

Shulin Hu

March 26th, 2020

Weibo – How is China's second largest social media platform being used for social research?

3 comments | 76 shares

Estimated reading time: 7 minutes

*Social media research in the global north is primarily focused on western social media platforms, notably Twitter and Facebook. In this post **Shulin Hu** describes how researchers have used data from Weibo, China's second most popular social media platform, to undertake a variety of research projects and provides resources for researchers looking to use Weibo as a tool for social research.*

Zoufan posted her last words on Weibo on 18, March, 2012. She was suffering from a major depressive disorder, and shortly after – committed suicide.

Weibo is a microblogging application, launched by Sina Corporation in 2009, based on user relationships to share, disseminate and get information. As of Q4 2019, it has over 516 million monthly active users (compared to Twitter's 300 million), making it only the second largest social media platform in China after WeChat.

A sign of the social media platforms importance, in April 2019, **the National Library of China announced that it will archive all public posts on Weibo for non-commercial uses**. This move will have a profound influence on digital-heritage retention and it is estimated that more than 200 billion textual posts and 50 billion pictures will be stored, as well as data relating to the sentiments underlying these posts.

Since Zoufan's last post, other Weibo users found her account and began to share their stories of depression as comments. There are now more than 1 million comments. This caught the attention of Tingshao Zhu and his colleagues from the Chinese Institute of Psychology, who started **investigating** this case and devising a strategy for how they could use the Weibo data to connect with patients and prevent other suicides. In a similar study about emotional states on Weibo, Sinan Yang & Jian Xu used **BDP** (A free online tool to help users deal with databases) to visualise the flow in polarity over time and quickly detect the location of peaks and wanes in the general population's emotional state.



As of Q4 2019, it has over 516 million monthly active users (compared to Twitter's 300 million), making it only the second largest social media platform in China



Another tool for data visualisation, **WeiboEvents**, allows users to study the spread of information and how accounts gain their popularity on Weibo. Feng Xian used it to study **@Yutu**, the Weibo account for the rover that reached the Moon in December 2013. Yutu literally means “jade rabbit”, which refers to the pet rabbit of the Moon goddess in Chinese mythology. On Weibo, **@Yutu** has over 730,000 followers and posts updates and news of its discoveries, as well as cute cartoons about its history and generally about the universe, explaining complex concepts in a visual way.

In February 2014, it briefly went quiet during the lunar night, but after recovering from some mechanical difficulties (which were actually happening to the real rover on the moon), it posted the message: "Hi, anybody there?", "I'm the rabbit that has seen the most stars!" This post attracted more than 840,00 reviews and 151,000 likes and also a fascinating dataset.



Feng Xian collated the data from the reviews and reposts and extracted the characters relating to emotional expressions as well as the emojis. He found that 60% of the users post compliments about its joyful 'personality' and 19% of users were encouraging the rabbit/rover to keep going (as if it were a real person) when the rover itself was facing technical problems on the Moon.

The reposting level also indicates a high penetration of Weibo content, from its targeted audience to potential audiences. The researcher classified the reposting on the microblog to six layers: the direct reposting number is 2231 (40%), the secondary reposting number is 1780 (32%), and the next four are 735 (13%), 231 (4%), 111 (2%), 490 (9%), indicating that after original reposts by some users, their friends will keep forwarding it based on social relationship circle.

The researchers also investigated the interaction model between the rover account and social media users, to find out how to balance the personified mood with scientific knowledge about the exploration of the universe. Specifically, instead of exhibiting the attitude of imparting professional knowledge as an emotionless machine, this account

established an equal relationship with the audience during the virtual interaction process, which helped mobilize their enthusiasm to participate.

In a **different study**, a group of researchers from Hong Kong used Sina Weibo data to analyse misinformation. They collected both Twitter and Weibo data to understand the levels and spread of Ebola misinformation in 2013-2014. The researchers had to write a script to scrape Weibo data, as an API was not available at the time. They found that only 2% of their sample contained misinformed treatment options, compared to perhaps 50%+ reported in other studies looking at the misinformation spread of Ebola treatments in Guinea, Liberia, and Nigeria during the same year. In the current context of Covid-19, lifelogs present an enormous opportunity to develop an online history, which can then be used in similar ways to understand the spread of virus and its social effects.



Weibo has a powerful ability, due to its large sample size, to study and track sentiment, affective states, online behaviours, and communications within the Chinese socio-cultural context



In 2014 Weibo released an official and free **API** for its raw data, a practical (and English) step-by-step guide to using the API is available **here**. One **application** of the API, was undertaken by a team of researchers from Wuhan University, who used the API to extract information about hot spots and movements across the city to help with urban planning. Although the API made their data collection easier, they were faced with the challenge of having to request permission from users for this type of data.

