

Unlisted: Variations in Location Sharing by Craigslist Users

Dara E. Seidl and Christopher Allen

ABSTRACT: Personal location is frequently collected from Internet users, but less studied is the extent to which individuals are willing to provide location in different contexts. This study examines the situational context that drives Craigslist users to post different levels of location, in particular, how the precision of location added to a post varies by posting category and geographical region. A mix of urban and rural Craigslist regions are sampled from California with sale categories that are considered more sensitive, as well as categories where offering location upfront may be more important, such as with apartments and garage sales. This study finds that Craigslist users in more densely populated regions are more likely to post location on maps and more likely to move a pin on the map. This study also confirms that users are more likely to keep location off maps when content may signal family or financial status, as with baby merchandise, electronics, tickets, and jewelry.

KEYWORDS: privacy, geoprivacy, VGI, California, location sharing

Introduction

Privacy preferences are known to vary according to context. In privacy law, certain types of information are classified as *sensitive*, and therefore receive more stringent protections (Nissenbaum, 2009). Examples are health, religious beliefs, sexual preferences, financial status, and political opinions. Location adds a personally identifying element which, when tied to sensitive information, may lead to emotional or physical harm. For example, the California Proposition 8 maps, previously available at eightmaps.com, marked the names and locations of donors in support of banning gay marriage. Such public identification of controversial political views caused outcry over privacy and sparked fears of in-person retaliation.

Today, personal information is routinely shared with a tagged location, often in exchange for some perceived benefit (Sun et al., 2015). Previous work on drivers of social network location sharing (Li et al., 2013) suggests that privacy competes with users' motivations to disclose location. For instance, posting about expensive jewelry in order to sell it may signal financial status, potentially increasing burglary risk if location is included. The online classifieds site, Craigslist, offers a unique opportunity to study location privacy preferences based on the multiple resolutions at which location may be shared. Users who post advertisements are able to share location at multiple levels, including numbered street address, intersection, street name, zip code, or moving a pin on a map. Users may uncheck the default of showing location on maps (Figure 1), though zip code is still collected by Craigslist to post an advertisement. This study examines variations in location sharing by California Craigslist users according to posting category and

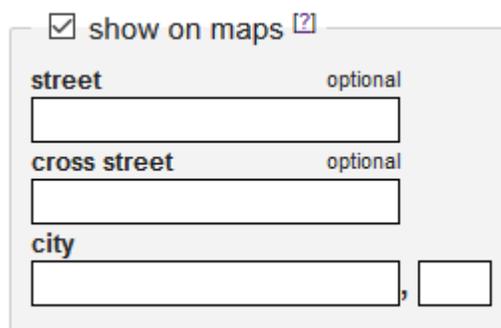
geographic region. Greater precision in location sharing is expected in less sensitive sale categories and in urban areas where location is shared by a higher density of individuals.

Categories of interest are baby+kid, electronics, free, jewelry, tickets, garage sale, apartments/housing, and rooms/shared. These categories include more sensitive indicators, such as expensive goods and the presence of children in a household, but also situations where providing an address is expected in a traditional offline sense, such as with garage sales and housing advertisements. Differences between groups are evaluated using chi-squared tests and logistic regression. Results suggest that those in more rural regions are more likely to hide precise locations when posting, as are those who are selling baby merchandise and expensive goods.

Method

Data Collection & Sampling

Craigslist offers a fitting opportunity to study location sharing preferences because it is pre-divided into topical and geographical posting categories and also allows users to consciously share location with varying precision. When creating a post, the default for location sharing is to show location on maps (Figure 1). Users must uncheck a box to omit location from maps, although zip code is always collected. Users must therefore make the effort to mask their location by not showing it on maps, or must deliberately choose to share additional location information beyond zip code, as shown in Figure 1. Another advantage of studying Craigslist is that the user base is likely to include individuals with more concern about privacy, compared to other forms of social media. The anonymous nature of Craigslist, compared to other platforms which involve names and photos, may encourage participation of those who have greater concern for privacy.



The image shows a screenshot of the Craigslist location sharing interface. At the top, there is a checkbox labeled "show on maps" which is checked. Below this, there are three input fields: "street" (optional), "cross street" (optional), and "city". The "city" field is followed by a small empty box, likely for a zip code.

