Craigslist Scams and Community Composition: Investigating Online Fraud Victimization

Vaibhav Garg
Drexel University

Shirin Nilizadeh
Indiana University

Abstract—Offline, crime and resulting victimization is not individual incidence. It is also hampered or encouraged by the community in which it is situated. Are community characteristics relevant for victimization online? This paper examines the prevalence of Craigslist-based (automobile) scams across 30 American cities. Our methodology analyses historical scam data and its relationship with economic, structural, and cultural characteristics of the communities that are exposed to fraudulent advertising. We find that Craigslist scams are not random, but targeted towards specific communities. The resulting policy insight is for creating public awareness campaigns addressing educated white males, as they are the most vulnerable.

Keywords-Craigslist; scam; policy; governance; cybercrime;

I. INTRODUCTION

Craigslist is an online platform for classified advertisements. Its impact on local economy is non-negligible. For example, such advertisements reduce housing rental vacancy rates [1]. The success of Craigslist is limited by persistent scams. The possibility of being scammed could prevent transactions from being completed. In economic terms, the expected value of loss due to a scam can be thought of as a probable tariff towards future transactions. Given the positive effects of Craigslist for the economy, local governments should aim to minimize this expected loss.

Thus, local governments should either decrease the probability of victimization or limit the magnitude of losses. The obvious solution is to decrease the number of attackers, for example through prosecution. Such deterrence is prohibitively expensive [2], especially considering the limited resources of local law enforcement and associated public bodies. (The difficulties encountered in the application of consumer fraud statutes online are well-documented [3].)

The alternative is to alleviate the number of attacks. Arguably, more scammers would be attracted to, and therefore post more fraudulent advertisements for a community, which in the past has more actively responded and been victimized. From an attacker perspective, posting the scam advertisement could be done through automated agents and thus requires minimum investment. However, follow up email exchanges require human engagement and thus become exponentially expensive [4]. If these follow up emails do not result in actual victimization for certain neighborhoods, these scammers would rationally choose not to participate.

There is limited understanding of online fraud victimization. Investigations have concentrated on individual behavioral explanations [5], [6]. However, victimization is a function of both individual vulnerability as well as the frequency of exposure to crime. Offline, criminological theories attribute this exposure to neighborhood characteristics, e.g. routine activity [7]. These theories are rarely examined online. Available studies use survey based methodologies, the limitations of which are well documented [8].

A notable exception is Hann et al.'s investigation of past spam data, which indicated that spam is not random but targeted [9]. (Their study had limitations as the email accounts were created only for the purpose of research, did not make online transactions, and spam received was unsolicited.) However, there is no similar examination of (Craigslist-based) scams, especially considering community characteristics. This examination is much needed for a comprehensive public policy and local governance response to online fraud victimization [10].

This paper presents a small sample study that examines whether economic, structural, and cultural characteristics of a community explain the incidence of Craigslist-based scams. We present an empirical investigation with automobiles scams, an example of advance fee fraud. Section II discusses the relevant criminological theories, and the various factors therein, which have been found relevant for victimization offline. Section III presents our methodology, specifically the operationalization of key variables and respective data sources. In section IV we present the results. Section V discusses the implications of our results and provides insights for public policy. Section VI concludes.

II. BACKGROUND & RELATED WORK

Current cybercrime research considers victimization as an individual behavioral problem [5], [6]. However, both exposure to crime and resulting victimization are a function of community composition. For example, social disorganization theory, dating to the early 1900s, investigates the concentration of crime in specific neighborhoods. It analyzed slum neighborhoods, where the level of criminal activity stayed constant irrespective of the resident ethnic group [11]. This theory indicates that neighborhoods with more homogeneous populations are less likely to be subjected

to crime [12], as the collective efficacy of informal social controls through unsupervised peer groups is alleviated by heterogeneity [13]. A potential hypothesis, for American cities, would be that cities that are predominantly white would be exposed to less Internet-based scams.

A second criminological theory, routine activity theory, considers crime to be a function of attacker motivation, availability of targets, and absence of guardianship [7]. For younger and more educated individuals online purchases are more of a routine activity and thus correlates with higher levels of fraud [7]. Simultaneously, those on the extreme ends of the educational spectrum, i.e. those without a high school degree as well as those with a graduate education, are less likely to be approached by fraudsters [14]. Then a lower percentage of high school graduates should reflect a less number of Craigslist scams. Simultaneously, a higher percentage of individuals with a bachelors degree should also be negatively correlated with the number of scam postings on Craigslist.

