Check for updates

Article

new media & society
2017, Vol. 19(9) 1476–1497
© The Author(s) 2016
Reprints and permissions:
sagepub.co.uk/journalsPermissions.nav
DOI: 10.1177/1461444816642210
journals.sagepub.com/home/nms
SAGE

Privacy protection and self-disclosure across societies: A study of global Twitter users

Hai Liang
The University of Hong Kong, Hong Kong

Fei Shen
City University of Hong Kong, Hong Kong

King-wa Fu
The University of Hong Kong, Hong Kong

Abstract

Privacy is a culturally specific phenomenon. As social media platforms are going global, questions concerning privacy practices in a cross-cultural context become increasingly important. The purpose of this study is to examine cultural variations of privacy settings and self-disclosure of geolocation on Twitter. We randomly selected 3.3 million Twitter accounts from more than 100 societies. Results revealed considerable cultural and societal differences. Privacy setting in collectivistic societies was more effective in encouraging self-disclosure; whereas it appeared to be less important for users in individualistic societies. Internet penetration was also a significant factor in predicting both the adoption of privacy setting and geolocation self-disclosure. However, we did not find any direct relationships between cultural values and self-disclosure.

Keywords

Boundary regulation, cross-cultural comparison, geolocation, privacy protection, self-disclosure, Twitter

Corresponding author:
Hai Liang, Journalism and Media Studies Centre, The University of Hong Kong, Room 114, Eliot Hall, Pokfulam Road, Hong Kong.
Email: rainfireliang@gmail.com

With the diffusion of Internet technologies, online privacy becomes an eminent issue that faces all Internet users. Unintentional personal information leakage could lead to a series of negative consequences such as credit damage, unsolicited emails or phone calls, or even financial loss. Many social media users have expressed serious concerns about the leakage of personal information online (Young and Quan-Haase, 2013). However, only 30% of all US adults have taken simple steps to protect their privacy online, such as changing their privacy settings on social media (Rainie and Madden, 2015). There is a serious disparity between people's reported high levels of privacy concerns and their protection practices (Barnes, 2006; Stutzman and Kramer-Duffield, 2010).

To a certain extent, such a disparity exists because of the way the Internet is used today; that is, the cyberspace is dominated by social media platforms. Privacy practices on social media platforms are often paradoxical. On the one hand, Internet users are often motivated to disclose personal information to present a unique identity that differentiates themselves from others (boyd, 2008) and to accumulate social capital in online social networks (Ellison et al., 2011). On the other hand, social media companies retain a large amount of personal information collected from its users, and the information can easily be abused (boyd, 2006; Papacharissi and Gibson, 2011).

To help with this dilemma, almost all popular social media platforms allow users to customize their privacy settings. Users can create deterministic rules specifying which part of the content will be shared, and to whom the content will be accessible. When being in control of their privacy, individuals tend to disclose more information (Stutzman et al., 2011). Studies have been conducted to understand privacy protection behavior on social media platforms (e.g. boyd and Marwick, 2011; Madden et al., 2013; Stutzman et al., 2011; Stutzman and Kramer-Duffield, 2010), but since most of the previous studies focused their attention on individual-level factors, it remains unclear whether privacy protection behaviors and self-disclosure vary across societies of different cultures. Researchers have shown that levels of online privacy concern are conditional and multicultural—the expression of privacy varies significantly across cultures (Cho et al., 2009). As social media platforms are going global, questions concerning macro societal-level factors seem to be increasingly important.

This study aims to answer two major questions. First, by conceiving privacy as boundary regulation (Altman, 1975; Petronio, 1991, 2002), we seek to examine the relationship between privacy protection and self-disclosure at the individual level. Second, we further explore the role of macro-societal factors in determining privacy protection and self-disclosure. To answer our research questions, we collected data from Twitter globally. We decided to use Twitter as our data source for four reasons. First, Twitter is one of the most popular social media platforms worldwide. Second, Twitter is public to anyone by default, but users can adjust privacy settings to make their accounts protected if they wish to restrict access to only approved followers. Finally, Twitter is more generous than other social media platforms in terms of the data availability via Application Programming Interfaces (APIs).

Privacy setting as boundary regulation

In highly contextual situations, privacy could be considered as a dynamic process in which individuals selectively control access to personal information (Altman, 1975).

Privacy is a boundary regulation process, whereby people close or open themselves to others in accordance with the need for disclosure and the need for privacy (Petronio, 1991, 2002). Boundary regulation is a central metaphor in communication privacy management (CPM) theory (Petronio, 1991, 2002). According to CPM theory, privacy boundary draws the line between private information and public information. Individuals create and apply rules to manage if and how information will be shared or concealed.

Boundary regulation theory has been applied to interpreting privacy practices on social media (Choi and Bazarova, 2015; Stutzman and Hartzog, 2012). Content-sharing behavior is potentially in conflict with the need to reduce privacy risk in the cyberspace. For social media platforms like Twitter, this conflict could be even more intense, because the primary purpose of Twitter usage is to share information. Without the flow of information among individuals, a social network becomes a static and a social environment (Papacharissi and Gibson, 2011). However, the public nature of Twitter might pave the way for privacy problems, such as the disclosure of offline activities (see Humphreys et al., 2014) and the locations where these activities take place (Elwood and Leszczynski, 2011).

There are many variations in the way people manage their private information on social media platforms (Child and Petronio, 2011). The privacy setting of Twitter is binary in nature. The default is that all tweets are public; that is, anyone on the Internet can access the content. In addition, all public tweets are searchable on Twitter or major search engines such as Google (Casey, 2010). However, users have the option to allow designated people to access their tweets. If a user wants protected tweets, he or she must manually approve who may see their tweets.

Choi and Bazarova (2015) conceptualized privacy setting on Twitter as a form of privacy boundary management. Protected Twitter accounts tend to disclose more personal and intimate information, because this boundary is a form of audience representations. Audience representations are the heuristic cues embedded in social media platforms that suggest the potential audience of a Tweeter, and they influence users' perceptions of audience as being bounded versus unbounded. Loosely defined boundaries (e.g. public Twitter account) imply less control over information and its disclosure to less familiar and unbounded audience.

Many types of personal information have been studied on social media, such as personally identifiable information, offline activities (Humphreys et al., 2014), photo tagging (Rui and Stefanone, 2013), and geolocation (Friedland and Sommer, 2010). Among these, geolocation poses a newer and possibly more serious privacy threat, for example, facilitating identification and disclosure (Elwood and Leszczynski, 2011) and mounting privacy attacks (Friedland and Sommer, 2010). Twitter's geo-tagging service allows its users to attach their location information to tweets. This feature is turned off by default, and users need to opt in to use it. Once a user has enabled the location service, a location will be attached automatically to their tweets. When using Twitter on mobile devices, tweets will contain the precise location information including latitude and longitude statistics. Humphreys et al. (2014) found that Twitter users are more likely to include location information when tweeting about offline activities.

A large number studies have argued the importance of understanding privacy management on social media (boyd and Hargittai, 2010; Stutzman et al., 2011). Disclosure choices of social media users are influenced by specific structures within which people

negotiate their individual preferences (Choi and Bazarova, 2015). Altman (1975) and Petronio (2002) suggest that exerting control over disclosure through rule-making reinvigorates one's motivation to engage in disclosure. Stutzman et al. (2011) found that people who personalized their privacy settings tended to disclose more information on Facebook. In this sense, we expect that users with public accounts are less likely to add geo-tags to their tweets, because people are reluctant to disclose location-based activities to unfamiliar users:

H1. Users who have protected their accounts are more likely to add geolocation information in tweets than those who have public accounts.

