

## Social Network Users' Self-Disclosure Intention: A Case Study of Sina Weibo (Postprint)

**Authors:** Zang Guoquan, Kong Xiaohuan, Zhang Kailiang, Yu Zhengjie

**Date:** 2023-04-01T16:02:54+00:00

### Abstract

[Purpose/Significance] Social network users' self-disclosure holds strategic significance for social networks built upon user-generated content as their business foundation, as the quantity and quality of such content depend on users' willingness to self-disclose. Therefore, investigating the factors influencing social network users' self-disclosure willingness provides valuable references for platforms to formulate privacy policies and enhance self-disclosure levels, thereby fostering the healthy and rapid development of social network platforms.

[Method/Process] By referencing established research frameworks, we construct a research model for social network users' self-disclosure willingness. Sina Weibo is selected as a representative social network platform, and Python-based web scraping techniques are employed to collect user Weibo data for analyzing self-disclosure willingness.

[Results/Conclusion] The semantic content of Weibo posts, location tags, and data permissions all significantly influence users' self-disclosure willingness; specifically, operations such as hiding location tags and configuring data permissions substantially enhance this willingness. Social network users' self-disclosure willingness constitutes a personal trait affected by demographic factors including gender, age, and educational attainment.

### Full Text

#### Research on Social Network Users' Self-Disclosure Willingness—A Case Study of Sina Weibo

Zang Guoquan<sup>1,2</sup>, Kong Xiaohuan<sup>1</sup>, Zhang Kailiang<sup>3</sup>, Yu Zhengjie<sup>1</sup>

<sup>1</sup>School of Information Management, Zhengzhou University, Zhengzhou 450001

<sup>2</sup>Zhengzhou Data Science Research Center, Zhengzhou 450001 <sup>3</sup>School of Politics and Public Administration, Zhengzhou University, Zhengzhou 450001

## Abstract

**[Purpose/Significance]** Users' self-disclosure is of strategic significance to social network platforms that rely on user-generated content, as both the quantity and quality of such content depend on users' willingness to self-disclose. Therefore, studying the factors influencing social network users' self-disclosure willingness can provide references for platforms to formulate privacy policies and enhance disclosure levels, thereby promoting healthy and rapid platform development. **[Method/Process]** Drawing upon existing research frameworks, this study constructs a research model for social network users' self-disclosure willingness. Using Sina Weibo as a representative platform, Python crawler technology was employed to obtain user microblog data for analysis. **[Result/Conclusion]** The semantic content, location tags, and data permissions of microblog posts all affect users' self-disclosure willingness. Operations such as hiding location tags and setting data permissions significantly enhance users' self-disclosure willingness. Social network users' self-disclosure willingness is a personal trait influenced by demographic factors including gender, age, and education level.

**Keywords:** social network; self-disclosure; disclosure willingness; Sina Weibo

---

## 2 Literature Review

Self-disclosure refers to individuals voluntarily and truthfully sharing their thoughts and behaviors with others during social interactions, forming the foundation of human communication and social relationships. Social networks facilitate relationship building and maintenance, naturally encouraging user self-disclosure. Platforms such as WeChat, Weibo, Douyin, and Kuaihou have become deeply integrated into daily life. However, frequent privacy breaches (e.g., the leak of 159 million LinkedIn users' sensitive data and the Cambridge Analytica scandal involving 87 million Facebook users) have raised widespread concerns, making sustained user self-disclosure a primary challenge for platforms. Users are fundamental to social networks, whose growth depends on the quantity and quality of self-disclosed content. Historical self-disclosure data represents authentic behavioral records. This study uses Sina Weibo as a case study, employing crawler technology to obtain historical microblog data and examining user self-disclosure willingness and its influencing factors from three dimensions: semantic content, location tags, and data permissions. Understanding these factors helps platforms develop privacy policies and enhance disclosure levels, while also guiding users (especially adolescents) in proper privacy settings and appropriate self-disclosure.

Existing empirical research on self-disclosure is summarized in Table 1 from three perspectives: data acquisition methods, key influencing factors, and research contexts. Current studies primarily attribute influencing factors to "perceived risk and benefit (privacy calculus theory)," "attitude, subjective norms, and perceived control (theory of planned behavior)," and "privacy rules, trust,

and perceived control (communication privacy management theory).” Most research combines privacy calculus theory with other frameworks, using surveys to measure subjective perceptions. However, surveys have limitations: fixed questions lack flexibility, may introduce bias, and self-reported data may not reflect actual attitudes or behaviors. Some studies combine experiments with surveys, but these remain subjective and costly.