Estimates of income and poverty are also relevant to victimization [15]. Offline, high income neighborhoods would be more targeted as the risk of crime, such as burglary, would have a higher payoff. Similarly, online cities with higher per capita income are likely to be targeted.

Gender composition of the city may also be relevant, as women are more risk averse than men [16]. Thus, they are less likely to fall for scams. Simultaneously, women perceive higher risk online purchases and are less satisfied with e-commerce [17]. Thus, women are less likely to respond to Craigslist based automobile scams [18]. A potential hypothesis is that a higher population percentage of women would lead to fewer scams being posted, as the probability of victimization is lower.

III. Methodology

Craigslist automobile scams are a typical instantiation of advance fee frauds. A possible scam in Craigslist's automobile section would include a scammer who posts lucrative advertisements for automobiles. The quoted price for such vehicles is irrationally low; arguably setting an unrealistically low price allows the scammer to filter out the individuals that are unlikely to be adequately gullible [4]. Despite the extant signal of an irrationally low price the scammer would receive responses from potential victims.

The scammers would then attempt to con these individuals with stories that would explain the reason for the unreasonably low price. For example, the seller might pretend to be an international traveller who inexplicably has to move back to his or her home country on short notice. Transactions are carried out using third-party agents, such as BidPay, Squaretrade, or PayPal. Victims advance the money to the scammers without getting the vehicle in return.

Our unit of analysis is limited to a city. In Craigslist the U.S. classifieds are listed under 413 cities. Of these, we focused on the 30 largest metropolitan areas as identified by Craigslist: Atlanta, Austin, Boston, Chicago, Cleveland, Dallas, Denver, Detroit, Honolulu, Houston, Kansas City, Las Vegas, Los Angeles, Miami, Minneapolis, Nashville, New York, Orange County, Philadelphia, Phoenix, Portland, Raleigh, Sacramento, San Diego, Seattle, San Francisco, St. Louis, Tampa, and Washington DC. We assume that the advertisements posted on each city's Craigslist website are primarily targeted towards individuals living in those cities. We do *not* assume that the scammers are from the same jurisdictional region as the city for which they post their respective advertisements.

Table I
MACRO-LEVEL FACTORS AND THEIR SOURCES

Factors	Source	Year
Population	CB	2010
Population % of Women	CB	2010
Population % of Majority Ethnicity (White)	CB	2010
High school graduates, % of persons age 25+	CB	2006-2010
Per capita income in past 12 months (2010 $)	CB	2006-2010
Persons below poverty level, %	CB	2006-2010
CB=Census Bureau; FBI= FBI Crime Database		

We concentrated on the *cars+trucks* section of Craigslist, which publishes automobile classifieds under two categories: *by-owner* and *by-dealer*. We examined the by-owner classifieds, leaving the investigation of the by-dealer section to future work. Data collection started before the Thanksgiving break on 11/19/2010 and continued for 60 days. Our crawler collected advertisements that were posted earlier. Thus, the dataset constitutes advertisements for a period of 3.5 months. Craigslist has a user-flagging system to identify illegal and inappropriate postings. When a certain number of users flag a posting, it is replaced by the *flagged for removal* message and finally after some days it is removed. We considered *flagged* ads to be scam and spam ads. However, an advertisement may be flagged for other reasons. For example, the advertisement may be placed in the wrong section on Craigslist. Thus, our data is noisy. We assume that this noise is uniformly distributed across the dataset. Note that flagged data does not include all scams, as some scams may not have been flagged by the requisite number of people to be flagged for removal. However, we assume that our data represents the relative distribution of scams on Craigslist. The total number of unique advertisements observed was 2,424,092, of which 42,185 were flagged.

After collecting data we classified the flagged ads based on the city where they were posted and then derived the number of total ads and flagged ads per city. Thus, we derived two dependent variables: 1) total number of flagged advertisements, and 2) percentage of flagged advertisements.

We had two primary sources for the independent variables: 1) the U.S. Census Bureau and 2) FBI Crime Data. A list

of variables that we considered, their sources, and year is given in Table I. These variables were considered based on the research discussed in section II[1]. The final regression equations are given by 1 and 2, where N and P correspond to the number and percentage of flagged advertisements respectively.

$$N = \epsilon_N + \beta_{N_1} * Population + \beta_{N_2} * Women + \beta_{N_3} * White$$
$$+ \beta_{N_4} * Education + \beta_{N_5} * Income + \beta_{N_6} * Poverty$$
(1)

$$P = \epsilon_P + \beta_{P_1} * Population + \beta_{P_2} * Women + \beta_{P_3} * White$$
$$+ \beta_{P_4} * Education + \beta_{P_5} * Income + \beta_{P_6} * Poverty$$
(2)

