

# A spatial-temporal public opinion analysis of IP location disclosure on Chinese social media platforms using Weibo data

Hongyu Zhang

hongyu.zhang@mcgill.ca

Department of Geography, McGill University  
Canada

Grant McKenzie

grant.mckenzie@mcgill.ca

Department of Geography, McGill University  
Canada

## ABSTRACT

On April 28, 2022, the Chinese social media platform, Weibo implemented a new feature that automatically adds a user's location (determined by IP addresses) to all posts and comments. In this work, we analyze users' reactions to this implementation. Exploratory spatial-temporal analysis was conducted on a wide range of content with the goal of understanding the general trends and major themes of the discussion. A Latent Dirichlet Allocation (LDA) topic model was used to extract implicit topics from the discourse. Results indicate that both supporters and opponents of the mandatory location disclosure participated in the discussion, with females more involved than males. Location privacy concerns were also interpreted through hashtags and LDA-derived topics, and the variation between local and overseas Chinese opinions was compared. The findings of this study will aid policymakers in understanding public concerns about mandatory location disclosure and help developers implement privacy-aware designs in the context of contemporary China.

## CCS CONCEPTS

• **Human-centered computing** → **Empirical studies in collaborative and social computing**; • **Security and privacy** → *Human and societal aspects of security and privacy.*

## KEYWORDS

IP location, public opinion, Weibo, location privacy, spatial-temporal analysis, topic model

## 1 BACKGROUND AND MOTIVATIONS

In April 2022, all of China's major online platforms implemented the functionality for displaying user locations based on internet protocol (IP) addresses, in addition to already tightened censorship measures such as real-name registration systems [6]. The feature was first added by Weibo, commonly referred to as China's version of Twitter, on March 4, 2022<sup>1</sup> to combat misinformation about the Russia-Ukraine crisis. Originally, the function was only tested on selected users and microblogs with keywords such as Russia, Ukraine, and Kyiv. Either provinces/regions in China or countries (for overseas IP addresses) were displayed when users posted or commented on Weibo. The announcement initially received limited attention on the platform, due to the small-scale implementation. A much larger-scale reaction was captured on the *hot search list* when the feature was fully implemented on Weibo and other social media platforms such as WeChat and Douyin by the end of April.

While the state media argued that no personally identifiable information is revealed by the feature with published locations at the regional level<sup>2</sup>, some netizens still expressed concerns about users' location privacy being compromised. Mixed with voices of support to increase transparency on online platforms [6], it has been difficult to determine Weibo users' level of location privacy concerns by simply glancing through the related posts and comments. The implementation of this feature and subsequent response offer an unprecedented opportunity to investigate how, where, and who are impacted by this feature. What are the spatial-temporal trends and major themes of Weibo users' discussion on this new feature? Is there a difference of opinions between local and overseas Chinese? This short paper investigates these questions through exploratory spatial-temporal analysis and topic modelling using Weibo posts and their associated comments.

## 2 EXPLORATORY SPATIAL-TEMPORAL ANALYSIS

Through the search function and cookies on Weibo.com, we scraped a total of 59,051 microblogs and 113,175 comments about IP location from March through May 2022 using modified Python scripts<sup>3</sup>. Attributes such as publication time, username, posted content, self-disclosed location<sup>4</sup>, IP location, gender, and birthday were collected. We conducted the analysis below to understand whether the sample was representative, whether gender bias existed, and how the discussion evolved over space and time.

### 2.1 Spatial Analysis

We first calculated the sum of microblog posts and comments to examine the spatial distribution of public participation in China (Figure 1). An east-west divide by the Heihe-Tengchong line can be observed with more participation along the east coast and Guangdong province top the list. Adding population<sup>5</sup> into account, Figure 2 shows that areas with higher economic development levels (e.g., Beijing and Shanghai) also had more discussion per capita than under-developed regions such as Guizhou and Guangxi. A similar comparison between overseas Chinese<sup>6</sup> is shown in Figure 3. An ideology divide can be noticed as overseas Chinese in Southeast Asia (e.g., Malaysia, Indonesia, and Thailand) posted less about the new feature in comparison to those in western countries (e.g., U.K., U.S., and Canada). A closer examination of the texts from the U.S.

