

Demographic inequalities in digital spaces in China: The case of Weibo

Wenqing Qian^{1,2}, Ole Hexel², Emilio Zagheni², Ridhi Kashyap^{3,4}, Ingmar Weber⁵

¹Institute for Social Research, University of Michigan, ²Max Planck Institute for Demographic Research, ³Department of Sociology, Leverhulme Centre for Demographic Science, ⁴Nuffield College, University of Oxford, ⁵Saarland Informatics Campus, Saarland University

Abstract

We provide the first systematic, large-scale exploration of user counts by age, gender, and location on Weibo, the world's largest microblogging platform, using data from its advertising platform. Digital platforms potentially give access to information, economic opportunities, and social participation to underserved groups, and can be used for information dissemination during emergency situations and humanitarian response. The trace data that they produce is potentially more timely, of higher spatial or temporal resolution, and available for difficult-to-reach populations, than data from traditional sources. Whether this potential is fully realized however depends on who actually participates in these online spaces. We contribute to a growing body of research investigating the demographic characteristics of social media platforms, and on the population generalizability of digital trace data in non-Western, non-English-language contexts. We show that Weibo users are disproportionately young, female, and urban. Use of Weibo is higher in areas that are more economically prosperous and with a more educated population.

Introduction

We provide the first systematic exploration of user demographics on Weibo, the world's largest micro-blogging platform (but see Gao et al. (2012); Zhang and Pentina (2012); Rizwan et al. (2018); Yuan, Wei, and Lu (2018) for earlier studies of subpopulations). Weibo claims 252 million average daily active users and half a billion monthly active users (Weibo Corporation 2023). Describing the Weibo user population contributes to our understanding of digital inequalities and of non-English-language, non-Western platforms. Furthermore, understanding demographic inequalities on Weibo is crucial for evaluating the generalizability of digital (trace) data generated from the use of the platform.

Digital inequalities matter because they may create, reinforce, or counteract offline inequalities (Hargittai 2021a; Robinson et al. 2015). Digital inequalities can stem from differences in the material and economic conditions of access – sometimes called the “first-level” digital divide – or differences in usage and skill – the “second-level” digital divide (Hargittai 2021b; Karaoglu, Hargittai, and Nguyen 2021).

The interaction of various social and demographic characteristics, chief among them age, gender, education, and income, with these two levels is a matter of ongoing research (Hargittai 2021a). For example, online platforms have the potential to provide more democratic modes of access to information, jobs, and networks to marginalized populations. However, this process is neither automatic nor uniform, and may further vary across different types of platforms. Kashyap and Verkroost (2021) show, for instance, that women are underrepresented on LinkedIn at levels consistent with more traditional labour market indicators, but that underrepresentation is lower for high-status women in very gender-inegalitarian countries. The benefits of digital platforms do not accrue automatically and not to everyone equally.

While representativeness along conventional demographic and socio-economic dimensions is increasingly well documented on different social media platforms, such as Facebook or LinkedIn, cross-country variation is less well studied. More recent work increasingly includes countries outside of North America and Western Europe in their samples and case studies (Gil-Clavel and Zagheni 2019; Mejova et al. 2018; Kashyap and Verkroost 2021; Kashyap et al. 2020; Mejova et al. 2018). However, even with these studies, the focus is often on a handful of major, usually English-speaking platforms. Studying Weibo's user population has the twofold advantage of extending the diversity of platforms studied and providing a way to study Chinese internet users who may be restricted in their access to other platforms such as Facebook.

Understanding the representativeness of digital populations across demographic characteristics is of particular importance when digital platforms become instrumental to information dissemination during emergency situations or humanitarian response, or when they are used to study hard-to-reach groups or to minimize the lag between data collection, analysis, and policy guidance (Kashyap and Zagheni 2023). For example, Chinese internet users have turned to Weibo for information dissemination during natural disasters (Qu et al. 2011; Liu, Zhang, and Zhang 2020) and to request help during the early phase of the Covid-19 pandemic (Yang et al. 2022). Grow et al. (2020) describe advantages and challenges of using Facebook to implement surveys during the Covid-19 pandemic, when timely data was essential but face-to-face surveys impossible or hard to