IV. RESULTS

We began by normalizing the dependent variables. Specifically, we shifted the mean to zero but subtracting it from the data. We also scaled the spread by the standard deviation, ie. divided by the s.d. The sample size of the data set under analysis is relatively small, n=30. Thus, we also checked to see if the dependent variables were normally distributed. We used the Shaprio-Wilkie test. Null hypothesis, i.e. that the variable is normally distributed, could not be rejected for either the number of flagged ads, p-value=0.21, or the percentage of flagged advertisements, p-value=0.07. We also checked to see if the residuals of the regression models given by equations 1 and 2 were normally distributed as a check for heteroskadasticity. Null hypothesis could not be rejected; the p-values were 0.19 and 0.89 respectively.

Table II
CORRELATION

Variable	N	P
Population	0.35	-0.17
Population % of Women	-0.48**	-0.20
Population % of Majority Ethnicity (White)	0.28	0.50**
High school graduates, % of persons age 25+	-0.34	0.07
Per capita income	0.24	0.02
Persons below poverty level, %	-0.07	0.03
0.5<*<0.01<**<0.001<***< ≈0		

Pearson's correlation coefficients between the independent variables and the two dependent variables is given in table II. Ordinary Least Squares provided estimates for equations 1 and 2. The results are given in tables III and IV.

V. DISCUSSION

Population is not significantly correlated with either with the number or percentage of flagged advertisements. It is also not a significant dimension in either of the regression models. Thus, merely having a lager pool of potential victims does not correlate with more scams.

[1]Not all variables were considered due to the presence of multicollinearity, which was assumed if Variance Inflation Factor was greater than 5.

Table III
OLS REGRESSION: NUMBER OF FLAGGED POSTS

Variable	Estimate	Std. Error
(Intercept)	1425.03***	79.40
Population	186.30	99.84
" % of women	-306.74*	114.48
" % of Maj. Ethnicity (White)	91.10	97.89
High School Grad., % of persons age 25+	-373.66*	136.47
Per capita income	449.69**	131.36
Persons below poverty level, %	231.70	141.18
0.5<*<0.01<**<0.001<***< ≈0		

Residual standard error: 427.6 on 22 degrees of freedom
Multiple R-squared: 0.6786, Adjusted R-squared: 0.5909
F-statistic: 7.74 on 6 and 22 DF, p-value: 0.0001471

Table IV
OLS REGRESSION: % OF FLAGGED POSTS

Variable	Estimate	Std. Error
(Intercept)	1.642e-02***	6.211e-04
Population	-1.563e-04	7.810e-04
" % of women	-5.671e-05	8.955e-04
" % of Maj. Ethnicity (White)	2.155e-03*	7.657e-04
High School Grad., % of persons age 25+	5.242e-04	1.068e-03
Per capita income	3.251e-04	1.028e-03
Persons below poverty level, %	1.474e-03	1.104e-03
0.5<*<0.01<**<0.001<***< ≈0		

Residual standard error: 0.003344 on 22 degrees of freedom
Multiple R-squared: 0.336, Adjusted R-squared: 0.1549
F-statistic: 1.856 on 6 and 22 DF, p-value: 0.1344

Population percentage of women is significantly and negatively correlated with the number of flagged advertisement. While the relationship with percentage of flagged advertisements is not significant the correlation is still negative. Gender composition of the population is also significant in the regression model for the number of flagged advertisements. The negative sign of the estimate indicates an inverse relationship. Simultaneously, the magnitude of the estimate is more than twice of the standard error, indicating that this inverse relationship between higher proportion of women and lower exposure to scams, persists. Thus, there are evident gender differences in exposure to victimization.

Counterintuitively, cities with more racially homogenous population saw more number of scams. The correlation with number of flagged advertisements is not statistically significant, but is positive. The correlation with percentage of flagged advertisements is, however, both positive and highly statistically significant. Racial composition is also the only statistically significant dimension in the regression model for % of flagged advertisements; racial homogeneity was positively related to percentage of scams postings. Here again the magnitude of the estimate is much greater than that of the standard error, suggesting that the relationship persists. Thus, unlike physical crime exposure to online fraud may be directly proportional to racial homogeneity.

The % of high school graduates is significantly and nega-

tively correlated with the number of flagged advertisements. This factor is also statistically significant in the respective regression model. The estimate is negative, indicating an inverse relationship with exposure to scam. The magnitude of the estimate is significantly higher than that of the standard error; so this inverse relationship likely persists. This estimate magnitude is also larger than for % of women, indicating that %of high school graduates have a stronger influence on exposure to scam. This finding also indicates a difference between online and offline fraud. Offline fraudsters would prefer to approach individuals with high school diplomas [14]; online this relationship is inverted [7].