Figure 1: Default checked map sharing option for Craigslist posts.

California is selected as the study area because of its great variation in Internet use between rural and urban areas (File and Ryan, 2014) and its strong tradition of privacy. An inalienable right to privacy is included in the state constitution, and California recently passed a sweeping law protecting digital privacy rights, including mobile location, from government surveillance. This makes California a particularly interesting region in which to study geoprivacy. The great variation between populations within the state suggests that the results may be applied elsewhere in the country.

Previous human behavior studies have randomly sampled Craigslist posts according to state, city, and posting category within a selected time period (Bevan et al., 2016; Grov, 2010). The present study adopts the procedure of previous work in sampling 12 Craigslist-defined regions from the 29 listed in the state of California to include a range of population densities. The final 12 are: Bakersfield, Inland Empire, Los Angeles, Mendocino, Merced, Orange County, Sacramento, San Diego, SF Bay, Siskiyou, San Luis Obispo, and Yuba-Sutter. In June 2015, data were sampled for the up to the most recent 100 posts in each selected category (baby, electronics, free, jewelry, tickets, garage sale, apartments, and rooms) for each of the 12 regions, comprising 6,805 posts between 30 April 2016 and 15 June 2016. The limit of 100 posts was set to prevent over-representation of posts from highly populated regions. No attempt was made to identify or remove scam posts, so data may include spam. Collected data include only the product category, metropolitan area, and precision level of the location provided. No personally identifying information was recorded from the site.

Analysis

For the categorical variables collected in this study, contingency tables are used to compare group counts, and chi-squared analysis is applied to test for significant differences between categories. This is appropriate given the non-parametric and categorical nature of the data. Next, this study focuses on the correlates of more precise location sharing and location withholding. In general, it is assumed that the goal of moving a point on the map would be to increase the accuracy of the location shared; however, it cannot be ruled out that users may move the point to hide their true locations. Therefore, location precision is not considered a truly ordinal variable in this study, and binomial logistic regression is instead used to model location sharing at a binary level. The regression is tested on 3 outcome variables: omitted map location, adjusted map point, and location provided with precision above zip code. Regression models tested that returned the lowest AIC are presented in the results.

Results

Of the 6,805 initial posts sampled, there was a generally even split across most regions and sale categories, with the notable exceptions of Siskiyou for regions and garage sale for categories (Figure 2). This is due to lower population in the former and lower popularity as a sale category in the latter. These groups are excluded from consideration in regression. The majority of posts (69%) are shared on maps at the zip code level or not shared on maps at all. This means that only 31% of users opted to provide any location information beyond zip code. Of those who opted to post a more precise location on maps, 11% moved the pin on the map to specify a location. Figure 3 displays the frequency of posts by precision level. There were only 2 cases of location shared with a valid street name but an invalid zip code, so these were excluded from the rest of the study for a final count of 6,803 posts.

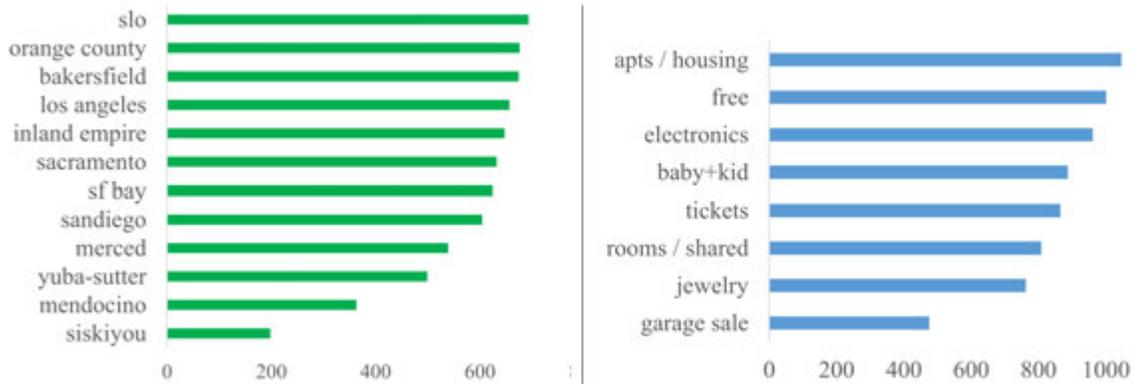


Figure 2: Frequency of sampled posts by region and category.