<sup>1</sup>See the announcement at <https://weibo.com/1934183965/Liait9YAp>

<sup>2</sup>See this post as an example at <https://weibo.com/2087169013/LCQJjtYEz>

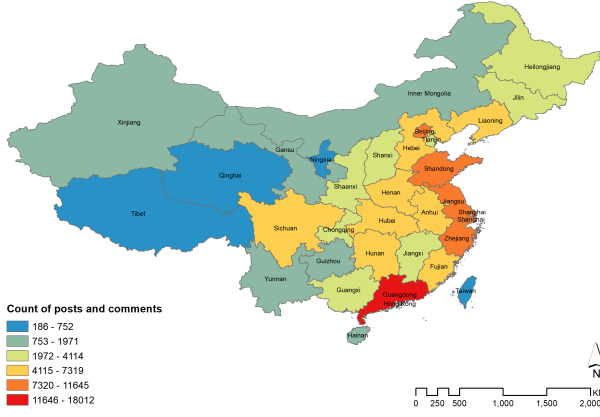
<sup>3</sup>See the original scripts at <https://github.com/Python3Spiders/WeiboSuperSpider>

<sup>4</sup>Some users choose to report locations on their profile pages.

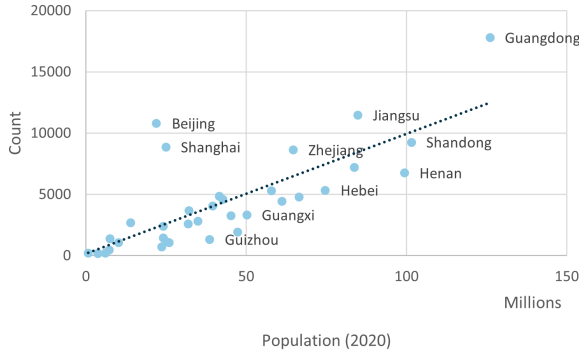
<sup>5</sup>Population data from National Bureau of Statistics (China), National Statistics (Taiwan), Census and Statistics Department (Hong Kong), and Statistics and Census Service (Macau)

<sup>6</sup>Population estimation from [https://en.wikipedia.org/wiki/Overseas\\_Chinese](https://en.wikipedia.org/wiki/Overseas_Chinese)

revealed that many posts were part of an online meme<sup>7</sup> that teases this new function and may involve manipulated IP locations, so the actual number of posts from the U.S. was likely lower than the skewed figure in Figure 3.



**Figure 1: Spatial distribution of microblog posts and comments in China**



**Figure 2: Correlation between local Chinese population and sum of posts and comments**

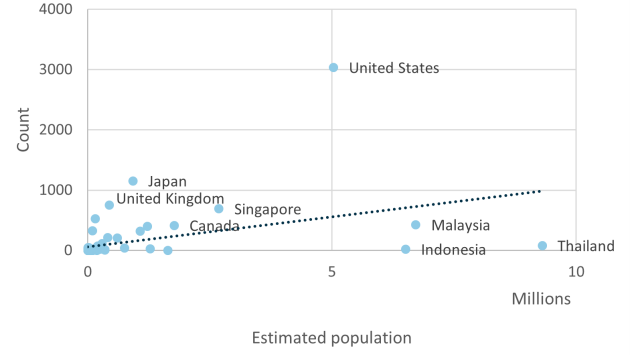
We next examined whether both genders participated in the discussion equally in our sample dataset. According to Statista<sup>8</sup>, the male-to-female (M:F) ratio of Weibo users was very close to 1 in 2021. However, the statistic does not tell the story of active users. Referencing the work of Yuan et al. [7], we computed the normalized M:F ratio  $(M:F)_N$  using the formula below:

$$(M:F)_N = \frac{(M:F)_W}{(M:F)_C}$$

where  $(M:F)_W$  and  $(M:F)_C$  represent the male-to-female ratios of our sample Weibo data and 2020 census data respectively. The