conduct and likely too slow. Others have used Facebook to recruit very specific or highly mobile populations (Schneider and Harknett 2022; Pöttschke and Braun 2017). In contrast to data collected through participants recruited via platforms, digital trace data that are passively generated by the use of platforms have also been increasingly used for studying social and demographic processes (Cesare et al. 2018; Kashyap and Zagheni 2023). In such cases, digital trace data can be more timely, granular, and affordable than census, administrative data, or surveys. They have been used, to mention but a few examples, to “nowcast” migrant stocks in the United States (Alexander, Polimis, and Zagheni 2022), track gender inequality on global, national, and subnational scales (Kashyap et al. 2020; Fatehkhia, Kashyap, and Weber 2018; Mejova et al. 2018), and improve estimation of development indicators (Weber, Kashyap, and Zagheni 2018), often in contexts where data collection is difficult or impossible and where other data collection methods incur considerable delays. To better assess the generalizability and biases of these data sources for population measurement, however, requires analysis of who uses platforms and who does not. The possibility of bias and limited representativeness is especially acute when the user population of a particular platform is not itself the population of interest (Blank and Lutz 2017; Hargittai 2015; Gil-Clavel and Zagheni 2019). Understanding who participates in which digital spaces and how they participate is important for our understanding both of these spaces in themselves and how they can be used to the benefit of all.

Data

We combine aggregate counts of daily active users by age and sex obtained from the Weibo advertising platform with population counts by age and sex from the 2020 census and additional socio-economic data (see Table 1). Our unit of analysis are administrative units at the prefecture levels ($n_{pref} = 335$).

	Min.	Median	Max.
Weibo Sex Ratio (WSR)	33.3	53.8	100.0
General Sex Ratio (GSR)	92.3	103.1	131.0
Total pop. (million)	0.1	3.2	32.1
Urban residents	17%	58%	100%
GDP pc (1000s CN¥)	3.4	53.3	180.9
Disp. income pc (1000s CN¥)	13.7	27.1	72.2
Education (avg. years), women	4.3	8.8	12.2
Education (avg. years), men	5.2	9.4	12.2

Table 1: Descriptive statistics for 335 prefecture-level administrative units

Weibo audience estimates Weibo provides a platform to advertisers on which they can define populations by certain demographic and other criteria in order to receive an estimate of the audience size and the expected cost of an ad campaign. We build a demographic profile of Weibo’s user population by systematically varying the characteristics of

interest (age, gender, and location) and retrieving the estimated number of users (audience counts).

We retrieve the total number of users separately for men and women in 5-year intervals from 15 years to 80 years across all provinces and prefectures. Weibo does not report user counts for ages 80 and above because of platform limitations and we exclude users below 15 years because persons under 13 years are not formally allowed to sign up and to align the Weibo data better with external demographic data on age and sex population composition. Weibo reports user counts in increments of 1 000, bottom-codes estimates smaller than 2 000, and top-codes estimates over 41×10^6 . Bottom-coding results in a considerable number of missing or uninformative user counts for certain age \times gender combinations: 52% of such combinations are missing at the prefecture level and 17% at the province level.

We cannot independently verify Weibo’s audience estimates. Users are required to report a date of birth during sign-up. To the best of our knowledge, Weibo does not verify this information at any moment. Similarly, gender can be but does not have to be reported by the user. Moreover, Weibo reports on its advertising platform that it algorithmically assigns age and gender to certain users. It is unclear whether this happens only to accounts with no self-reported information or whether self-reported information is potentially replaced. Documentation on the advertising platform and patents registered by Weibo employees indicate that assignment of demographic characteristics is based on followed accounts and interests.

For this paper, we use data collected on December 14 and 19, 2022. A single run of data collection takes several hours at the most granular level (gender \times 5-year age intervals \times location by prefecture). Our data represent thus a snapshot of the Weibo user population. This has the advantage that differences between groups, i.e. by gender, age, or location, are unlikely to be confounded by time. On the other hand, this limits the generalizability of our results beyond this period. We also collected user counts for a limited number of user population definitions (both broad age \times gender aggregates and specific locations) over several weeks, in order to gauge the stability of the numbers, and we found little variation.

Population data Population counts by age, sex, and location are from the Seventh National Population Census carried out in 2020. Data on education, income, and employment are from provincial statistical yearbooks.