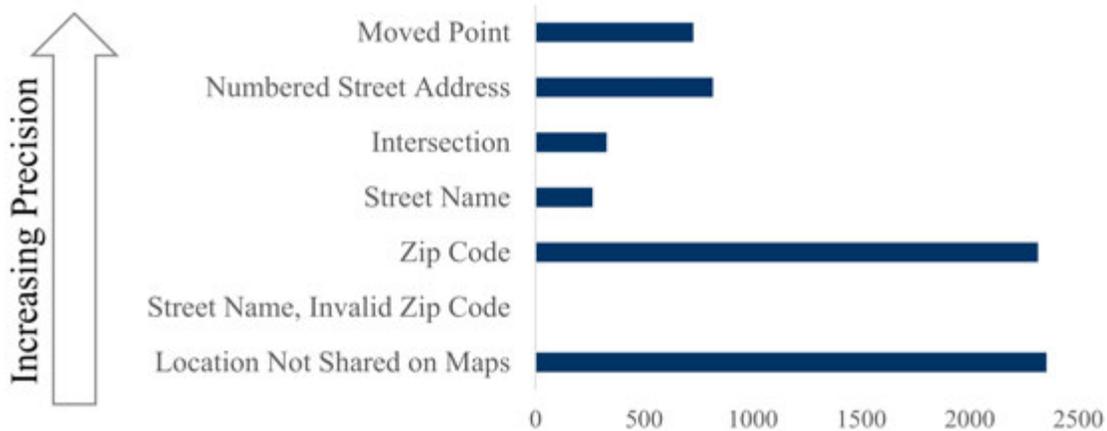


Figure 3: Frequency of sampled posts by geocoding precision level.

Chi-square analysis of contingency tables for posting category and region revealed significant differences between both sets of groups ($p < 0.00$). The results for category are shown in Table 1, while the results for region are depicted in Table 2. Contingency mosaics (Figure 4 and Figure 5) offer a color-coded glimpse at these group differences. As hypothesized, a higher percentage (at least 45% by category) of posts omitted map location in the baby, electronics, jewelry, and tickets categories. At least 44% percent also kept location off maps in the more rural regions of Merced and Yuba-Sutter, higher than for urban regions. Compared to the other categories, a higher percentage of posts included location information beyond zip code for apartments, garage sale, rooms, and free offers. These results demonstrate greater location sharing behavior in less sensitive categories and in more urban regions.

Table 1: Contingency table and chi-squared test for posting category

<i>Category</i>	<i>Point</i>	<i>Address</i>	<i>Intersection</i>	<i>Street</i>	<i>Zip</i>	<i>No Map</i>	<i>TOTAL</i>
apts / housing	158	351	63	50	239	185	1046
baby+kid	68	22	26	18	328	425	887
electronics	86	55	27	10	375	408	961
garage sale	66	149	38	109	92	19	473
jewelry	36	35	21	2	263	405	762
rooms / shared	128	76	79	40	277	208	808
tickets	49	36	17	10	368	385	865
free	136	93	56	24	373	319	1001
TOTAL	727	817	327	263	2315	2354	6803
Chi-squared: $X^2 = 1968.96$ df = 35 p < 0.00							

Table 2: Contingency table and chi-squared test for geographical region

<i>Region</i>	<i>Point</i>	<i>Address</i>	<i>Intersection</i>	<i>Street</i>	<i>Zip</i>	<i>No Map</i>	<i>TOTAL</i>
bakersfield	47	87	27	24	255	234	674
inlandempire	61	49	18	18	259	242	647
losangeles	101	85	58	23	194	195	656
mendocino	35	69	10	8	149	92	363
merced	44	41	13	19	158	264	539
orangecounty	96	67	20	49	241	201	674
sacramento	63	68	34	26	201	240	632
sandiego	55	95	42	22	238	152	604
sfbay	85	81	51	28	177	202	624
siskiyou	29	20	3	6	70	70	198
slo	79	102	24	22	226	240	693
yubasutter	32	53	27	18	147	222	499
TOTAL	727	817	327	263	2315	2354	6803
Chi-squared: $X^2 = 326.16$ df = 5 5p < 0.00							