There are many more studies and research teams using Sina Weibo data, analysing behaviours, trends and the spread of information through the network. Another

difference from studies of Twitter and Facebook is the emoji feature on Weibo and the ease of using this response, along with 'likes' to understand people's tendencies and trends in expressing emotions in response to events or posts.

Ultimately, Tingshao Zhu and his colleagues from the Chinese Institute of Psychology wanted to prevent suicides, currently more than 300k Chinese ending their lives every year. So they built an algorithm, trained with manually tagged data from the responses to the Zoufan posts, to recognize people at high risk of suicide among numerable updates on Weibo and classify the severity automatically. His team aims to use this algorithm, combined with their training in psychology to identify people at high risk of suicide and reach out to provide the support they need.

It should be noted that using this kind of highly sensitive data runs the very real risk of invading peoples privacy. In 2014, a British project used a similar tool, *Samaritans radar*, to guard against potential mental illness. If it recognised negative content published by a user, then it informed the user's friends in Twitter automatically. This application was strongly opposed and ultimately only ran for ten days. In the Weibo case, it was decided to limit the level of contact between the project's AI and potential sufferers. To date they have found 4222 users with depressive disorders and provided further advice for them.

The above cases make clear that compared to other social media platforms, Weibo has a powerful ability, due to its large sample size, to study and track sentiment, affective states, online behaviours, and communications within the Chinese socio-cultural context. However, in so doing, it also highlights risks and controversies surrounding online privacy and safety, which surround such research. As social media research on the social network gathers pace, how these issues will be addressed and resolved remains an open question.

Data collection tools/applications for Weibo:

Tool	Developer	Functions	Access from:
Weibo Events	PKUVIS	Data visualisation; Social network Analysis	http://vis.pku.edu.cn/weibova/weiboevents/index.htm

Tool	Developer	Functions	Access from:
WEIBOREACH	Zhiweidata. Co. Ltd	Content Analysis; Thematic Analysis; User Profile	http://www.weiboreach.com/index-main.html
BDP	Haizhi Network Technology Co. Ltd.	Coding dashboard; Data Visualisation	https://me.bdp.cn/home.html
GOOSEEKER	Tianju Co.Ltd.	Content Analysis; Semantic analysis	http://www.gooseeker.com/land/weibo.html
WEIBOSTATS	KAWO Co. Ltd.	Analytics dashboard for industry intelligence	https://weibostats.com/

Datasets of Weibo (mainly for academic use):

<https://github.com/Lab41/sunny-side-up/wiki/Chinese-Datasets> (Chinese language corpora for sentiment analysis, including 226 million posts on weibo)

http://networkrepository.com/soc_sinaweibo.php (Characterizing Tweeting behaviors of weibo users via public data streaming)

<http://coai.cs.tsinghua.edu.cn/hml/dataset/> (About emotional conversation)

<https://archive.ics.uci.edu/ml/datasets/microblogPCU> (Can be used to study machine learning methods)

<https://aminer.org/data-sna#Weibo-Net-Tweet> (Helpful for analysing weibo users' relationships, including 300,000 original microblogs and 23 million retweets)

<https://hub.hku.hk/cris/dataset/dataset107483> (Sampling timelines of more than 350,000 Chinese microbloggers who have more than 1k followers)

A similar version of this post appeared on the SAGE Ocean Blog as, [How researchers around the world are making use of Weibo data.](#)

Note: This article gives the views of the author, and not the position of the LSE Impact Blog, nor of the London School of Economics. Please review our [comments policy](#) if you have any concerns on posting a comment below

Image Credit, [Rabbits watching the moon](#) via Public Domain Vectors.



About the author



Shulin Hu

Shulin Hu is a MSc student from the Digital Humanities programme at University College London, where she worked with the SAGE Ocean team. She graduated from Communication University of China, majoring in journalism and communication. She has also worked as an intern with Tencent's Marketing and PR team, Caixin Media, Xinhua News Agency, China Review News Agency of Hong Kong, and the CSR department of BMW Brilliance Automotive Ltd.

Posted In: Social media

3 Comments

Related Posts

Big data

Using LinkedIn for Social Research

JULY 9TH, 2019

2 { ... }

Digital scholarship

Using Twitter as a data source: an overview of social media research tools (2019)

JUNE 18TH, 2019

24 { ... }

Academic communication

"Why I've deleted my Twitter account #exhaustionrebellion" ~ by Mark Carrigan

DECEMBER 6TH, 2019

4 { ... }

Academic communication

#ScientistsWhoSelfie – How sharing selfies can build trust in science

JULY 3RD, 2019



Visit our sister blog [LSE Review of Books](#)



© LSE 2022

We use cookies on this site to understand how you use our content, and to give you the best browsing experience. To accept cookies, click continue. To find out more about cookies and change your preferences, visit our [Cookie Policy](#).

Continue

u