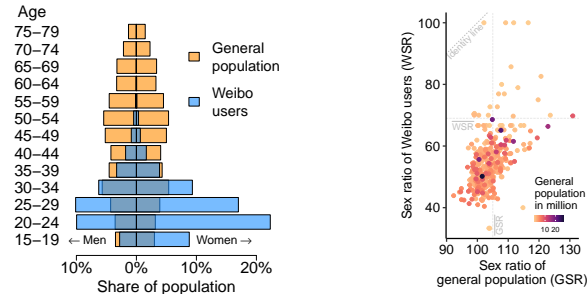
Our main demographic characteristics of interest are age and sex. To compute the sex composition of Weibo, we compute the sex ratio of Weibo users, WSR:

$$WSR = \frac{\text{male users}}{\text{female users}} \times 100 \quad (1)$$

A $WSR > 100$ indicates male users outnumbering female users and vice versa. We also calculate the sex ratio of the general population, GSR , in an analogous way. We further compute sex ratios for different age groups. This definition of the sex ratio has long been used in demography to study sex imbalances in populations (see Brian and Jaisson (2007) for an overview of different variants).

Results

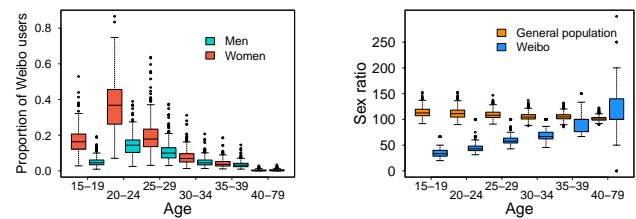
Young women are a plurality of Weibo users. The sex ratio of the Weibo user population is considerably lower than that of the general Chinese population. 64% of Weibo users are women; this corresponds to a *WSR* across all age groups and prefectures of 69, compared to a *GSR* of 104.9 (figure 1a).



(a) Women under 30 years are a plurality of Weibo users (b) Prefecture WSR and GSR are positively correlated

Figure 1: Weibo users are younger and more female than their prefecture population but WSR is positively correlated with GSR (Spearman's $\rho = 0.481$)

The low overall *WSR* is due to the combination of the young age of Weibo's user base (40% of its users are between 14 and 25 years old) with an especially low *WSR* among the youngest users (15 – 19 years: 31.7; 20 – 24 years: 44.7). Overall, the *WSR* is positively correlated with the *GSR* (Spearman's $\rho = 0.481$; Figure 1b) but its mean is much lower. The difference between the *WSR* and *GSR* is greatest in the youngest age groups and progressively decreases (Figure 2b).



(a) Women are more likely to use Weibo than men, especially under 30 years (b) The age groups with the lowest *WSR* have the highest *GSR*

Use of Weibo and the gender gap are highest among younger users. Weibo users are overwhelmingly young (40% are between 14 and 25 years old) and, among the young, women are overrepresented compared not only to their age group among their prefecture's general population but also compared to other age groups on Weibo (Figure 2a). For example, 15 to 19-year old women are 4 to 6 times as likely to use Weibo as their male peers, and 20 to 24-year old women around 3 times. In addition, it is in these

age groups that men are most overrepresented in the general population. Consequently, the gap between *WSR* and *GSR* is greater (Figure 2b). Whether this age pattern is an age, cohort, or period effect, cannot be conclusively answered with cross-sectional data. Yet, we contend that the association with education discussed below points to this being partly a cohort effect.

Weibo use is broadly associated with economic and social development. Mapping the levels of Weibo use across prefectures results in a clear geographical pattern: the proportion of Weibo users is highest in the great coastal agglomerations around Beijing, Shanghai, and Guangzhou (Figure 3). This conforms to the general pattern of economic and social development in China: more economically prosperous areas with more educated populations have a higher share of Weibo users.