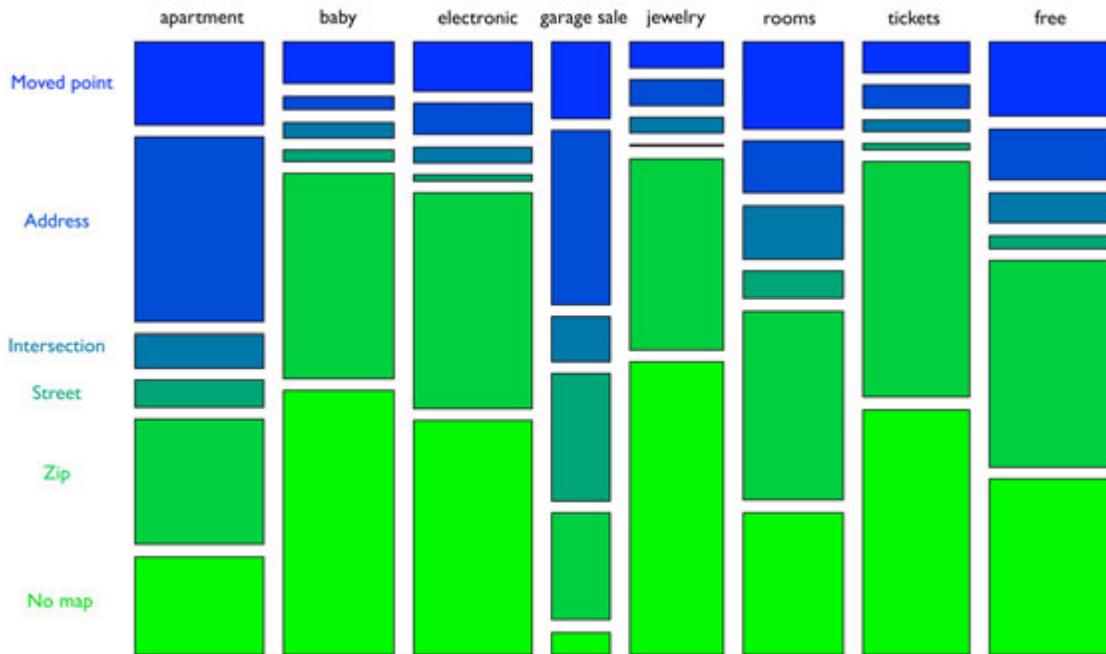


Figure 4: Contingency mosaic for location precision and posting category

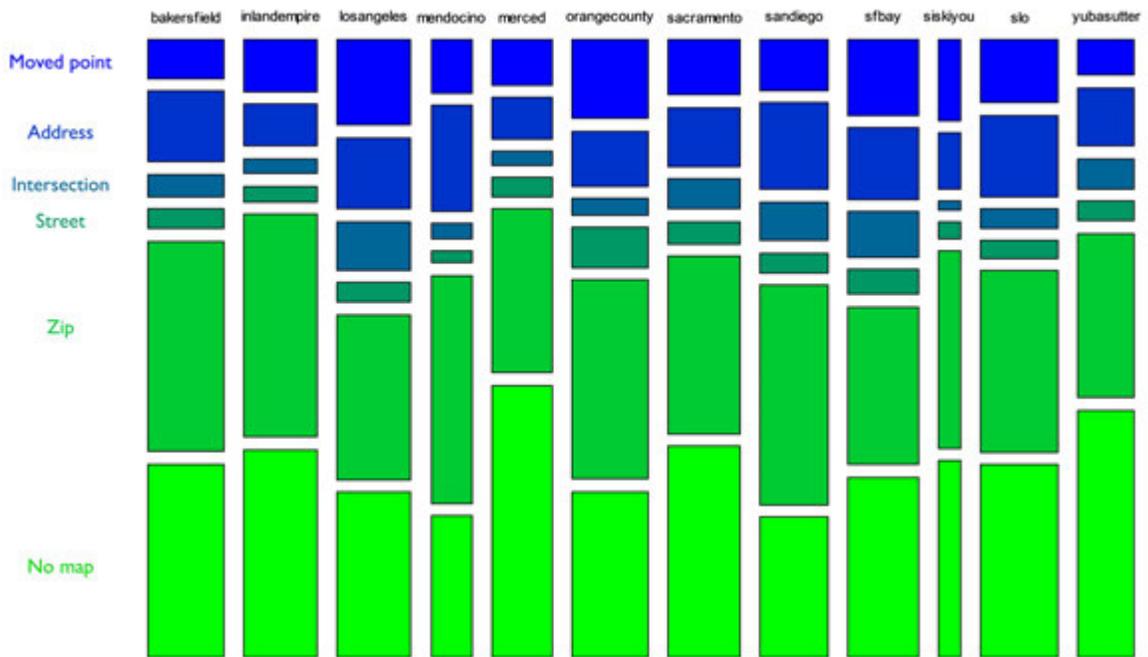


Figure 5: Contingency mosaic for location precision and geographical region

Figures 6 and 7 depict maps of two location sharing options of interest: omission of map location and adjustment of map point. These options are assumed to represent opposite ends of the location-sharing spectrum, as long as the map point is moved to improve the

accuracy of street address geocoding (rather than a displacement to an incorrect location). The dark green areas where the percentage of posts with omitted map location is higher are clearly in the inland, more rural locations (Yuba-Sutter, Merced, Inland Empire), while the coastal cities exhibit more location sharing (San Diego, Los Angeles, Orange County, Bay Area, and the outlier of Mendocino). A similar pattern is shown in Figure 7, where greater comfort with map point adjustment is shown in the coastal cities of the Bay Area (SF Bay), Los Angeles, and Orange County, with the outlier of Siskiyou.