Individual-level predictors of privacy setting and self-disclosure

Social media users differ in their privacy setting and self-disclosure behavior, which could be a function of individual-level factors. Previous studies hinted at a few possible antecedents including social media network size, user activity, and user experience.

Network size, such as number of followers on Twitter or Facebook, could affect self-disclosure and privacy setting. Both risks and benefits can rise as network size grows. According to boundary regulation theory, a larger network usually implies less familiar and unbounded audience and thus is associated with higher privacy risks (Rui and Stefanone, 2013). In this sense, users with larger networks are more likely to protect their accounts and disclose less private information. Meanwhile, social media users have a strong motivation to expand their social network for higher social capital (e.g. Choi and Bazarova, 2015) or larger probability of being retweeted (e.g. Bakshy et al., 2012; Suh et al., 2010). To maintain the relationships in a large network, users need to keep their accounts public and increase self-disclosure (see Rui and Stefanone, 2013).

Empirical studies found that larger network size is associated with more sophisticated privacy controls (Rui and Stefanone, 2013; Stutzman and Hartzog, 2012; Stutzman and Kramer-Duffield, 2010) and larger amounts of self-disclosure on Facebook (Rui and Stefanone, 2013; Young and Quan-Haase, 2013). The case of Twitter is more complicated. Choi and Bazarova (2015) found that network size is negatively associated with self-disclosure intimacy for protected users, whereas the relationship is positive for public users. They explained that people use self-disclosure to build social capital on Facebook and public Twitter, whereas they decrease self-disclosure for privacy protection on protected Twitter. Therefore, we hypothesize the following:

H2. The positive relationship between network size and self-disclosure is stronger for public Twitter users than for protected users.

The intensity of activity on social media, for example, the frequency of status updating, is expected to influence privacy-setting behavior. According to Lewis et al. (2008), there are three major reasons for active users to protect their account on Facebook. First, peer influence may be amplified by social media activity. The more frequently a user

browses online, the more likely the user may be affected by peers who adopted privacy protection. Second, active users may disclose more personal information by chance and thus are more inclined to protect their accounts. Finally, active users are more aware of the accessibility of others' personal information. Therefore, they would become more sensitive to the accessibility of their own account and upgrade privacy settings accordingly. Findings obtained from Facebook might be replicable on Twitter. Therefore, we hypothesize the following:

H3a. Active users are more likely to protect their profiles than inactive users.

H3b. Active users are more likely to disclose their geolocation in tweets than inactive users.

Familiarity with privacy practices grows with user experience. Experienced users tend to adopt more sophisticated privacy tools and disclosure strategies. Bellman et al. (2004) reported that Internet users' privacy concerns decrease with Internet experience. However, Cho et al. (2009) found that the length of Internet use is only positively associated with privacy concern but not related to privacy protection. Increased social media experience could also lead to heightened privacy concerns, because experienced users are more aware of how their data could be collected and used without permission (Singh and Hill, 2003). If this holds, privacy concerns will lead to active privacy protection and self-disclosure (Stutzman et al., 2011). Therefore, we hypothesize the following:

H4a. Users registered earlier are more likely to protect their accounts.

H4b. Users registered earlier are more likely to disclose geolocation in tweets.

The role of national culture

National culture is the collective mindset distinguishing members of one nation from another (Hofstede, 1980, 1991). Sense of privacy is culturally sensitive because patterns of interpersonal interaction vary from culture to culture (Altman, 1977). Culture is one of the five primary factors that influence the way people develop their own privacy rules (Petronio, 2002). In contrast, very few studies conducted systematical comparisons of privacy practices across cultures (see Cho et al., 2009).

Researchers in the past have looked into the relationship between Hofstede's cultural indicators and people's level of concerns for privacy (e.g. Bellman et al., 2004; Cho et al., 2009; Milberg et al., 2000). Cultural values have significant impacts on information technology use (Calhoun et al., 2002). Hofstede's four indices of national culture were frequently used: individualism (IND), power distance, uncertainty avoidance (UAI), and masculinity. The four indices were rarely included simultaneously in the same model, because these indices are highly correlated with one another (see Cho et al., 2009; Krasnova et al., 2012).

Culture might impact privacy protection and self-disclosure in different ways. First, cultural values, privacy protection, and self-disclosure are closely intertwined. Individuals

from individualistic cultures tend to place more value on private life, whereas collectivistic societies are more willing to accept organizational interference into the private life of an individual. Researchers found that people in highly individualistic societies exhibited higher levels of privacy concerns (Cho et al., 2009; Liu et al., 2004; Milberg et al., 2000) and are more likely to adopt proactive self-protections (Cho et al., 2009; Rui and Stefanone, 2013). UAI measures the extent to which a society feels uncomfortable with ambiguity and tries to avoid these situations. High levels of UAI are associated with anxiety, stress, and concerns for security. Therefore, privacy concerns are positively related to UAI (Bellman et al., 2004; Milberg et al., 2000) and are a major predictor of privacy protection online (e.g. Youn, 2009). However, empirical results are not always consistent as expected, partly because most studies used non-probability samples or collected data in only a few countries (Bellman et al., 2004; Cho et al., 2009; Krasnova et al., 2012; Milberg et al., 2000). Given the mixed results, this study uses a global random sample of Twitter users to test the following hypotheses anew:

H5a. Users from highly individualistic societies are more likely to protect their Twitter accounts than those from the societies with low individualism.

H5b. Users from societies with high uncertainty avoidance are more likely to protect their Twitter accounts than those from societies with low uncertainty avoidance.

For self-disclosure, people from individualistic societies have a strong tendency toward keeping secrets and preserving privacy (Petronio, 2002). In contrast, people from collectivist societies exhibit more trust and prefer disclosing personal information to other members within the community (Miltgen and Peyrat-Guillard, 2014). However, Rui and Stefanone (2013) argued that IND is positively related to self-disclosure in cyberspace to the extent that people from highly individualistic societies disclose more in order to compete for public attention as a personal achievement. Existing empirical evidence remains ambiguous. Rosen et al. (2010) found that social media users with individualistic cultural identities share more digital photos, but Rui and Stefanone (2013) found that Singaporeans shared more photos on Facebook, and Americans updated statuses more frequently. Therefore, we ask the following:

RQ1. What is the relationship between cultural values and self-disclosure of geolocation in tweets?

National culture might further moderate the relationship between privacy protection and self-disclosure. According to *H1*, users who protect their accounts from public access are more likely to disclose their geolocation information, because they might infer audience representations from the boundary defined by the privacy setting on Twitter. Individuals from different cultures may understand privacy boundaries very differently. People from collectivistic cultures are expected to strongly differentiate between in-group and out-group members. Individualists, on the other hand, are less likely to see the difference between in-group and out-group members (Krasnova et al., 2012). Therefore,

social media users in individualistic societies might perceive little boundary difference between the protected and public accounts, whereas users in collectivistic societies consider the difference important. Following this rationale, users in collectivistic societies are more inclined to disclose personal information when their accounts are protected:

H6. The difference in self-disclosure between protected and public accounts is smaller in individualistic societies than in collectivistic societies.

Method

Data collection

We collected a random sample of Twitter accounts using the Twitter API. We employed the method proposed by Liang and Fu (2015) and Zhu et al. (2011) to generate random Twitter user IDs. Twitter ID is a unique numeric value. Twitter users can change their screen names, but they can never change their Twitter IDs. A list of random Twitter IDs represent a random sample of Twitter users. Using this method, we identified 3,328,793 valid Twitter accounts. The random sample could represent the population of Twitter users as of November 2014. We collected the profiles of all sampled users in January 2015.