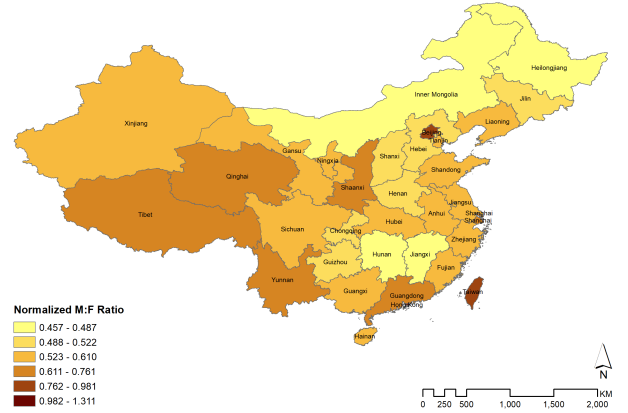
<sup>7</sup>The meme literally translates to “I am in the U.S., just went off the plane”, which is used as an introduction on social media to indicate the users’ “noble” identities.

<sup>8</sup>See <https://www.statista.com/statistics/1287809/sina-weibo-user-gender-distribution-worldwide/>



**Figure 3: Correlation between overseas Chinese population and sum of posts and comments**

average  $(M:F)_N$  was 0.62, indicating that women were more active in expressing their opinions on this topic in general. This trend was consistent across Chinese regions except for Hong Kong and Taiwan, where the  $(M:F)_N$  of the former was 1.31 and the latter was 0.98. Figure 4 demonstrates the gender difference on a map. The result was similar to [7], which discovered that female users were more likely to share their locations on Weibo. It is therefore not a surprise that women were more engaged in this discussion as females also disclose their locations on the platform more frequently [7].



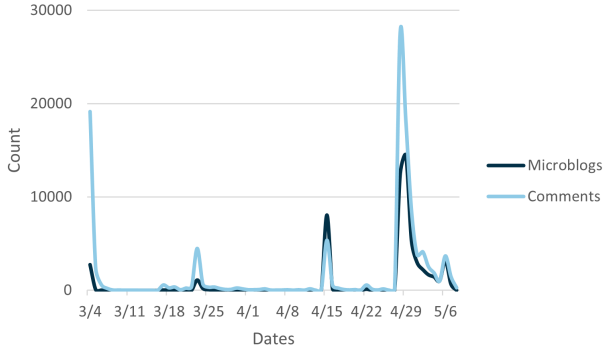
**Figure 4: Normalized M:F ratio by province and region**

## 2.2 Temporal Analysis

We next analyzed the temporal trend of related discourse. Figure 5 displays the count of collected posts and comments over time. Five peaks can be observed from the graph, namely the initial announcement (March 4), the feature update (March 23), the planned feature expansion to multiple platforms (April 15), the feature release across Chinese social media (April 28), and the popularity of IP proxy services reported by the media (May 6).

**Table 1: Selected hot topics over time**

Time periods	Topics with hashtags	Number of mentions
Early March	Weibo plans to launch real geographic location function	2720
	Russia Ukraine	49
Mid April	Douyin accounts will display IP locations	6363
	Xiaohongshu accounts will display IP locations	1900
	Toutiao accounts will display IP locations	124
	Douyin accounts' IP locations do not support manual ON/OFF	314
End of April	Weibo fully launch IP location function	10204
	Weibo's IP location function is to defend the good order of the Internet	4953
	WeChat's public platform is about to show users' IP locations	3156
	Zhihu announced the fully launch of displaying users' IP locations	1293
	What does the launch of the platforms' IP location function mean	3121
	Weibo accounts' IP locations do not support manual ON/OFF	55
Early May	After displaying IP locations, the deformed IP proxy industry went viral	2335
	Entertainment	57