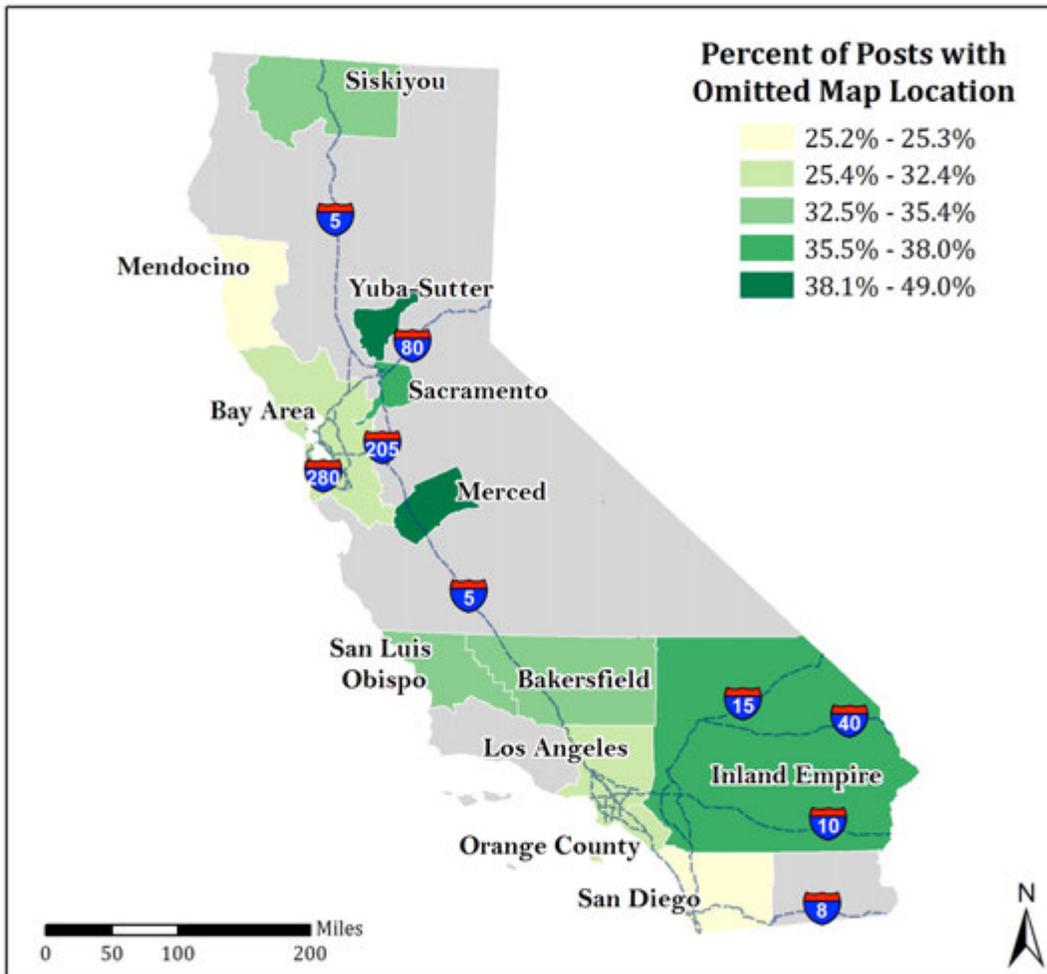


Figure 6: Percentage of posts with omitted map location by Craigslist region

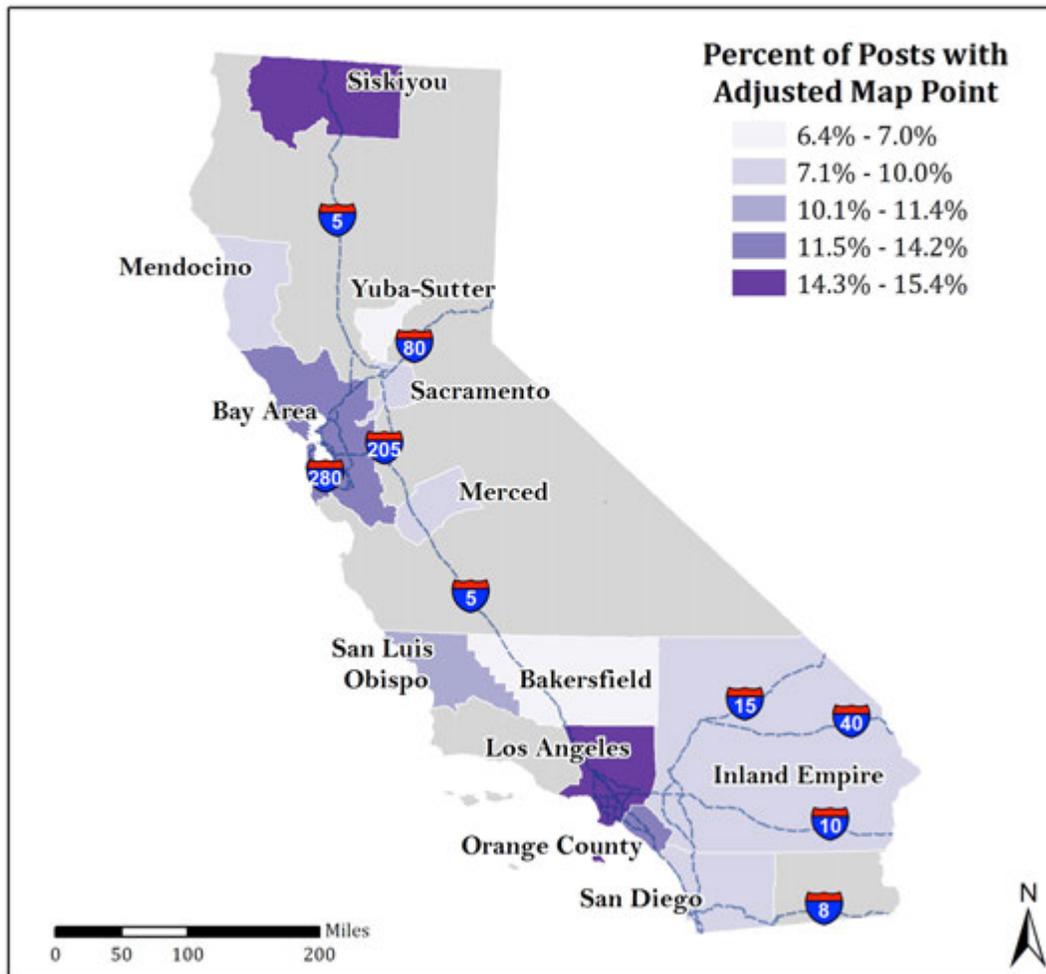


Figure 7: Percentage of posts with adjusted map point by Craigslist region

The binomial logistic regressions for omitted map location, adjusted map point, and location provided beyond zip code also yield significant results for all variables shown and confirmed both a rural-urban and sensitive information trend. These results are displayed in Table 3. The strongest positive predictors for omitted map location are baby+kid, jewelry, tickets, and electronics. Negative regional predictors include Los Angeles, Orange County, San Diego, and Mendocino.

Table 3: Binomial logistic regression results

Omitted Map Location	Estimate	Std. Error	z	p
<i>(Intercept)</i>	-3.08	0.24	-13.08	0.00
<i>baby+kid</i>	3.06	0.24	12.52	0.00
<i>apartments</i>	1.61	0.25	6.49	0.00
<i>free</i>	2.40	0.24	9.82	0.00