Measures

All Twitter account variables were obtained through the Twitter API. For each ID, the following fields were recorded for constructing different variables: privacy protection ("protected"), geo-information disclosure ("geo_enabled"), activity frequency ("statuses_count"), account age ("created_at"), network size ("followers_count," "friends_count"), and society ("location"). The Twitter API provides information about the profiles of all accounts including the protected ones.

Privacy protection and geo-information disclosure. Privacy protection is coded to be 1 or 0, with 1 indicating the user chose to set his or her Twitter account to be private. In our sample, about 5.4% (181,072) of the accounts chose to protect their tweets. Similar to privacy protection, geo-information disclosure indicates whether a user added geolocation to their tweets. About 9.6% (319,422) of all accounts enabled the geolocation setting.

Activity frequency. Twitter activity frequency was operationalized as the number of tweets. The mean number of tweets posted by an individual account in our sample is 447 (Mdn = 1, *SD* = 3637). Nearly 40% of the users did not post any tweets.

Network size. Social media network size includes two indicators: number of followers and number of followings. The mean number of followers is 79 (Mdn = 1, *SD* = 5961), and the mean number of followings is 74 (Mdn = 8, *SD* = 654). More than 40% of users did not have any followers, and 24% of users did not follow any accounts.

Account age. Account age was calculated by the number of days between date of data collection and account registration time. The mean age is 805 days (Mdn = 722, *SD* = 538)—more than 2 years.

Country. Information extracted from the "location" field was used to generate country or territory information. These self-declared locations are usually unstructured (e.g. I live in California). We, thus, applied an automatic geocoder provided by the Data Science Toolkit website (http://www.datasciencetoolkit.org/) to convert these unstructured texts into society labels. We identified 233 country-level locations in our sample. We excluded countries with fewer than 150 users to ensure that a fairly large amount of users could be used to represent the country or location. Finally, our data include 473,441 users from 104 societies. The distribution of the number of users in each country is shown in Appendix 1. To test the accuracy of the geocoder, we extracted the coordinates of the geo-enabled tweets posted by 6234 public users, which is the maximum number of users we can obtain within our sample to the extent that the Twitter profile API only returns the most recent status update for each user. By comparing the results generated by the geocoder according to users' profile information and according to the coordinates embedded in users' tweets, we found that 82% (5084/6234) of the results were matched (at the country level). A large portion of the mismatched cases were due to users' traveling behavior. Travelers especially like to post with geo-enabled tweets. Therefore, the accuracy level of the geocoder we used could be considered pretty high.

Cultural dimension. We included two of Hofstede's cultural dimensions in the study: IND versus collectivism and UAI. Data on these two dimensions were harvested from Hofstede's official website (http://geert-hofstede.com/index.php). In our sample, the mean scores for IND and UAI are 39.38 (*SD* = 22.76, *N* = 82) and 64.73 (*SD* = 22.03, *N* = 82), respectively. In regression analysis, the two indicators were rescaled (to be divided by 100).

Internet penetration. Internet penetration data in 2014 were collected from Internet World Stats (*M* = 54.59, *SD* = 26.86). This variable was controlled in our analysis because Internet penetration is highly correlated with important demographic factors, such as age, gender, and education (e.g. Chinn and Fairlie, 2007), which are found to be significant predictors of online privacy practices (e.g. Cho et al., 2009).

Analytic strategies

In our sample, only 14% of users' society information was identifiable (*N* = 473,441). List-wise deletion is one of the most common techniques for handling missing data, but it requires the data to meet the assumption of missing completely at random (MCAR)—the probability of missing should be unrelated to the independent variables and the dependent variables as well. In our case, it is highly possible that users who did not report their location information are biased toward less self-disclosure and more privacy protection. Obviously, the use of list-wise deletion will be inappropriate.

Instead, this study treated missing cases as non-responses. To adjust for non-response bias, analogous to survey data collection, we employed the logistic propensity model

(see Little, 1986). First, we created a new variable to indicate missing location information (1: responded, 0: missing, $N = 3{,}328{,}793$). Second, we used all available variables to predict disclosure of location information and computed the response probability for each case (p_i). Finally, the weight of user i is given by $w_i = 1/p_i$, which means we gave users who are less inclined to disclose a higher weight to balance the missing cases. The weighted proportion of the protected accounts using the 473,441 sub-sample is 5.4% that is the number estimated from the full sample, suggesting our weighting approach to be valid.

The generalized logistic multi-level regression (Snijders and Bosker, 2012) was employed to test our hypotheses. In our study, each user nested under the same country could be influenced by the unique characteristics of that particular society. We chose logistic as the link function because our dependent variables are binary responses (i.e. protected or not, geo-enabled or not). Although we have identified 104 societies, only 82 societies have the cultural dimension measures. Therefore, in our multi-level models, a total of 460,232 users were nested under 82 countries. All Twitter measures are Level-1 variables. All societal-level predictors are Level-2 variables (i.e. IND, UAI, and Internet penetration).

Results

Descriptive statistics

According to Figure 1, regional variations of privacy setting (A) and disclosure of geolocation (B) could be seen clearly. Users from Southeast Asia, South Asia, and Central Africa were more inclined to protect their Twitter accounts. The maximum percentage is 15% in Singapore. Eastern Europeans were less inclined to protect their accounts. The minimum percentage is 2% in Russia. Users from Southeast Asia, South Europe, and several African and South American countries are more inclined to enable geolocation in their tweets, in contrast to users from North America, Russia, and Australia. The maximum percentage is 18% in Kenya, while the minimum is 6.6% in Russia. Chi-Square tests showed that these regional differences were statistically significant both for privacy protection ($\chi^2_{weighted}(104, N = 473{,}441) = 2{,}767{,}697,\ p < .01$) and geolocation disclosure ($\chi^2_{weighted}(104, N = 473{,}441) = 1{,}908{,}494,\ p < .01$).

Although the national variations of privacy protection and self-disclosure are both statistically significant, according to Figure 1, cultural values appear less likely to be the deterministic predictors. Users in similar cultures (e.g. China, Taiwan, and Japan) exhibited different levels of privacy protection, while users in different cultures (e.g. Malaysia and Spain) presented a similar level of self-disclosure.

Concerning *H1*, the correlation at the society level is significant yet weak (*Spearman's rho* = 0.31, $p < 0.001$, $N = 104$). At the individual level, 38.1% of protected user accounts have enabled geolocation, whereas 8.0% of public user accounts did this ($\chi^2(1, N = 3{,}328{,}793) = 179{,}431,\ p < .01$). The mismatch between aggregate- and individual-level results suggests that other variables might be important, and multi-level modeling is necessary.