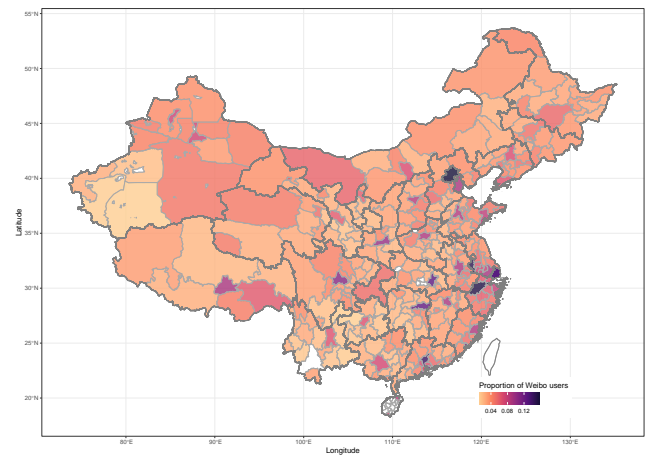


Figure 3: Proportion of population using Weibo

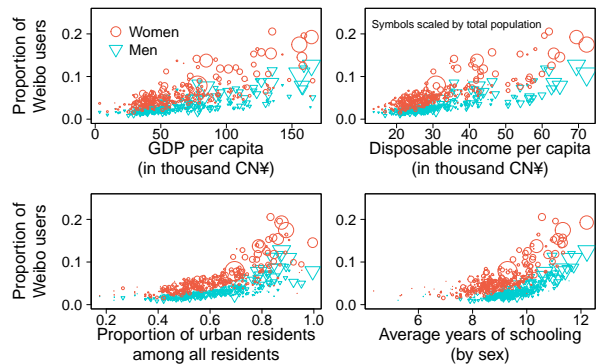


Figure 4: Proportion of Weibo users in prefecture by GDP per capita, disposable income per capita, proportion of urban residents, and average years of schooling

Figure 4 presents the share of Weibo users, separately for men and women, among the prefecture residents by GDP per capita, average disposable income, proportion of urban residents, and average years of schooling. The proportion

of Weibo users is positively correlated with these four indicators. The proportion follows similar patterns for men and women, but women show higher Weibo use across all indicators and all levels. However, the association between Weibo use and education shows a non-linear pattern: it seems stronger in prefectures with average years of schooling above 9, the duration of compulsory schooling.

Discussion

Our results present the first systematic description of the demographic characteristics of the user population of Weibo, the largest micro-blogging platform in the world. They also add to our knowledge of non-English-speaking and non-Western digital populations. We find that the proportion of women among Weibo users is substantially greater than their share of the general population. This confirms previous research on limited subsets of Weibo users (Liu and Li 2015; Zhang and Pentina 2012; Gao et al. 2012) but contrasts with the overrepresentation of men on Facebook in other Asian countries (Gil-Clavel and Zagheni 2019; Kashyap et al. 2020; Mejova et al. 2018).

Weibo users are younger and more often female than the population of their place of residence. Weibo use in general is positively correlated with indicators of economic and social development. We cannot conclusively say whether this reflects a link between Weibo use and personal economic wealth or whether socio-economic indicators measure infrastructure quality which in turn makes access easier for all residents of that area. Since 95% of Weibo users access Weibo via their smartphone and 99.6% of Chinese residents with access to the internet use the internet via their phones, it seems plausible that Weibo use is not so much conditioned by individual economic and material obstacles to access (“first-level” inequality (Hargittai 2021b; Karaoglu, Hargittai, and Nguyen 2021)) than by infrastructure and by “second level” factors, such as education.

“Second level” factors may play a role in Weibo use insofar as Weibo remains a predominantly text-based platform. Education may thus be the barrier to use and help explain why Weibo use is strongly correlated with educational composition. This would be consistent with the observed age gradient and the rapid educational expansion that has occurred after market reforms in China (Yeung 2013). Another potentially relevant “second level” factor is type of use. Research suggests a diversification of platform-specific use with digital expansion (DiMaggio et al. 2004): different user populations look to different platforms for their news, entertainment, or socializing. In China, the gender composition of different platforms is reported to vary drastically: from > 90% female users on Xiaohongshu (resembling Pinterest or Instagram) to > 90% male users on Douyu, a gaming live-streaming platform.

When considering digital platforms for information dissemination or targeting and coordination of relief efforts during emergencies, it is important to consider who is underrepresented on the platform. In the case of Weibo, the proportion of the population using the platform declines with a prefecture’s level of economic development, education, and

urban population. This suggests that population-level socioeconomic inequalities still pattern digital inequalities in the Chinese context, even as digitalization and mobile penetration have expanded. Weibo use is also very rare among people 40-year old and above.

While the cross-sectional nature of our data rules out a definitive answer to the question whether the overrepresentation of young women is an age, cohort, or period effect, we can sketch a few possibilities. For the age structure and gender imbalance to be due to an age effect would require women to have a higher propensity than men to start using Weibo at younger ages and to leave Weibo at older ages. This could be because they use Weibo for socializing and networking more than men do and that this activity is more frequent between the ages of 15 and 30 years. Women’s use of Weibo may change, especially as they assume family or care-related responsibilities, which may lead to the gender gap shrinking at ages above 30. A cohort effect could arise from educational expansion or from “early adopter” effects: more recent cohorts are more educated, with gender inequalities in education that have traditionally disfavoured women decreasing, or even reversing in urban areas (Yeung 2013; Wang 2021). These urban areas are also those where we find high levels of overall Weibo penetration. More educated cohorts are also more familiar with new technologies and different internet platforms. Together, we speculate that a combination of both age and cohort effects may help explain the observed age and gender patterns of Weibo use.