<i>jewelry</i>	3.26	0.25	13.26	0.00
<i>tickets</i>	2.95	0.24	12.06	0.00
<i>room</i>	2.07	0.25	8.36	0.00
<i>electronics</i>	2.84	0.24	11.66	0.00
<i>losangeles</i>	-0.20	0.10	-2.10	0.00
<i>orangecounty</i>	-0.25	0.09	-2.60	0.00
<i>merced</i>	0.52	0.10	5.36	0.00
<i>sandiego</i>	-0.53	0.10	-5.14	0.00
<i>mendocino</i>	-0.49	0.13	-3.79	0.00
AIC: 8028.3				
Adjusted Map Point	Estimate	Std. Error	z	p
<i>(Intercept)</i>	-2.01	0.07	-30.07	0.00
<i>baby+kid</i>	-0.52	0.14	-3.80	0.00
<i>apartments</i>	0.22	0.10	2.15	0.03
<i>jewelry</i>	-1.07	0.18	-5.95	0.00
<i>tickets</i>	-0.92	0.16	-5.85	0.00
<i>sfbay</i>	0.37	0.13	2.84	0.00
<i>losangeles</i>	0.50	0.12	4.04	0.00
<i>orangecounty</i>	0.42	0.12	3.34	0.00
<i>bakersfield</i>	-0.36	0.16	-2.24	0.02
<i>yubasutter</i>	-0.49	0.19	-2.58	0.01
AIC: 4489.0				
Beyond Zip Code	Estimate	Std. Error	z	p
<i>(Intercept)</i>	-0.82	0.06	-14.50	0.00
<i>baby+kid</i>	-1.22	0.10	-11.66	0.00
<i>apartments</i>	0.92	0.08	11.81	0.00
<i>free</i>	-0.34	0.08	-4.16	0.00
<i>jewelry</i>	-1.50	0.12	-12.60	0.00
<i>tickets</i>	-1.49	0.11	-13.44	0.00
<i>sfbay</i>	0.74	0.10	7.42	0.00
<i>losangeles</i>	0.80	0.10	8.31	0.00
<i>orangecounty</i>	0.51	0.10	5.17	0.00
<i>slo</i>	0.47	0.10	4.83	0.00
<i>sacramento</i>	0.42	0.10	4.07	0.00
<i>sandiego</i>	0.53	0.10	5.18	0.00
AIC: 7533.1				

For the outcome variable of map point adjustment, the negative predictors include baby+kid, jewelry, tickets, Bakersfield, and Yuba-Sutter. Posts from SF Bay, Los Angeles, and Orange County are significantly positively correlated with map point adjustment. As for providing a location beyond zip code, a similar trend is seen. The strongest negative predictors of this location sharing are baby+kid, jewelry, and tickets. The strongest positive predictors are apartments, Los Angeles, and SF Bay. The model with the lowest AIC value is for adjusted map point, suggesting that this outcome variable is most adequately predicted.