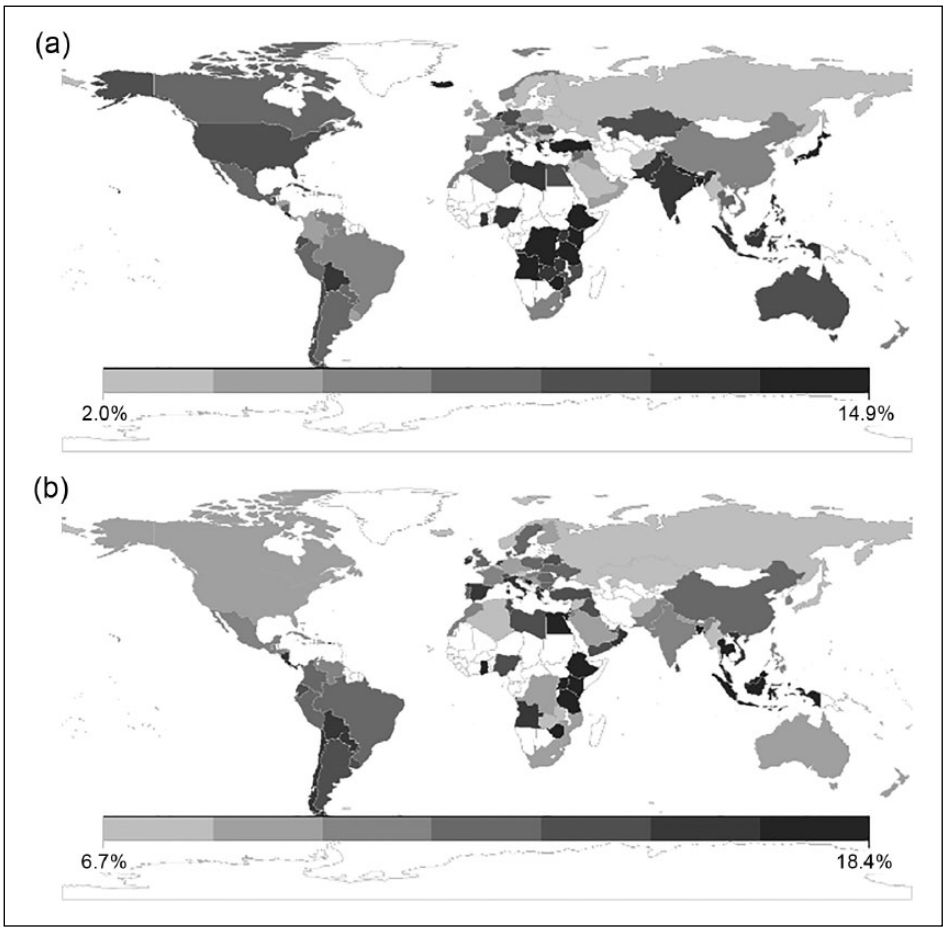


Figure 1. The proportions of (a) protected users and (b) geo-enabled users across 104 societies.
Note: Both proportions were weighted by the logistic propensity score. White areas are those countries with fewer than 150 cases in our sample.

Multi-level logistic regression analyses

Table 1 presents the weighted coefficients for predicting privacy protection and self-disclosure of geolocation. Model fits are reasonably good—the explained variations are 16.7% and 25.4%, respectively. Concerning *H1*, Model 2 further confirmed that protected users are more likely to add geolocation when tweeting. The odds of geo-disclosure for protected users are about 25 times higher than that of public users ($B = 3.205$, $Z = 0.354$, $p < .01$). Therefore, *H1* was fully supported.

H2 through *H4* focus on the impacts of individual-level factors. First, in terms of network size, users with more followers were more likely to keep their account public, whereas users with more followings were more likely to protect their accounts. This

Table 1. Multi-level logistic regression predicting protection and disclosure.

	Model 1: Protected vs public		Model 2: Geo-enabled vs not	
	Estimate (SE)	*Z*	Estimate (SE)	*Z*
log no. of followers	−0.624** (0.003)	−230.96	0.024** (0.002)	11.62
log no. of followings	0.301** (0.002)	142.94	0.180** (0.002)	96.86
log account age (days)	0.574** (0.003)	167.32	0.234** (0.002)	98.93
log no. of tweets	0.293** (0.002)	193.09	0.241** (0.001)	201.78
Protected vs public			3.205** (0.354)	9.06
log no. of followers × protected			−0.344** (0.005)	−70.55
log no. of followings × protected			−0.062** (0.005)	−12.80
Internet penetration%	−0.728** (0.178)	−4.08	−0.369* (0.176)	−2.09
IND	0.070 (0.213)	0.33	−0.017 (0.193)	−0.09
UAI	−0.380* (0.170)	−2.23	−0.028 (0.148)	−0.19
Protected × IND			−1.456** (0.561)	−2.60
Protected × UAI			−0.359 (0.362)	−0.99
Intercept	−6.267** (0.142)	−44.02	−4.610** (0.124)	−36.95
Var. of intercepts across societies	0.102 (0.319)		0.077 (0.278)	
Var. of protected across societies			0.635 (0.797)	
Log-likelihood	−678,924.4		−923,807.3	
Explained variation	16.7%		25.4%	
No. of users	460,232		460,232	
No. of societies	82		82	

Note: All results were weighted by the logistic propensity score. SE: standard error.
**$p < .01$, *$p < .05$.

indicates that followers and followings should have different functions for privacy practices on Twitter. As predicted in *H2*, privacy protection significantly moderates the relationship between network size and self-disclosure of geolocation ($B = -0.344$, $Z = -70.55$, $p < .01$). Figure 2(a) shows that the number of followers is positively associated with geo-disclosure for public users ($B = 0.024$, $Z = 11.62$, $p < .01$), whereas the relationship

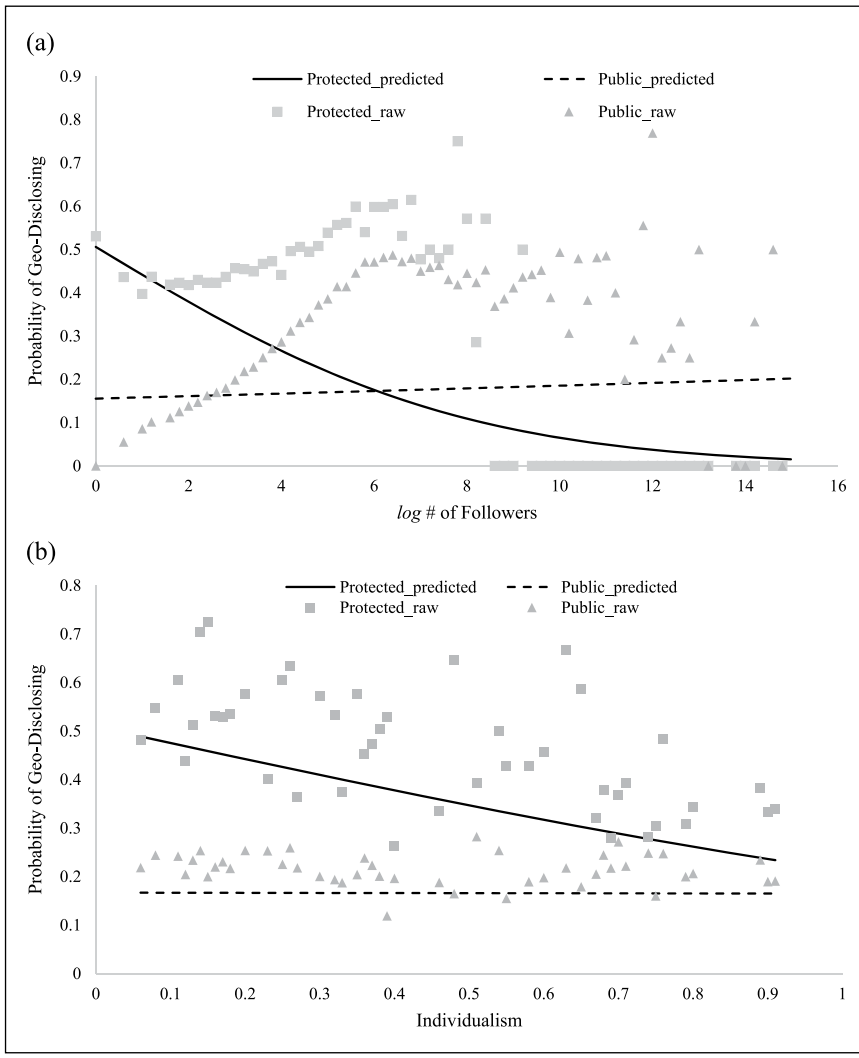


Figure 2. The probability of geolocation disclosure against (a) number of followers and (b) individualism index. The predicted probabilities were calculated based on the models in Table 1. The raw probabilities are the empirical percentages.