To overcome these limitations, this study employs crawler technology to extract historical self-disclosure data, offering two advantages: (1) platforms retain vast amounts of authentic behavioral records for deep mining, and (2) crawled data is large-scale and reliable, overcoming subjectivity and more objectively reflecting behavioral characteristics. Based on A. Adams’ privacy perception model (information sensitivity, information reception, and information use), this study examines self-disclosure willingness from semantic, spatial (type and level), and data permission dimensions.

---

### 3 Research Model for User Self-Disclosure Willingness

A. Adams et al. used grounded theory to develop a privacy perception model from three dimensions—information sensitivity, information reception, and information use—which has become an effective framework for privacy research. J.S. Cunningham et al. applied this model to digital photo sharing, while C. Jennett et al. studied credit loan applications. In social networks, self-disclosure involves posting information, but information users are unpredictable. Therefore, this study adopts only the “information sensitivity” and “information reception” dimensions.

#### 3.1 Semantic Dimension

The semantic dimension categorizes self-disclosure text based on content features. Existing research shows information type significantly influences forwarding willingness—users prefer hedonic over utilitarian information, and non-emotional over emotional content in hot topics. Drawing on Hou Min’s classification and user posting purposes, this study categorizes self-disclosure content into: life records, news/entertainment, work/study, and emotional expression.

#### 3.2 Spatial Dimension

The spatial dimension analyzes location tag features across spatial level and type.

*Spatial Level Sub-dimension:* Platforms allow custom location precision from city to specific venues. Users can protect privacy through personalized settings. This study defines three levels: city, street, and specific location.

*Spatial Type Sub-dimension:* When spatial level is “specific location,” type becomes relevant. Research shows sensitivity varies by location type—public entertainment venues are least sensitive, health-related locations more sensitive,

and residences most sensitive. Given students comprise 21% of Chinese internet users, this study defines four types: public entertainment venues, medical/health facilities, education/learning institutions, and personal residences.

### 3.3 Data Permission Dimension

Data permission refers to user control over data recipient scope. Studies show that granting users choice over recipients reduces privacy concerns and increases disclosure willingness. Most platforms offer permission settings (e.g., Weibo: public, followers, friends circle, private; Facebook: only me, friends, friends of friends, public). For simplicity, this study uses two levels: “visible to all” and “visible to some.”

Based on these dimensions, the research model is constructed as shown in Figure 1 [Figure 1: see original paper].

---

## 4 Data Acquisition and Processing

According to Sina’s official statistics, as of September 2020, Weibo had 511 million monthly active users and 224 million daily active users, making it a crucial social media platform in China.

### 4.1 Acquisition Method and Data Items

User data can be obtained via Weibo API (requiring authorization with request limits) or web crawlers. This study adopts web crawler technology. Required data includes: (1) basic user data (gender, birth date, education, registration time) and (2) historical microblog data (post text, location tags, total posts).

### 4.2 Sample Selection

Samples were screened for: (1) **Account Activity**: registered \$ \$1 year with \$ \$100 posts; (2) **Account Operator**: individual users (not official, celebrity, or professional accounts), excluding those with >5,000 followers; (3) **Representativeness**: demographics should align with the 2018 Weibo User Development Report.

### 4.3 Data Processing

#### 4.3.1 Data Cleaning

Using Python’s Pandas library: (1) **Basic data cleaning**: removed incomplete, logically erroneous, and duplicate data. Logical errors included inconsistent age/registration time or age/education combinations. (2) **Text cleaning**: removed non-original posts, “@username” mentions, and URLs.

#### 4.3.2 Data Classification

(1) **Semantic Classification**: Using Jieba for word segmentation and stop-word removal, TF-IDF for feature weighting, and LDA for topic modeling to

cluster posts into four semantic categories (life records, news/entertainment, work/study, emotional expression). Table 2 shows sample LDA results.