Intuitively, per capita income is significantly and positively correlated with the exposure to online scam. It seems reasonable that scammers would approach individuals with enough expendable income. However, number of individuals below poverty level did not have a measurable impact on exposure to scam, indicating that the distribution of income or income inequality is not relevant.

VI. Conclusion & Future Work

We find that *Craigslist (automobile) scams are targeted* and influenced by community characteristics and composition. Communities with higher proportion of *educated white males* specifically are most exposed to online fraud. The likely explanation is purchasing behavior. Local government initiatives, such as public awareness campaigns and education efforts should take this vulnerability into account.

Our results are limited by the nature of the online scam studied. Individuals that buy automobiles online may not be representative of other fraudulent transactions enabled by Craigslist scams. This study is also limited to 30 American cities. Thus, the results are likely not generalizable.

The obvious first step for future research is to analyze a larger set of scams over a broader range of cities ideally over different advertising platforms. For example, Craigslist is popular in United States but other platforms would be more relevant in European countries.

Acknowledgments

References

[1] K. Kroft and D. G. Pope, "Does online search crowd out traditional search and improve matching efficiency? evidence from craigslist," August 2010, available at: http://www.sole-jole.org/12104.pdf.

[2] R. Anderson, C. Barton, R. Böhme, R. Clayton, M. van Eeten, M. Levi, T. Moore, and S. Savage, "Measuring the cost of cybercrime," in *Proceedings of the Workshop on the Economics of Information Security*. Springer, 2012.

[3] D. S. Wall, "Policing cybercrimes: Situating the public police in networks of security within cyberspace," *Police Practice and Research*, vol. 8, no. 2, pp. 183–205, 2007.

[4] C. Herley, "Why do nigerian scammers say they are from nigeria?" in *Proceedings of the Workshop on the Economics of Information Security*. Springer, 2012.

[5] A. Rege, "What's love got to do with it? Exploring online dating scams and identity fraud," *International Journal of Cyber Criminology*, vol. 3, no. 2, pp. 494–512, 2009.

[6] F. Stajano and P. Wilson, "Understanding scam victims: seven principles for systems security," *Communications of the ACM*, vol. 54, no. 3, pp. 70–75, 2011.

[7] T. C. Pratt, K. Holtfreter, and M. D. Reisig, "Routine online activity and internet fraud targeting: Extending the generality of routine activity theory," *Journal of Research in Crime and Delinquency*, vol. 47, no. 3, pp. 267–296, 2010.

[8] J. Junger-Tas and I. H. Marshall, "The self-report methodology in crime research," *Crime and justice*, pp. 291–367, 1999.

[9] I.-H. Hann, K.-L. Hui, and Y.-L. Lai, "Who gets spammed?" *Communications of the ACM*, vol. 49, no. 10, pp. 83–87, 2006.

[10] H. N. Pontell, "Identity theft: Bounded rationality, research, and policy*," *Criminology & Public Policy*, vol. 8, no. 2, pp. 263–270, 2009.

[11] C. Shaw and H. McKay, "Juvenile delinquency and urban areas," *Chicago, Ill*, 1942.

[12] R. Sampson and W. Groves, "Community structure and crime: Testing social-disorganization theory," *American journal of Sociology*, pp. 774–802, 1989.

[13] R. Sampson, S. Raudenbush, and F. Earls, "Neighborhoods and violent crime: A multilevel study of collective efficacy," *Science*, vol. 277, no. 5328, p. 918, 1997.

[14] R. Titus, F. Heinzelmann, and J. Boyle, "Victimization of persons by fraud," *Crime & Delinquency*, vol. 41, no. 1, pp. 54–72, 1995.

[15] R. J. Sampson, "Neighborhood and crime: The structural determinants of personal victimization," *Journal of Research in Crime and Delinquency*, vol. 22, no. 1, pp. 7–40, 1985.

[16] J. Byrnes, D. Miller, and W. Schafer, "Gender differences in risk taking: A meta-analysis." *Psychological bulletin*, vol. 125, no. 3, p. 367, 1999.

[17] S. Rodgers and M. A. Harris, "Gender and e-commerce: an exploratory study," *Journal of Advertising Research*, vol. 43, no. 3, pp. 322–329, 2003.

[18] E. Garbarino and M. Strahilevitz, "Gender differences in the perceived risk of buying online and the effects of receiving a site recommendation," *Journal of Business Research*, vol. 57, no. 7, pp. 768–775, 2004.