is negative for protected users ($B = 0.024-0.344 = -0.320$, $p < .01$). The slope for the positive relationship in Figure 2(a) is rather flat, suggesting a small overall effect size, but the main concern of interest is the conditional impacts of network size on self-disclosure for different groups of users (*H2*). Notice that the predicted probability values are different from the observed ones in Figure 2(a). Such discrepancy doesn't necessarily mean a lack of fit. The predicted probability reflects the "net" impact of network size on

self-disclosure when all other variables were controlled, whereas the raw probability values show the pure impact of the predictor.

We used the number of tweets to measure Twitter activity frequency and used account age to measure user experience. The number of tweets (activity) was positively associated with privacy protection ($B = 0.293$, $Z = 193.09$, $p < .01$) and self-disclosure ($B = 0.241$, $Z = 201.78$, $p < .01$). Similarly, user experience was positively associated with privacy protection ($B = 0.574$, $Z = 167.32$, $p < .01$) and self-disclosure ($B = 0.234$, $Z = 98.93$, $p < .01$). Both *H3* and *H4* were fully supported.

H5 focuses on the role of national culture in privacy practices. As Model 1 presents, users in high UAI societies were less likely to protect their accounts ($B = -0.380$, $Z = 0.170$, $p < .05$), which was opposite to our hypothesis. IND was not significantly related to privacy protection. Therefore, neither *H5a* nor *H5b* was supported.

Regarding *RQ1*, there are no direct relationships between cultural values and self-disclosure of geolocation in Model 2. It is possible that the correlations between cultural values and self-disclosure have been mediated by privacy protection. To examine this, we excluded privacy protection and reran Model 2. The coefficients remain non-significant.

As expected in *H6*, national culture could influence self-disclosure indirectly. IND significantly moderated the relationship between privacy protection and self-disclosure ($B = -1.456$, $Z = 0.561$, $p < .01$). Figure 2(b) shows that the gap of geo-disclosure between protected and public users was larger in collectivistic societies ($B = 3.205$) than that in individualistic societies ($B = 3.205 - 1.456 = 1.749$). The difference of predicted probability at IND = 0.06 is 0.32 ($p < .01$), whereas the difference is 0.07 at IND = 0.91 ($p > .05$).

In addition to cultural values, Internet penetration was found to be the most significant national-level variable in predicting both privacy protection ($B = -0.728$, $Z = -4.08$, $p < .01$) and self-disclosure of geolocation ($B = -0.369$, $Z = -2.09$, $p < .05$). Users from Internet-developed countries were more inclined to keep their accounts public and less inclined to add geo-tags to their tweets.

Discussion

Using a large-scale and representative social media dataset, this study investigated the privacy practices and self-disclosure of global Twitter users. In particular, we found that using the privacy setting indeed increases the likelihood of disclosing geolocation in tweets. This relationship is believed to be an effective solution for the well-known privacy paradox on social media (Barnes, 2006)—individuals with strong privacy concerns were found to disclose large amounts of personal information online. The rationale is that people perceive self-disclosure within controlled boundaries as safe. The major contribution of this study is to expand this argument to the global setting and argued for the importance of societal-level differences in terms of people's social media use behavior. Indeed, we identified considerable cultural and societal differences.

First, privacy setting in collectivistic societies was more effective in encouraging self-disclosure; whereas it appears less important for users in individualistic societies. As we explained earlier, this may be because people from collectivistic cultures perceived in-group and out-group differences to be larger. In terms of CPM, the same privacy boundary

could mean different things for collectivists and individualists. People in individualistic cultures tend to create wider boundaries than those in collectivistic cultures. A post hoc analysis suggested that IND was positively associated with the number of followers (*log*) even when privacy protection is controlled for ($B = 0.379$, $Z = 41.57$, $p < .01$).

Second, our expectation of cultural values' impact on privacy practices was not supported, and we found the opposite. UAI was negatively associated with privacy protection. Previous studies on cross-cultural comparisons found online privacy concerns to be positively related to UAI (e.g. Bellman et al., 2004; Milberg et al., 2000). The contradictory results might be caused by different privacy-related variables: privacy concern and privacy protection behavior. Cho et al. (2009) found that UAI tendency is positively correlated with behavioral avoidance and negatively correlated with proactive protections.

Third, we also discovered that cultural values were not directly related to geolocation disclosure. The relationship between culture and self-disclosure is conditional (Gudykunst et al., 1996; Rui and Stefanone, 2013). We found that IND was negatively related to geo-disclosure only for the protected users. This implies that the underlying mechanism behind IND and self-disclosure is trust (Miltgen and Peyrat-Guillard, 2014; Petronio, 2002) other than self-achievement (Rosen et al., 2010; Rui and Stefanone, 2013), at least for the protected users. If the trust mechanism works, users in collectivistic societies who exhibit more trust will disclose more personal information. If the self-achievement mechanism works, users in individualistic societies who consider public attention as a personal achievement will disclose more personal information.

Internet penetration, though it was treated as a control variable, was found to be a significant predictor in both models. Users in societies with higher penetration rate were more likely to keep their accounts public and less likely to disclose geolocation in tweets. This finding could help explain the variation presented in Figure 1. Lower penetration areas (e.g. Southeast Asia and Africa) are more likely to protect their accounts and add geolocation information in tweets. Internet penetration matters because, in areas with low penetration, the early adopters are mostly social elites (Chinn and Fairlie, 2007), and their privacy practices might be different from those of grassroots users in a society with high Internet penetration. The elites might be more aware of online privacy and more capable of protecting themselves. In this sense, the observed variations of privacy practices across societies were mainly caused by the differences in user compositions rather than cultural values. If this is true, we could expect that these differences will phase out when the global Internet penetration gap shrinks.

Concerning individual-level factors, the number of followers and the number of followings exhibited different effects on privacy protection. A large number of followers bring both privacy risks and benefits (Rui and Stefanone, 2013). Users have to weigh the benefits against the risks to decide whether protection is needed. A negative relationship between the number of followers and privacy protection suggests that Twitter users generally consider a large number of followers as beneficial. We also found for public users, having more followers indicates higher probability of geo-disclosure. This is because public users are motivated to self-disclose for building social capital (Choi and Bazarova, 2015) and for increasing the possibility of being retweeted (Bakshy et al., 2012; Suh et al., 2010). The number of followings implies different behavioral motivations. Unlike being followed, following other users is more voluntary behavior. All other things being

equal, users with more followings are more likely to be passive information consumers. They are more interested in what other people are saying than tweeting about themselves. These users are less likely to be motivated by being retweeted in the wide and unbounded network, and they are more inclined to protect their account.