While our study provides the first large-scale exploration of Weibo combining digital trace from Weibo’s advertising platform and census data at the prefecture level to understand demographic inequalities on the platform, we acknowledge limitations of our approach. Our analysis is at the aggregate level and does not allow us to say anything specifically about individual behaviours. Microdata collected via surveys or qualitative methods are required to understand individual-level correlates and patterns of use. Furthermore, we use aggregate user counts provided by Weibo’s advertising platform. For many age*prefecture combinations, these data are sparse: 38% of women and 33% of men belong to age*prefecture bins with missing Weibo user counts. This concerns mostly older users: fewer than 5% of prefectures are missing counts for users under 35. Moreover, a Spearman ρ of 0.28 between WSR and population size indicates that this is not likely to influence our data. Another limitation results from Weibo possibly assigning user characteristics according to unknown criteria. While users may report age, gender, and other characteristics (during account creation or on their public profile), we do not know how Weibo imputes these characteristics for users who do not do so. In addition, we ignore how Weibo’s distinguishes between accounts of distinct physical persons and other accounts (e.g. bots or organizational accounts). These concerns, however, are not unique to advertising data from Weibo but are more generally applicable to social media advertising data (Weber, Kashyap, and Zagheni 2018). Future research could evaluate the accuracy of demographic targeting on Weibo’s advertising platform by comparing it with surveys conducted via the platform, as has been done for Facebook (Grow et al. 2022).