**Figure 5: Number of related microblogs and their associated comments over time**

A Python script was written to count the number of mentions of topics with hashtags by dates. Selected hot topics are presented in Table 1, which match the temporal trend illustrated in Figure 5. “Russia Ukraine” appeared in early March as the feature was originally designed for filtering misinformation about the Russia-Ukraine crisis. In April, platform names such as Douyin<sup>9</sup>, Xiaohongshu<sup>10</sup>, Zhihu<sup>11</sup>, Toutiao<sup>12</sup>, and WeChat were frequently mentioned, which indicates the extensiveness of the feature implementation. Users soon turned to IP proxy services to protect their location privacy, and the high demands of these services was identified by the media in early May. Although there were debates about whether disclosing IP locations at the provincial level violates users’ location privacy, the privacy of some Chinese celebrities was already compromised as many posts under the “Entertainment” tag were about tracking where the celebrities were using Weibo’s new feature<sup>13</sup>.

<sup>9</sup>The local version of TikTok

<sup>10</sup>The Chinese equivalent of Instagram

<sup>11</sup>A Q&A website that is similar to Quora

<sup>12</sup>A news feed platform

<sup>13</sup>See this example: [https://www.sohu.com/a/545303325\\_120884185](https://www.sohu.com/a/545303325_120884185)

### 3 TOPIC MODELLING

While the spatial-temporal analysis provided a general picture, the actual contents of posts require additional attention. Textual analysis was necessary to further investigate the public opinions. Multiple studies have successfully implemented automated textual analysis with Weibo data. The majority of the related literature (e.g., [1], [3], [4], [5]) used Latent Dirichlet allocation (LDA) to cluster and understand public opinions. As a generative Bayesian probabilistic model [2], LDA is capable of discovering implicit topics from text sets.

After cleaning the data, Chinese text segmentation tool, Jieba<sup>14</sup> was used to extract features. A self-defined dictionary was written to help with the segmentation task. Stop words were removed by searching through stop word dictionaries, and only nouns, verbs, and adjectives that are longer than one Chinese character were kept. Next, the hyperparameters were tuned using a randomly-sampled subset through grid search. The best coherence score was found when the number of topics ( $n$ ) was three. We also picked twenty-five as the additional number because we wanted to examine more topics, and the results deteriorated with a higher number. We then built our LDA models using posts and comments from local Chinese users. The  $u\_mass$  coherence scores were -3.5823 ( $n = 3$ ) and -10.8251 ( $n = 25$ ). Among the generated topics, most were neutral or contained mixed reactions, which matched the general impression by journalists that both frustration and relief were observed on the platform [6]. To understand whether there was a regional difference, topic distributions of posts and comments from overseas Chinese were inferred using the previously built models. Since this paper focuses on location privacy, we selected topics that expressed users’ concerns about the new feature. Table 2 shows the relevant topics derived from the two LDA models and the average probabilities of the local ( $\bar{P}_C$ ) and overseas ( $\bar{P}_O$ ) text sets. The focuses of the two groups were different according to the bold probabilities. While local Chinese users mentioned “privacy”, more attention was paid to expressing emotions. For example, topic 3-3 indicates users’

<sup>14</sup>See <https://github.com/fxsjy/jieba>

Table 2: Selected LDA-derived topics

n	#	Keywords	$\bar{P}_C$	$\bar{P}_O$
3	1	Weibo, display, IP location, location, function, user, comment, Douyin, exhibit, address	0.4571	<b>0.4691</b>
3	2	network, IP proxy, platform, feeling, seller, pay, mean (v.), like, business, data	0.3240	<b>0.3426</b>
3	3	privacy, opinion, IP location, irrelevant, quality, region, Internet, expose, on the Internet, province	<b>0.2188</b>	0.1883
25	4	expose, try, useless, read, cannot wait, IP location, IP address, open, Douyin, enhance	0.0312	<b>0.0345</b>
25	8	unwilling, mandatory, location, version, Kyiv, avoid, delete, appearance, self-certification, collusion	0.0338	<b>0.0393</b>
25	11	Internet, real-name system, feel, on the Internet, hide, fun, game, devil, cannot open, IP location	<b>0.0450</b>	0.0384
25	25	information, social, media, keyword, fake, login, cost, overwhelm, mark (v.), give up	0.0333	<b>0.0336</b>

