reason was that with the deepening of media convergence, people were used to it, so the elastic coefficient of the public's media convergence public attention declined.

4.3. Characteristics of Spatial Distribution of Media Convergence Public Attention

Baidu search index value has a great relationship with the total search volume in this area. With a smaller population in the region, the search index is lower and vice versa. Therefore, the search index cannot effectively represent the real scope of concern of the region on media convergence. Based on media convergence public attention, the analysis index of media convergence public attention per million population in the region is calculated. The population data of each province in 2020 has not been released yet.

Table 1. Provincial Gini coefficient of media convergence public attention from 2013 to 2020

Year	Provincial Gini coefficient
2013	0.9123
2014	0.8083
2015	0.8154
2016	0.8235
2017	0.7608
2018	0.6743
2019	0.5682
2020	0.6586

4.5. Geographical Concentration Changes of Media Convergence Public Attention

Geographical concentration index can effectively reflect the agglomeration of things in spatial geography. In this paper, the geographic concentration index is used to quantitatively test the spatial geographic concentration of media convergence public attention.

Table 2. Geographic concentration index of media convergence public attention from 2013 to 2020

Year	Geographic concentration index
2013	0.2431
2014	0.2309
2015	0.2317
2016	0.2327
2017	0.2257
2018	0.2166
2019	0.2066
2020	0.2151

It can be seen from table 2 and figure 4 that:

(1) From the geographical concentration index of 2013-2020, the values are all small, which indicates that the concentration of media convergence public attention in geography is low, and it is still relatively scattered in different cities in the country. Under the guidance of the national policy in the field of media, media convergence has been developed rapidly in various provinces and cities, and has been widely concerned by the masses throughout the country.

(2) On the whole, the national geographic concentration index shows a downward trend, showing that the difference of media convergence public attention between regions is gradually narrowing. With the continuous development of media convergence, more and more provinces and cities pay more attention to the whole industrial layout of media convergence. Especially in recent years, with the rapid development of mobile media and other new media, the arrival rate of media has been improved. At the same time, it also

makes it more convenient for the general public to search and pay attention to the relevant information of media convergence from the internet. This situation of mutual growth and prosperity further promotes the balance of media convergence and public attention.

(3) It is worth noting that, as shown in Figure 4, after the two milestone events mentioned above in 2014 and 2018, the downward trend of geographical concentration index did not continue. After the two milestone events of 2014 and 2018, the geographical concentration index did not decrease but increased, showing an expanding trend, which shows that the phenomenon of regional differentiation is likely to occur under the stimulation of government policies.

5. CONCLUSION AND SUGGESTIONS

5.1. Conclusion

In view of the integrity and availability of the data, this paper analyzes the internet development level, local economic development level, local government policy support and education level from 2013 to 2019. SPSS statistics 22 was used to test the intensity of multicollinearity in Multiple Linear Regression model by variance expansion factor (VIF). If the value is less than 10, it is considered that there is no collinearity between variables.

Table 3. GLS regression results of model (4)

Item	Result
Constant value	-3.781
Internet development level	-0.325
Per Capita GDP	2.849
Education level	2.094
Government policy support	2.466
$R^2 = 0.927 \quad F = 6.335$	

According to the regression results in Table 3, the coefficient of determination, or R^2 , indicates the value of the variance of the dependent variable that is explained by the independent variables. The coefficient of determination was 0.927, and the F test value was 6.335. The obtained value of the coefficient of determination indicates that the independent variables of the model (per capita GDP, government policy support and education level) could justify about 92.7% of media convergence public attention. A comparison of the intensity of the correlations shows that per capita GDP, education level, and government policy support have an important relationship with media convergence public attention with coefficients of 2.849, 2.094, 2.466. The direction of the relationship is positive, which is consistent with the hypothesis of this paper. Per capita GDP has the strongest relationship with media

convergence public attention with a coefficient of 2.894. Media convergence is an industry that needs huge capital investment, which is directly related to the economic development, and its infrastructure needs a lot of investment from the government and relevant departments. However, the internet development level is not consistent with the hypothesis of this paper, which has a negative relationship with the media convergence public attention. The reason is probably that most users who are interested in media convergence are young. With the emergence of new media, they are more likely to use new media such as Tiktok, Headlines Today, WeChat, Micro-blog and other media to search for information about media convergence.

5.2. Suggestions

(1) Strengthen macro policy guidance to promote the healthy development of media convergence industry. Macro policy has greatly promoted the deep development of media convergence, and also promoted the public's attention to media convergence. However, we should also pay attention to the negative effects of macro policies that may lead to geographical agglomeration, which is not conducive to the balanced development of media convergence.