It is important to note that the interpretation of our empirical findings is grounded upon a few assumptions, which deserve further elaboration. First, to conceptualize privacy setting as a form of privacy boundary management, we assume that many if not most private users are the ones who choose to protect their accounts out of privacy concerns. If this assumption is untenable (e.g. it is possible that some users do not understand the openness of information sharing on Twitter and simply keep their accounts private for other purposes), the findings could become a mere reflection of the differences between people who don't know what Twitter is and those who do. The best way to solving this problem is to look into the motivations of the private users. Although we do not have survey data about this at hand, our analysis partially hints at the answer to this question. Foremost, Model 2 in Table 1 shows that private setting indeed predicts self-disclosure of geolocation, which implies that users think the two functions are inherently related. In addition, users might have the chance to misunderstand the openness feature of Twitter at the very beginning, and to adapt their behaviors later on through peer interaction on the platform. In this study, we included the account age (days since registration) as a control variable in the analysis to avoid this confounding effect. According to Table 1, old users are more inclined to keep their accounts private. Therefore, multiple pieces of evidence seem to suggest that the private users keep their accounts intentionally for privacy concerns.

Second, the current study focuses on the behavioral expression, instead of the psychological demand of privacy concerns and self-disclosure. Privacy concerns and privacy protection behavior are two different constructs. Privacy concern has been demonstrated an ineffective predictor of privacy behaviors (Barnes, 2006; Stutzman and Kramer-Duffield, 2010). Since privacy setting on Twitter is opt-in setting, it requires additional effort, knowledge, and skills to adopt this function. There exist a certain amount of users who feel the need but do not possess the knowledge and skills to set up privacy protections appropriately. In other words, the percentages we presented based on behavioral indicators could be lower than the percentage of people who desire for more privacy. However, this descriptive bias has little impact on our testing of hypotheses. Our main interests are focusing on the relationship between the privacy setting behavior and geo-disclosure behavior on Twitter. Even if the percentages were underestimated, the correlation between two items should remain robust, because if users have difficult time configuring their privacy settings, it is highly likely that they have trouble using the geo-tag function.

Limitations and future research

Several limitations of our study should be noted. First, we used an unobtrusive method to collect representative Twitter user data. While the validity of data obtained through an unobtrusive method is high, the disadvantage is that it is difficult to know users' demographic backgrounds, which is important in predicting privacy practices (e.g. Cho et al.,

2009; Stutzman and Kramer-Duffield, 2010). In addition, we were not able to ascertain whether the Twitter accounts included in our sample were operated by an institution (e.g. media outlets) or an individual. It is reasonable to believe that individual accounts and institutional accounts differ in terms of their tweeting and account-setting behaviors.

It is nevertheless important to point out that our sample includes a certain portion of accounts which were largely inactive or even accounts generated by robots, whose behaviors could be different from everyday user behavior. However, after removing the inactive/likely robot accounts—defined as accounts with 0 followers and 0 followings, we found the regression results remain roughly the same, which attests to the robustness of our findings (see Appendix 2). There is one easily noticeable difference at the descriptive level for the two dependent variables though: the percentages of Chinese-protected users and geo-enabled users turn out to be higher after the inactive account removal (see Appendix 3). Despite the fact that Twitter is blocked in China, there is a considerable amount of Chinese Twitter users who use circumvention tools to get around the firewall. The heightened percentages for China might suggest that "active" Twitter users in China are more likely to protect their accounts compared to users from other regions.

Second, our study focused on the role of national culture in privacy practices and self-disclosure. However, this does not mean that other societal variables should be ignored. For instance, Internet penetration plays a vital role in our models. Bellman et al. (2004) found that government involvement in regulation directly influences information privacy across countries. The aggregate level of engagement in Twitter community could also be a potential predictor for privacy practices. For societies with alternative social media platforms, such as Weibo in China, people there could be less engaged and thus show less concern for privacy. Future studies might include other national-level predictors.

Third, it is also important to acknowledge that the nature of our data is static but not dynamic. It will be highly possible that there are reciprocal and multiple causal relationships between the key variables we examined, for instance, between network size, posting activity, privacy protection, and self-disclosure. The number of followers is highly correlated with the number of tweets ($r = 0.80$, $p < .01$) and the number of followings ($r = 0.79$, $p < .01$). In traditional linear regression models, this indicates the possibility of a high level of multicollinearity. If this is the case, our test of hypotheses involving the number of followers (*H2*) and the number of tweets (*H3*) might incur biased estimates. To ensure the technical soundness of our analysis, we calculate the Kappa condition value as suggested by Baayen (2008). It turns out the values are 15.78 and 16.56 for Model 1 and Model 2, respectively (smaller than 30), which suggests a moderate collinearity.

In terms of reciprocal causal relationship, it is possible that self-disclosure relates to privacy protection through other mechanisms than what we proposed. For example, people who self-disclose more will be more aware of their privacy and therefore set more strict privacy settings. Or according to the reinforcing spiral framework (Slater, 2007), which assumes a bi-directional mutual influence between communication technology adoption and the attitudinal and behavioral outcomes, privacy setting and self-disclosure could be mutually causal process. Yet, the direction of causal flow cannot change the moderation role of the cultural indicators—our main interest of inquiry.

Nevertheless, one promising direction of future studies is to track users of different types of privacy settings and see how the correlates of privacy settings change across time, for instance, to use the privacy setting at time 1 to predict the change of the geolocation setting at time 2.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This research was supported by the Small Project Funding from The University of Hong Kong (201409176011) and the Public Policy Research Fund, Hong Kong Government (2013.A8.009.14A).

References

Altman I (1975) *The Environment and Social Behavior: Privacy, Personal Space, Territory, and Crowding*. Monterey, CA: Brooks/Cole.

Altman I (1977) Privacy regulation: culturally universal or culturally specific. *Journal of Social Issues* 33(3): 66–84.

Baayen RH (2008) *Analyzing Linguistic Data: A Practical Introduction to Statistics Using R*. Cambridge: Cambridge University Press.

Bakshy E, Rosenn I, Marlow C, et al. (2012) The role of social networks in information diffusion. In: *Proceedings of the 21st international conference on world wide web*, New York, 16 April, pp. 519–528. New York: ACM.

Barnes SB (2006) A privacy paradox: social networking in the United States. *First Monday* 11(9). Available at: http://firstmonday.org/article/view/1394/1312

Bellman S, Johnson EJ, Kobrin SJ, et al. (2004) International differences in information privacy concerns: a global survey of consumers. *Information Society* 20(5): 313–324.

boyd D (2006) Friends, friendsters, and myspace top 8: writing community into being on social network sites. *First Monday* 11(12). Available at: http://firstmonday.org/article/view/1418/1336

boyd D (2008) *Taken out of context: American teen sociality in networked publics*. PhD Thesis, University of California, Berkeley, CA.

boyd D and Hargittai E (2010) Facebook privacy settings: who cares? *First Monday* 15(8). Available at: http://firstmonday.org/article/view/3086/2589

boyd D and Marwick AE (2011) Social privacy in networked publics: teens' attitudes, practices, and strategies. In: *A decade in Internet time: symposium on the dynamics of the Internet and society*, Oxford, 22 September.

Calhoun KJ, Teng JTC and Cheon MJ (2002) Impact of national culture on information technology usage behaviour: an exploratory study of decision making in Korea and the USA. *Behaviour & Information Technology* 21(4): 293–302.

Casey D (2010) Replay it: Google search across the Twitter archive. In: *The official Google blog*. Available at: googleblog.blogspot.hk/2010/04/replay-it-google-search-across-twitter.html

Child JT and Petronio S (2011) Unpacking the paradoxes of privacy in CMC relationships: the challenges of blogging and relational communication on the Internet. In: Wright KB and Webb KM (eds) *Computer-Mediated Communication in Personal Relationships*. New York: Peter Lang, pp. 21–40.