Conclusions

This study confirms that a difference in context promotes differences in location sharing when posting advertisements on Craigslist. Posts are more likely to have map location omitted when they portray more sensitive information, such as the presence of children in a household or expensive and valuable goods. Another consideration is that apartment and room advertisements may be more likely to share precise location because location is a necessary part of what is being advertised. Housing ads that share location immediately mean less work for those who are searching for housing to determine their level of interest. It may be that the benefit of quickly finding someone to rent an apartment outweighs privacy considerations when posting location.

An important contribution of this study is the finding that posts from more rural regions are more likely to have map location omitted than those originating in urban regions. Possible explanations include that Craigslist users in rural regions are more concerned with privacy or exhibit less trust in sharing location online. Another motivation may be the lower anonymity inherent to less densely populated regions. A point in a dense urban center is much less unique than a point on a rural farm with few neighbors. It is possible that concern for anonymity is reflected in posts of rural origin. A next step would be to test whether the urban-rural dichotomy in location sharing exists at a finer resolution than the regional characteristics depicted here.

The results of this study also shed light on patterns of map use and comfort with map technology. There is a clear geographical divide with coastal cities more likely to have posts with the map point adjusted. This may reflect greater familiarity with interactive mapping technology, compared to areas of lower internet and smartphone penetration. It remains to be tested whether adjusting this map point is used to point others to a more accurate location or to mask one's true location. For this reason, it may be misleading to present a map point adjustment as the highest precision level. It also may be the case that a Craigslist user does not add location to a map but instead includes location information in the content of the post that can be identifying. This is particularly true of housing-related posts, which may refer to building characteristics.

The finding that location sharing varies according to contextual sensitivity has implications for designers of location-based social networks (LBSN). It suggests, similar to the findings of Benisch et al. (2011), that users may want greater control over the level

of location they share with others and the option to fine-tune these options according to context. When given the opportunity to share location at six different levels on Craigslist, many users take advantage of the options, rather than simply submitting to the collection of GPS coordinates as a condition of use, as with other modern applications. Finally, while posting zip code on the map is the default setting for Craigslist, approximately one third of users uncheck the option, effectively masking their locations from other users. The motivations for this active location hiding merit further exploration.

References

- Benisch, M., Kelley, P.G., Sadeh, N. and Cranor, L.F. (2011) Capturing Location-Privacy Preferences: Quantifying Accuracy and User-Burden Tradeoffs. *Personal and Ubiquitous Computing*, 15, 7, pp. 679-694.
- Bevan, J.L., Galvan, J., Villasenor, J. and Henkin, J. (2016) “You've Been on My Mind Ever Since”: A Content Analysis of Expressions of Interpersonal Attraction in Craigslist.org's Missed Connections Posts. *Computers in Human Behavior*, 54, pp. 18-24.
- File, T. and Ryan, C. (2014) Computer and Internet Use in the United States: 2013. *American Community Survey Reports*.
<https://www.census.gov/history/pdf/2013computeruse.pdf> Last visited 7/15/2016.
- Grov, C. (2010) Risky Sex-And Drug-Seeking in a Probability Sample of Men-For-Men Online Bulletin Board Postings. *AIDS and Behavior*. 14, 6, pp. 1387-1392.
- Li, J., Zhong, C., Cao, Z., Pei, W. and Zhou, X. (2013) Examining Privacy Concern in Social-Driven Location Sharing: An Empirical Study on Chinese Popular SNSs. In *Computer Software and Applications Conference (COMPSAC), 2013 IEEE 37th Annual* (pp. 668-677). IEEE.
- Nissenbaum, H. (2009) *Privacy in Context: Technology, Policy, and the Integrity of Social Life*. Stanford University Press.
- Sun, Y., Wang, N., Shen, X.L. and Zhang, J.X. (2015) Location Information Disclosure in Location-Based Social Network Services: Privacy Calculus, Benefit Structure, and Gender Differences. *Computers in Human Behavior*, 52, pp. 278-292.

Dara E. Seidl, Department of Geography, San Diego State University – UC Santa Barbara, San Diego, CA , 92182

Christopher Allen, Department of Geography, San Diego State University – UC Santa Barbara, San Diego, CA , 92182