(2) Increase the coverage of media convergence publicity in new media. Considering the marginal effect of the public's access to media convergence information from the traditional internet, we should constantly innovate and increase the coverage in new media such as Tiktok, Headlines Today, Micro-blog and WeChat, and constantly enhance the influence of media convergence policy publicity.

This study uses the internet search index to analyze the public's attention to media convergence in China (excluding Hong Kong, Macao and Taiwan). This paper is innovative with reliable data source and covers an extended period of time. However, the selection of factors affecting the media convergence public attention is limited by data availability, which may lead to some limitations in the analysis results and conclusions. Further improvement is expected in the follow-up study.

ACKNOWLEDGMENT

This work was supported by New Media Communication Research Center of Minjiang University (FJMJ2019A03).

REFERENCES

[1] T. Zhang, Y. Yuan, W. Zeng, Can Investor Attention Help to Predict Stock Market Volatility? An Empirical Research Based on Chinese Stock Market High-frequency Data," *Chinese Journal of Management Science*, vol. v.28; No.193, no. 11, pp.

- 195-208, 2020. DOI: <https://www.doi.org/10.16381/j.cnki.issn1003-207x.2018.0509>
- [2] Y. Xu, L. Lu, H. Zhao, Dynamic Evolution and Spatial Differences of Network Attention in Wuzhen Scenic Area," *Economic Geography*, no. 7, 2020. DOI: <https://www.doi.org/10.15957/j.cnki.jjdl.2020.07.023>
- [3] R. Cao, R. Jiang, J. Xie, Y. Zhao Monitoring the internet public opinion and evolution mechanism for urban waterlogging using big data," *Journal of Xi'an University of Technology*, no. 2, pp. 151-158, 2020. DOI: <https://www.doi.org/10.19322/j.cnki.issn.1006-4710.2020.02.003>
- [4] B. Tu, L. Wei, Y. Jia, and J. Qian, Using baidu search values to monitor and predict the confirmed cases of covid-19 in china - evidence from baidu index," 2020. DOI: <https://www.doi.org/10.1186/s12879-020-05740-x>
- [5] M. Y. Huang, R. R. Rojas, and P. D. Convery, Forecasting stock market movements using google trend searches," *Empirical Economics*, vol. 59, no. 2, 2020. DOI: <https://www.doi.org/10.1007/s00181-019-01725-1>
- [6] A. Naccarato, S. Falorsi, S. Loriga, and A. Pierini, Combining official and google trends data to forecast the italian youth unemployment rate," *Technological Forecasting and Social Change*, vol.130, no. MAY, pp. 114-122, 2018. *Research on Influencing Factors of Network Attention 11*. DOI: <https://doi.org/10.1016/j.techfore.2017.11.022>
- [7] Brittany, Ward, Max, Ward, Boris, and Paskhover, Google trends as a resource for informing plastic surgery marketing decisions," *Aesthetic Plastic Surgery*, 2018. DOI: <https://www.doi.org/10.1007/s00266-017-1019-4>
- [8] S. P. Jun, H. S. Yoo, and S. Choi, Ten years of research change using google trends: From the perspective of big data utilizations and applications," *Technological Forecasting and Social Change*, vol. 130, no. MAY, pp. 69-87, 2018. DOI: <https://doi.org/10.1016/j.techfore.2017.11.009>
- [9] Carta, Salvatore, Medda, Andrea, Pili, Alessio, Reforgiato, Recupero, Diego, and Saia, Forecasting e-commerce products prices by combining an autoregressive integrated moving average (arima) model and google trends data," *Future Internet*, 2019. DOI: <https://doi.org/10.3390/fi11010005>

- [10] N. Kalatzis, I. Roussaki, C. Matsoukas, M. Paraskevopoulos, and S. Tonoli, Social media and google trends in support of audience analytics: Methodology and architecture," in DATA ANALYTICS 2018, The Seventh International Conference on Data Analytics, 2018.
- [11] X. Wan, Y. Li, Evolution from media fusion to the media convergence of cognition and conception," Science and Technology for China's Mass Media, vol. 000, no. 004, pp. 59-60,79, 2017. DOI: <https://www.doi.org/10.19483/j.cnki.11-4653/n.2017.04.009>
- [12] L. Ge, G. Zhang, On Types of News Audience in Omnimedia Environment," Henan Social Sciences, no. 5, pp. 83-94, 2016. DOI: <https://www.doi.org/CNKI:SUN:HNSH.0.2016-05-014>