worries about location-based stereotypes (“irrelevant”, “quality”, “region”). Topic 25-11, on the other hand, suggests users were on the fence: some wanted to “hide” from the “real-name system”, but others treated the new feature as a “fun game”. Overseas Chinese users also shared their negative feelings (e.g., topic 25-4 and 25-25), however, more actions were taken (e.g., “IP proxy” in topic 3-2 and “avoid”, “delete” in topic 25-8). Thus, critics living abroad may be more capable and active in protecting their location privacy, while local users may have few alternatives to the new location tagging functionality.

#### 4 CONCLUSION

The feature of displaying users’ locations based on their IP addresses gained a lot of Internet traffic when it was released in March and updated in April. The suddenly expanded feature implementation across Chinese social media platforms was a major factor that contributed to the lively discussion on Weibo. Overall, opposite opinions were identified as some voiced their support and others disapproved. Densely populated areas on the east coast of China as well as countries with mainstream western ideologies had relatively more users participated in the discussion. Except for Hong Kong and Taiwan, results showed that female users were the dominant force in the conversation. Users’ location privacy concerns were twofold. An analysis of hashtags suggests that IP proxy services went viral as users scrambled to respond to the pervasive implementation of location disclosure. Through the lens of LDA topic modelling, slight differences between the two discourses told the story of the proactiveness of overseas Chinese and the powerlessness of local Chinese in terms of location privacy protection.

This work presents a preliminary analysis and exploration of a timely and important topic. Although every attempt was made to collect a representative sample of data, this study does not cover the full scope of the opinions due to the limitations of the search function on Weibo.com and its API. We will continue analyzing these data, refining our research questions, and expanding on the types of analysis we have touched on thus far.

#### ACKNOWLEDGMENTS

This work was supported by the Fonds de recherche du Québec - Société et culture (FRQSC).

#### REFERENCES

- [1] Lu An, Chuanming Yu, Xia Lin, Tingyao Du, Liqin Zhou, and Gang Li. 2018. Topical evolution patterns and temporal trends of microblogs on public health emergencies: an exploratory study of Ebola on Twitter and Weibo. *Online information review* 42, 6 (2018), 821–846.
- [2] David M Blei, Andrew Y Ng, and Michael I Jordan. 2003. Latent dirichlet allocation. *Journal of machine Learning research* 3, Jan (2003), 993–1022.
- [3] Bingyang Chen, Xiao Wang, Weishan Zhang, Tao Chen, Chenyu Sun, Zhenqi Wang, and Fei-Yue Wang. 2022. Public Opinion Dynamics in Cyberspace on Russia-Ukraine War: A Case Analysis With Chinese Weibo. *IEEE Transactions on Computational Social Systems* (2022).
- [4] Xuehua Han, Juanle Wang, Min Zhang, and Xiaojie Wang. 2020. Using social media to mine and analyze public opinion related to COVID-19 in China. *International Journal of Environmental Research and Public Health* 17, 8 (2020), 2788.
- [5] Pengxiang Li, Hichang Cho, Yuren Qin, and Anfan Chen. 2021. # Metoo as a connective movement: examining the frames adopted in the anti-sexual harassment movement in China. *Social Science Computer Review* 39, 5 (2021), 1030–1049.
- [6] Xinmei Shen. 2022. Weibo’s new user location display appears to show that Western Tech Gurus are based in China. <http://www.scmp.com/tech/big-tech/article/3176272/weibos-new-policy-display-user-locations-prompts-some-humour-among>
- [7] Yihong Yuan, Guixing Wei, and Yongmei Lu. 2018. Evaluating gender representativeness of location-based social media: A case study of Weibo. *Annals of GIS* 24, 3 (2018), 163–176.