**(2) Spatial Classification:** - *Spatial Level:* Based on 2020 Administrative Division Codes, location tags were coded and classified as “no tag,” “city” (ending with “00”), “street” (ending with road/street/district/alley characters), or “specific location.” - *Spatial Type:* For “specific location” tags, fuzzy matching categorized venues into public entertainment, medical/health, education/learning, or personal residence.

---

## 5 Statistical Analysis and Results

### 5.1 Data Statistics

From December 10, 2020, to January 31, 2021, a Python crawler collected 61,902 user IDs using snowball sampling from follower lists. After screening and cleaning, 2,047 valid users with ~157,000 posts remained. Demographics are shown in Table 3 .

### 5.2 Data Analysis

Self-disclosure willingness was measured as the ratio of publicly posted microblogs to total posts. SPSS 22.0 analyzed demographic, semantic, spatial, and permission effects.

#### 5.2.1 Demographic Factors

Multiple linear regression results (Table 4 ) show:

- **Gender:** Female users’ willingness is significantly lower than males ( $\beta = -0.040$ ,  $P = 0.004$ ). Males may use social networks more to establish new relationships, particularly with strangers.
- **Age:** Significant negative effect ( $\beta = -0.014$ ,  $P = 0.035$ ). Older users have higher privacy concerns, while younger users view social networks as platforms for identity expression.
- **Education:** Significant negative effect ( $\beta = -0.026$ ,  $P = 0.005$ ). Higher-educated users exhibit greater privacy concern, reducing disclosure.

Thus, self-disclosure willingness is a personal trait shaped by demographics.

#### 5.2.2 Semantic Dimension

Public posts by semantic type averaged: life records (163), news/entertainment (212), work/study (148), and emotional expression (245). Paired t-tests (Table 5 ) show news/entertainment and emotional expression posts significantly exceed life records and work/study posts ( $P < 0.001$ ), with no significant difference within these pairs. This aligns with prior research—users prefer sharing hedonic information with lower privacy risks.

#### 5.2.3 Spatial Dimension

(1) **Location Tags:** Users posted significantly fewer location-tagged posts (355) than non-tagged posts (413,  $P = 0.026$ ), indicating preference for hiding location. Location data can reveal sensitive information and enable user profiling, reducing disclosure willingness.

(2) **Spatial Level:** Average public posts by level: city (153), street (111), specific location (91). Paired t-tests (Table 6) show significant differences (all  $P < 0.001$ )—lower spatial precision correlates with higher disclosure willingness.

(3) **Spatial Type:** For specific locations, average posts were: public entertainment (32), medical/health (15), education (25), personal residence (19). Paired t-tests (Table 7) show significantly higher willingness at public entertainment and education venues than medical/health facilities and residences ( $P < 0.001$ ), with no difference within these pairs. Public venues show higher willingness than private ones, with health locations raising health privacy concerns.

#### 5.2.4 Data Permission Dimension

Average public posts (“visible to all”): 768; restricted posts (“visible to some”): 945. Paired t-tests show significantly more restricted posts ( $P = 0.005$ ), confirming that permission control enhances disclosure willingness by reducing privacy concerns.

---

## Conclusion

This study reveals: (1) Self-disclosure willingness is a demographic-influenced personal trait; (2) Semantic content, location tags, and data permissions significantly affect willingness; (3) Hiding location tags and setting permissions enhance willingness.

**Practical Applications:** Platforms can improve user control by: (1) Setting default location tags to “none” or city-level, and default permissions to “visible to some”; (2) Empowering users with granular control options to reduce privacy concerns and sustain disclosure.

**Limitations:** (1) Low valid user rate (3.3%) due to incomplete profile data; (2) Analysis limited to posted content, not profile data disclosure; (3) Using public post ratio as willingness proxy may need validation. Future research should incorporate multiple indicators and examine profile data disclosure.