Chinn MD and Fairlie RW (2007) The determinants of the global digital divide: a cross-country analysis of computer and Internet penetration. *Oxford Economic Papers–New Series* 59(1): 16–44.

Cho H, Rivera-Sanchez M and Lim SS (2009) A multinational study on online privacy: global concerns and local responses. *New Media & Society* 11(3): 395–416.

Choi YH and Bazarova NN (2015) Self-disclosure characteristics and motivations in social media: extending the functional model to multiple social network sites. *Human Communication Research* 41(4): 480–500.

Ellison NB, Vitak J, Steinfield C, et al. (2011) Negotiating privacy concerns and social capital needs in a social media environment. In: Trepte S and Reinecke L (eds) *Privacy Online*. New York: Springer, pp. 19–32.

Elwood S and Leszczynski A (2011) Privacy, reconsidered: new representations, data practices, and the geoweb. *Geoforum* 42(1): 6–15.

Friedland G and Sommer R (2010) Cybercasing the joint: on the privacy implications of geo-tagging. Available at: www.usenix.org/legacy/events/hotsec10/tech/full_papers/Friedland.pdf

Gudykunst WB, Matsumoto Y, TingToomey S, et al. (1996) The influence of cultural individualism-collectivism, self construals, and individual values on communication styles across cultures. *Human Communication Research* 22(4): 510–543.

Hofstede G (1980) *Culture's Consequences: International Differences in Work Related Values*. Beverly Hills, CA: SAGE.

Hofstede G (1991) *Cultures and Organizations: Software of the Mind*. New York: McGraw Hill.

Humphreys L, Gill P and Krishnamurthy B (2014) Twitter: a content analysis of personal information. *Information, Communication & Society* 17(7): 843–857.

Krasnova H, Veltri NF and Gunther O (2012) Self-disclosure and privacy calculus on social networking sites: the role of culture intercultural dynamics of privacy calculus. *Business & Information Systems Engineering* 4(3): 127–135.

Lewis K, Kaufman J and Christakis N (2008) The taste for privacy: an analysis of college student privacy settings in an online social network. *Journal of Computer-Mediated Communication* 14(1): 79–100.

Liang H and Fu KW (2015) Testing propositions derived from Twitter studies: generalization and replication in computational social science. *PLoS ONE* 10(8): e0134270.

Little RJA (1986) Survey nonresponse adjustments for estimates of means. *International Statistical Review* 54(2): 139–157.

Liu C, Marchewka JT and Ku C (2004) American and Taiwanese perceptions concerning privacy, trust, and behavioral intentions in electronic commerce. *Journal of Global Information Management* 12(1): 18–40.

Madden M, Lenhart A, Cortesi S, et al. (2013) Teens, social media, and privacy. Available at: www.pewinternet.org/2013/05/21/teens-social-media-and-privacy

Milberg SJ, Smith HJ and Burke SJ (2000) Information privacy: corporate management and national regulation. *Organization Science* 11(1): 35–57.

Miltgen CL and Peyrat-Guillard D (2014) Cultural and generational influences on privacy concerns: a qualitative study in seven European countries. *European Journal of Information Systems* 23(2): 103–125.

Papacharissi Z and Gibson PL (2011) Fifteen minutes of privacy: privacy, sociality, and publicity on social network sites. In: Trepte S and Reinecke L (eds) *Privacy Online*. New York: Springer, pp. 75–89.

Petronio S (1991) Communication boundary management: a theoretical model of managing disclosure of private information between marital couples. *Communication Theory* 1(4): 311–335.

Petronio S (2002) *Boundaries of Privacy: Dialectics of Disclosure*. New York: State University of New York Press.

Rainie L and Madden M (2015) Americans' privacy strategies post-Snowden. Available at: http://www.pewinternet.org/2015/03/16/americans-privacy-strategies-post-snowden

Rosen D, Stefanone MA and Lackaff D (2010) Online and offline social networks: investigating culturally-specific behavior and satisfaction. In: *The 43rd Hawaii international conference on the system sciences (HICSS)*, Honolulu, HI, 5–8 January, pp. 1–10. New York: IEEE.

Rui J and Stefanone MA (2013) Strategic self-presentation online: a cross-cultural study. *Computers in Human Behavior* 29(1): 110–118.

Singh T and Hill ME (2003) Consumer privacy and the Internet in Europe: a view from Germany. *Journal of Consumer Marketing* 20(7): 634–651.

Slater MD (2007) Reinforcing spirals: the mutual influence of media selectivity and media effects and their impact on individual behavior and social identity. *Communication Theory* 17(3): 281–303.

Snijders TAB and Bosker RJ (2012) *Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modeling*. 2nd ed. London: SAGE.

Stutzman F and Hartzog W (2012) Boundary regulation in social media. In: *Proceedings of the ACM 2012 conference on computer supported cooperative work*, Seattle, WA, 11–15 February, pp. 769–778. New York: ACM.

Stutzman F and Kramer-Duffield J (2010) Friends only: examining a privacy-enhancing behavior in Facebook. In: *Proceedings of the SIGCHI conference on human factors in computing systems*, Paris, 27 April–2 May 2013, pp. 1553–1562.

Stutzman F, Capra R and Thompson J (2011) Factors mediating disclosure in social network sites. *Computers in Human Behavior* 27(1): 590–598.

Suh B, Hong L, Pirolli P, et al. (2010) Want to be retweeted? Large scale analytics on factors impacting retweet in twitter network. In: *IEEE second international conference on social computing*, Minneapolis, MN, 20–22 August, pp. 177–184. New York: IEEE.

Youn S (2009) Determinants of online privacy concern and its influence on privacy protection behaviors among young adolescents. *Journal of Consumer Affairs* 43(3): 389–418.

Young AL and Quan-Haase A (2013) Privacy protection strategies on Facebook: the Internet privacy paradox revisited. *Information, Communication & Society* 16(4): 479–500.

Zhu JJH, Mo Q, Wang F, et al. (2011) A random digit search (RDS) method for sampling of blogs and other user-generated content. *Social Science Computer Review* 29(3): 327–339.

Author biographies

Hai Liang is a post-doctoral fellow at the Journalism and Media Studies Centre, The University of Hong Kong. His research focuses on computational social science, social media, and political communication.

Fei Shen is an associate professor in the Department of Media and Communication, City University of Hong Kong. His research interests include public opinion, media effects, data mining, social movement, and consumer behavior.

King-wa Fu is an associate professor at the Journalism and Media Studies Centre, the University of Hong Kong. His research focuses on political participation and media use, computational media studies, mental health/suicide and the media, health communication, young people's Internet use, and statistics for journalism.

Appendix 1. Number of users in 104 societies.