References

- Alexander, M.; Polimis, K.; and Zagheni, E. 2022. Combining Social Media and Survey Data to Nowcast Migrant Stocks in the United States. *Population Research and Policy Review*, 41(1): 1–28.
- Blank, G.; and Lutz, C. 2017. Representativeness of social media in Great Britain: investigating Facebook, LinkedIn, Twitter, Pinterest, Google+, and Instagram. *American Behavioral Scientist*, 61(7): 741–756.
- Brian, E.; and Jaisson, M. 2007. *The descent of human sex ratio at birth: a dialogue between mathematics, biology and sociology*. Dordrecht: Springer.
- Cesare, N.; Lee, H.; McCormick, T.; Spiro, E.; and Zagheni, E. 2018. Promises and pitfalls of using digital traces for demographic research. *Demography*, 55(5): 1979–1999.
- DiMaggio, P.; Hargittai, E.; Celeste, C.; Shafer, S.; et al. 2004. From unequal access to differentiated use: A literature review and agenda for research on digital inequality. *Social inequality*, 1: 355–400.
- Fatehikia, M.; Kashyap, R.; and Weber, I. 2018. Using Facebook ad data to track the global digital gender gap. *World Development*, 107: 189–209.
- Gao, Q.; Abel, F.; Houben, G.-J.; and Yu, Y. 2012. A Comparative Study of Users' Microblogging Behavior on Sina Weibo and Twitter. In Masthoff, J.; Mobasher, B.; Desmarais, M. C.; and Nkambou, R., eds., *User Modeling, Adaptation, and Personalization*, 88–101. Berlin, Heidelberg: Springer.
- Gil-Clavel, S.; and Zagheni, E. 2019. Demographic Differentials in Facebook Usage around the World. *Proceedings of the International AAAI Conference on Web and Social Media*, 13: 647–650.
- Grow, A.; Perrotta, D.; Del Fava, E.; Cimentada, J.; Rappazzo, F.; Gil-Clavel, S.; and Zagheni, E. 2020. Addressing Public Health Emergencies via Facebook Surveys: Advantages, Challenges, and Practical Considerations. *J Med Internet Res*, 22(12): e20653.
- Grow, A.; Perrotta, D.; Del Fava, E.; Cimentada, J.; Rappazzo, F.; Gil-Clavel, S.; Zagheni, E.; Flores, R. D.; Ventura, I.; Weber, I.; et al. 2022. Is Facebook's advertising data accurate enough for use in social science research? Insights from a cross-national online survey. *Journal of the Royal Statistical Society Series A*, 185(S2): S343–S363.
- Hargittai, E. 2015. Is bigger always better? Potential biases of big data derived from social network sites. *The ANNALS of the American Academy of Political and Social Science*, 659(1): 63–76.
- Hargittai, E., ed. 2021a. *Handbook of Digital Inequality*. Cheltenham, UK Northampton, MA, USA: Edward Elgar Publishing.
- Hargittai, E. 2021b. Introduction. In Hargittai, E., ed., *Handbook of Digital Inequality*, 1–7. Cheltenham, UK Northampton, MA, USA: Edward Elgar Publishing.
- Karaoglu, G.; Hargittai, E.; and Nguyen, M. H. 2021. Inequality in online job searching in the age of social media. *Information, Communication & Society*, 1–19.
- Kashyap, R.; Fatehikia, M.; Tamime, R. A.; and Weber, I. 2020. Monitoring global digital gender inequality using the online populations of Facebook and Google. *Demographic Research*, 43: 779–816.
- Kashyap, R.; and Verkroost, F. C. J. 2021. Analysing global professional gender gaps using LinkedIn advertising data. *EPJ Data Science*, 10(1): 39.
- Kashyap, R.; and Zagheni, E. 2023. Leveraging Digital and Computational Demography for Policy Insights. In *Handbook of Computational Social Science for Policy*, 327–344. Springer.
- Liu, J.; and Li, S. 2015. Characteristics Study of Weibo Users' Interactions. *Wuhan University Journal of Natural Sciences*, 20(6): 499–504.
- Liu, T.; Zhang, H.; and Zhang, H. 2020. The Impact of Social Media on Risk Communication of Disasters—A Comparative Study Based on Sina Weibo Blogs Related to Tianjin Explosion and Typhoon Pigeon. *International Journal of Environmental Research and Public Health*, 17(3).
- Mejova, Y.; Gandhi, H. R.; Rafaliya, T. J.; Sitapara, M. R.; Kashyap, R.; and Weber, I. 2018. Measuring subnational digital gender inequality in India through gender gaps in Facebook use. In *Proceedings of the 1st ACM SIGCAS Conference on Computing and Sustainable Societies*, 1–5.
- Pötzschke, S.; and Braun, M. 2017. Migrant Sampling Using Facebook Advertisements: A Case Study of Polish Migrants in Four European Countries. *Social Science Computer Review*, 35(5): 633–653.
- Qu, Y.; Huang, C.; Zhang, P.; and Zhang, J. 2011. Microblogging after a Major Disaster in China: A Case Study of the 2010 Yushu Earthquake. In *Proceedings of the ACM 2011 Conference on Computer Supported Cooperative Work, CSCW '11*, 25–34. New York, NY, USA: Association for Computing Machinery.
- Rizwan, M.; Wan, W.; Cervantes, O.; and Gwiazdzinski, L. 2018. Using Location-Based Social Media Data to Observe Check-In Behavior and Gender Difference: Bringing Weibo Data into Play. *ISPRS International Journal of Geo-Information*, 7(5): 196.
- Robinson, L.; Cotten, S. R.; Ono, H.; Quan-Haase, A.; Mesch, G.; Chen, W.; Schulz, J.; Hale, T. M.; and Stern, M. J. 2015. Digital inequalities and why they matter. *Information, Communication & Society*, 18(5): 569–582.
- Schneider, D.; and Harknett, K. 2022. What's to Like? Facebook as a Tool for Survey Data Collection. *Sociological Methods & Research*, 51(1): 108–140.
- Wang, Y. 2021. Closing the gender gap in college attendance: Variation by family background in China over time. *Social Science Research*, 98: 102578.
- Weber, I.; Kashyap, R.; and Zagheni, E. 2018. Using advertising audience estimates to improve global development statistics. *ITU Journal: ICT Discoveries*, 1(2).
- Weibo Corporation. 2023. Weibo Reports Fourth Quarter and Fiscal Year 2022 Unaudited Financial Results.
- Yang, W.; Wu, Z.; Mok, N. Y.; and Ma, X. 2022. How to Save Lives with Microblogs? Lessons From the Usage of

Weibo for Requests for Medical Assistance During COVID-19. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, CHI '22. New York, NY, USA: Association for Computing Machinery.

Yeung, W.-J. J. 2013. Higher education expansion and social stratification in China. *Chinese Sociological Review*, 45(4): 54–80.

Yuan, Y.; Wei, G.; and Lu, Y. 2018. Evaluating gender representativeness of location-based social media: a case study of Weibo. *Annals of GIS*, 24(3): 163–176.

Zhang, L.; and Pentina, I. 2012. Motivations and Usage Patterns of Weibo. *Cyberpsychology, Behavior, and Social Networking*, 15(6): 312–317.