---

## References

- [1] JOURARD SM, LASAKOW P. Some factors in self-disclosure[J]. Journal of abnormal & social psychology, 1958, 56(1): 91-98. [2] Sohu.com. LinkedIn hacked, data of ~159 million users stolen[EB/OL]. [2021-01-14]. [https://m.sohu.com/a/292645813\\_{557054}](https://m.sohu.com/a/292645813_{557054}). [3] Baidu Baike. Cambridge

Analytica scandal[EB/OL]. [2021-01-14]. <https://baike.baidu.com/item/剑桥分析丑闻>. [4] KRASNOVA H, KOLESNIKOVA E, GÜNTHER O. Leveraging trust and privacy concerns in online social networks: an empirical study[C]//18th European conference on information systems. Illinois: AIS Library, 2010. [5] GUO Hailing, MA Hongyu, XU Zehui. Construction and empirical study of social media user information disclosure willingness model—taking WeChat users as an example[J]. Library and Information Service, 2019, 63(15): 111-120. [6] DINEV T, HART P. Internet privacy concerns and their antecedents—measurement validity and a regression model[J]. Behaviour & information technology, 2004, 23(6): 413-422. [7] ZHU Hou, WANG Ke, YAN Zhijun, et al. Research on SNS user privacy paradox based on privacy calculus theory[J]. Journal of Intelligence, 2017, 36(2): 134-139. [8] ZHANG Xing, CHEN Xing, HOU Delin. Study on influencing factors of online health information disclosure willingness: an integrated model of planned behavior theory and privacy calculus[J]. Information and Documentation Services, 2016(1): 48-53. [9] WANG Yuchao. Research on influencing factors of health privacy information disclosure willingness in online medical communities[J]. Journal of Information Resources Management, 2018, 8(1): 93-103. [10] XU H, TEOH H, TAN BCY, et al. The role of push-pull technology in privacy calculus: the case of location-based services[J]. Journal of management information systems, 2010, 26(3): 135-174. [11] LI Zhengren. Research on influencing factors of user privacy concern and privacy information diffusion patterns in mobile internet environment[D]. Beijing: Beijing University of Posts and Telecommunications, 2014. [12] LIU Bailing, XIA Huimin, LI Yanhui. Empirical study on influencing factors of mobile shopping user information disclosure willingness—based on equity theory and rational behavior theory[J]. Information Theory and Practice, 2017, 40(5): 87-93. [13] LI Yuguang. Research on influencing factors of patient medical information sharing willingness in online medical communities[D]. Harbin: Harbin Institute of Technology, 2015. [14] ZLATOLAS N, WELZER T, HÖLBL M, et al. A model of perception of privacy, trust, and self-disclosure on online social networks[J]. Entropy, 2019, 21(8): 772-789. [15] KO H, CHEN T. Understanding the continuous self-disclosure of bloggers from the cost-benefit perspective[C]//Proceedings of the 2nd conference on human system interactions. New York: IEEE Press, 2009. [16] LOIACONO ET. Self-disclosure behavior on social networking websites[J]. International journal of electronic commerce, 2015, 19(2): 66-94. [17] JIANG Youqi. Research on influencing factors of personal health information disclosure willingness in online medical websites[D]. Wuhan: Wuhan University, 2017. [18] ZIMMER JC, ARSAL R, AL-MARZOUQ M, et al. Knowing your customers: using a reciprocal relationship to enhance voluntary information disclosure[J]. Decision support systems, 2010, 48(2): 395-406. [19] PREMAZZI K, CASTALDO S, GROSSO M, et al. Customer information sharing with e-vendors: the roles of incentives and trust[J]. International journal of electronic commerce, 2014, 14(3): 63-91. [20] STERN T, SALB D. Examining online social network use and its effect on the use of privacy settings and profile disclosure[J]. Bulletin of science technology & society, 2015, 35(1): 25-34. [21] KIM E,