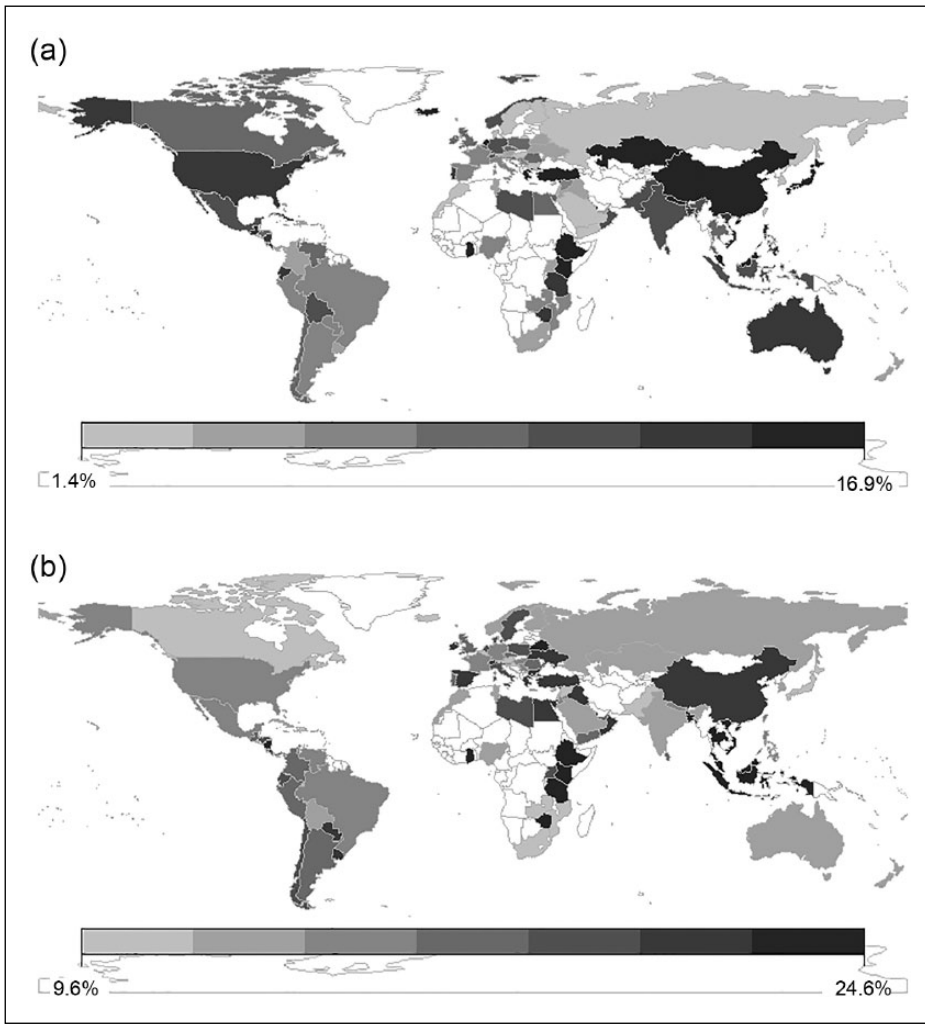
Society	No. of users	No. of weighted	Society	No. of users	No. of weighted	Society	No. of users	No. of weighted
United States	134,165	698,200	China	1402	12,273	Algeria	221	3424
Indonesia	47,729	254,218	Kenya	1617	11,565	Bulgaria	301	3410
Brazil	36,598	207,113	Ukraine	1608	11,196	Belarus	490	3346
Japan	10,076	156,345	Peru	1445	10,114	Finland	477	3341
India	18,351	154,313	Bangladesh	1096	9537	Uruguay	487	3272
United Kingdom	20,524	104,821	Singapore	1583	9384	Sri Lanka	391	3257
Turkey	12,156	101,552	Ecuador	1548	8892	Vietnam	341	3112
Russia	11,901	98,988	Kazakhstan	906	8821	Denmark	475	3059
Mexico	16,413	94,340	Taiwan	823	8434	Zambia	431	2969
Philippines	10,388	69,397	Kuwait	938	8025	Tunisia	328	2940
South Africa	8163	61,643	New Zealand	1200	7978	Switzerland	393	2674
Spain	10,041	56,442	Czech Republic	959	7894	Honduras	365	2598
Argentina	10,211	56,325	Guatemala	1176	7345	Nicaragua	374	2570
Colombia	8615	54,050	Sweden	1075	7231	Zimbabwe	302	2491
France	6876	50,056	Libya	322	6823	Israel	211	2458
Italy	6887	48,608	Romania	842	6518	Lebanon	238	2316
Canada	8821	45,989	Dominican Republic	994	6316	Croatia	280	2224
Saudi Arabia	2517	34,881	Greece	925	6142	Angola	230	2207
South Korea	3365	34,401	Serbia	701	6130	Uganda	279	2191
Nigeria	4334	33,000	Portugal	910	6096	Afghanistan	199	2180
Australia	6087	32,418	Ireland	1181	6040	Bahrain	250	2035
Germany	4203	27,157	Austria	858	5950	Democratic Republic of the Congo	194	1995
Venezuela	4937	26,500	Paraguay	973	5732	Ethiopia	243	1969
Malaysia	3818	26,028	Poland	766	5639	Republic of the Union of Myanmar	173	1925
Pakistan	2619	24,083	Mozambique	521	5183	Costa Rica	253	1901
Cuba	4505	23,966	Syria	284	5157	Bolivia	264	1782
Netherlands	3904	21,401	Norway	819	4861	Panama	314	1693
Chile	3522	18,332	Jordan	323	4848	Iceland	282	1672
Egypt	1819	17,998	Hong Kong	538	4378	Qatar	208	1671
Thailand	1710	17,976	Tanzania	482	4176	Slovenia	214	1399
Belgium	3100	17,456	Ghana	491	3934	Bosnia and Herzegovina	156	1348
Morocco	1325	14,128	Nepal	405	3733	Macedonia (FYROM)	152	1299
United Arab Emirates	1705	13,237	Puerto Rico	532	3642	Cyprus	212	1283
Yemen	584	13,065	Oman	299	3542	Jamaica	173	1044
Iraq	605	12,899	Hungary	424	3493	**Total**	**473,441**	**3,037,403**

Note: "Weighted" indicates the number of weighted users by the logistic propensity score.

Appendix 2. Multi-level logistic regression predicting protection and disclosure (likely robot accounts excluded).

	Model 1: Protected vs public		Model 2: Geo-enabled vs not	
	Estimate (SE)	*Z*	Estimate (SE)	*Z*
log no. of followers	–0.542** (0.003)	–155.97	–0.001 (0.002)	–0.56
log no. of followings	0.301** (0.003)	99.53	0.158** (0.002)	70.59
log account age (days)	0.505** (0.004)	166.47	0.256** (0.003)	94.75
log no. of tweets	0.310** (0.002)	172.28	0.254** (0.001)	197.72
Protected vs public			1.636** (0.212)	7.71
log no. of followers × protected			–0.335** (0.006)	–52.12
log no. of followings × protected			0.119** (0.007)	16.79
Internet penetration%	–0.459** (0.190)	–2.41	–0.432* (0.169)	–2.56
IND	–0.042 (0.226)	–0.18	0.007 (0.179)	0.04
UAI	–0.357* (0.182)	–1.96	–0.094 (0.141)	–0.67
Protected × IND			–0.927** (0.271)	–3.43
Protected × UAI			0.250 (0.256)	0.97
Intercept	–6.343** (0.154)	–41.13	–4.546** (0.120)	–37.85
Var. of intercepts across societies	0.102 (0.319)		0.077 (0.277)	
Var. of protected across societies			0.377 (0.613)	
Log-likelihood	–451,006.5		–751,923	
Explained variation	15.4%		20.8%	
No. of users	375,806		375,806	
No. of societies	82		82	

Note: All results were weighted by the logistic propensity score. ** $p < .01$, * $p < .05$.



Appendix 3. The proportions of (a) protected users and (b) geo-enabled users across 96 societies (likely robot accounts excluded).