LEE J, SUNG Y, et al. Predicting selfie-posting behavior on social networking sites: an extension of theory of planned behavior[J]. *Computers in human behavior*, 2016, 62(9): 116-123. [22] BANSAL G, ZAHEDI FM, GEFEN D. The impact of personal dispositions on information sensitivity, privacy concern and trust in disclosing health information online[J]. *Decision support systems*, 2010, 49(2): 138-150. [23] ADAMS A, SASSE MA. *Privacy in multimedia communications: protecting users, not just data*[M]. London: Springer, 2001. [24] CUNNINGHAM JS, MASOODIAN M, ADAMS A. Privacy issues for online personal photograph collections[J]. *Journal of theoretical & applied electronic commerce research*, 2010, 5(2): 26-40. [25] JENNETT C, BROSTOFF S, MALHEIROS M, et al. Adding insult to injury: consumer experiences of being denied credit[J]. *International journal of consumer studies*, 2012, 36(5): 549-555. [26] SHI Wenhua, CHEN Nan, ZHANG Qi. Research on the influence of relationship type and information type on social network users' forwarding willingness[J]. *Institutional Economics Research*, 2016(4): 162-180. [27] WAN Xiaoyu, YANG Teng, YUAN Ye, et al. Research on the influence of Weibo information emotion types on users' sharing willingness[J]. *Information Science*, 2019, 37(8): 97-107. [28] HOU Min. Research on user information sharing behavior on social network platforms—taking Shanghai college students' Weibo usage as an example[J]. *China Newspaper Industry*, 2020(10): 16-17. [29] JIN L, LONG X, JOSHI JBD. Towards understanding residential privacy by analyzing users' activities in foursquare[C]//*Proceedings of the 2012 ACM workshop on building analysis datasets and gathering experience returns for security*. New York: ACM Press, 2012. [30] YANG Shihong, CHEN Tangfa, ZHANG Rui. User location information sensitivity cognition under privacy risk estimation framework[J]. *Journal of Lanzhou University (Social Sciences)*, 2019, 47(4): 8-16. [31] China Internet Network Information Center. 47th Statistical Report on China's Internet Development[R/OL]. [2021-03-03]. [http://www.cac.gov.cn/2021-02/03/c\\_1613923423079314.htm](http://www.cac.gov.cn/2021-02/03/c_1613923423079314.htm). [32] TSAI JY, KELLEY P, DRIELSMAP, et al. Who's viewed you? The impact of feedback in a mobile location-sharing application[C]//*Proceedings of the SIGCHI conference on human factors in computing systems*. New York: ACM Press, 2009. [33] SADEH N, HONG J, CRANOR L, et al. Understanding and capturing people's privacy policies in a mobile social networking application[J]. *Personal and ubiquitous computing*, 2009, 13(6): 401-412. [34] Sina Finance. Weibo releases Q3 2020 financial report[R/OL]. [2021-03-03]. <https://finance.sina.com.cn/stock/usstock/c/2020-12-28/doc-iiznnezx9422515.shtml>. [35] Sina Weibo Data Center. 2018 Weibo User Development Report[EB/OL]. [2021-03-03]. <https://data.weibo.com/report/reportDetail?id=49935337?fr=aladdin>. [36] Ministry of Civil Affairs of China. 2020 Administrative Division Codes of China[EB/OL]. [2021-03-03]. <http://www.mca.gov.cn/article/sj/xzqh/2020/>. [37] STRANO MM. User descriptions and interpretations of self-presentation through Facebook profile images[J]. *Journal of psychosocial research on cyberspace*, 2008, 2(2): 1-10. [38] ZANG Guoquan, ZHOU Xiaoqian. Analysis of users' motivation to continuously use social networks under privacy concern distress[J]. *Information Studies: Theory & Application*, 2020, 43(3): 71-77. [39]

YANG Yang, WEN Xiuyan. The mediating effect of privacy protection willingness: privacy concern, privacy protection self-efficacy and precise advertising avoidance[J]. Press Circles, 2020(7): 41-52. [40] LU Jiayin, BAI Jie. Research on Chinese youth's online privacy concerns and influencing factors—based on a survey of 1,599 Communist Youth League members[J]. Shanghai Journalism Review, 2021(2): 69-79. [41] WU Biaobing, XU Helong, ZHANG Yu. Empirical research on Chinese public's privacy sensitivity[J]. Journal of Nanjing University of Posts and Telecommunications (Social Science Edition), 2015, 17(3): 82-90. [42] PREOTIUC-PIETRO D, COHN T. Mining user behaviors: a study of check-in patterns in location-based social networks[C]//Proceedings of the 5th annual ACM web science conference. New York: ACM press, 2013. [43] FIRE M, KAGAN D, ELISHAR A, et al. Social privacy protector—protecting users' privacy in social networks[C]//The second international conference on social eco-informatics. Venice: IARIA XPSpress, 2012.

---

### Author Contributions

Zang Guoquan: Conceptualization, methodology, final manuscript revision  
Kong Xiaohuan: Visualization, initial manuscript revision Zhang Kailiang:  
Data analysis, original draft writing Yu Zhengjie: Data acquisition, data  
processing

*Note: Figure translations are in progress. See original paper for figures.*

*Source: ChinaXiv — Machine translation. Verify with original.*