Supplemental Material. Towards Safe Cities: A Mobile and Social Networking Approach

Jaime Ballesteros, Bogdan Carbunar, Mahmudur Rahman, Naphtali Rishe, S.S. Iyengar

I. Geosocial Network Data

Figure 1(a) shows the distribution of the per-venue number of reviews of Miami-Dade venues, with a logarithmic y scale. It shows a long tail distribution, with around 2000 venues having 1 review but only 1000 venues having 2 reviews. We emphasize the low number of venues without reviews - only 177. Figure 1(b) shows the distribution of the number of venues with an aggregated rating ranging between 1 and 5: Yelp reviews are mostly positive as most aggregate ratings are at or above 4 stars.

II. Crime Data

We use a historical database of more than 2.3 million crime incidents reported in the Miami Dade county area since 2007 [1]. Each record is labeled with a crime type (e.g., homicide, larceny, robbery, etc), the time and the geographic location where it has occurred. We briefly document two problems we encountered when pre-processing this data. First, since records come from different Police departments, the crime type labels are non-uniform, (e.g., *murder* in Miami Beach vs. *homicide* in North Miami). Second, crime reports include many minor incidents (e.g., fire alarms issues), resulting in over 140 different crime types.

In order to standardize and eliminate ambiguities, we mapped crimes into 7 categories: Murder, Forcible Rape, Aggravated Assault, Robbery, Larceny/Theft, Burglary/Arson, Motor Vehicle Theft. We removed minor crime reports that did not fall into these categories. Due to the large number of records in the database, manual mapping was infeasible. Instead, we have experimented with two machine learning techniques for classifying each record: the Naive-Bayes (NB) classifier and the Decision Trees (DT) classifier [2]. In order to build our training and test sets, we manually annotated a random sample of 2000 records from different police departments. Then, we split this subset of records into training and test datasets, each containing 1000 records. We built our classifiers using the NLTK library [3]. The accuracy was measured using a simple metric that measures the percentage of inputs in the test set that the classifier correctly labeled. For instance, a crime type classifier that predicts the correct crime type 60 times in a test dataset containing 100 crime types, would have an accuracy of 60%. On our crime dataset, the NB classifier achieved an accuracy of 91% and the DT classifier an accuracy of 98%. Thus, we have used the outcome of the DT classifier. Figure 2 shows the crime set's distribution of the crime categories following the DT classification.

The authors are with the School of Computing and Information Sciences at the Florida International University, Miami, FL, USA. E-mail: {jball008,carbunar,mrahm004,,rishen,iyengar}@cs.fiu.edu.

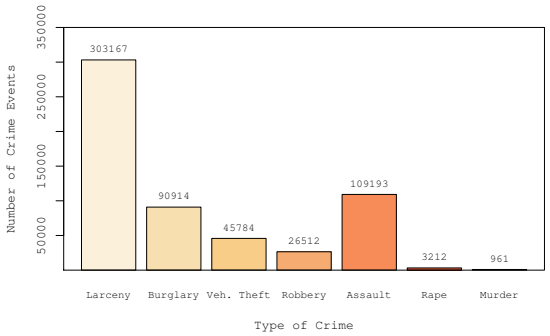


Fig. 2. Outcome of DT classifier – statistics of crime in Miami-Dade county. Distribution of number of crime events per type of crime.

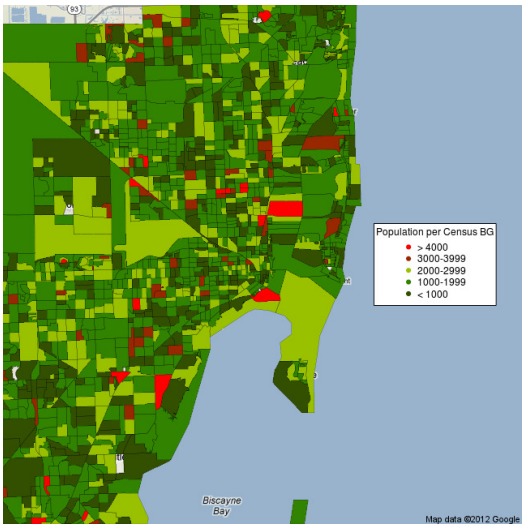
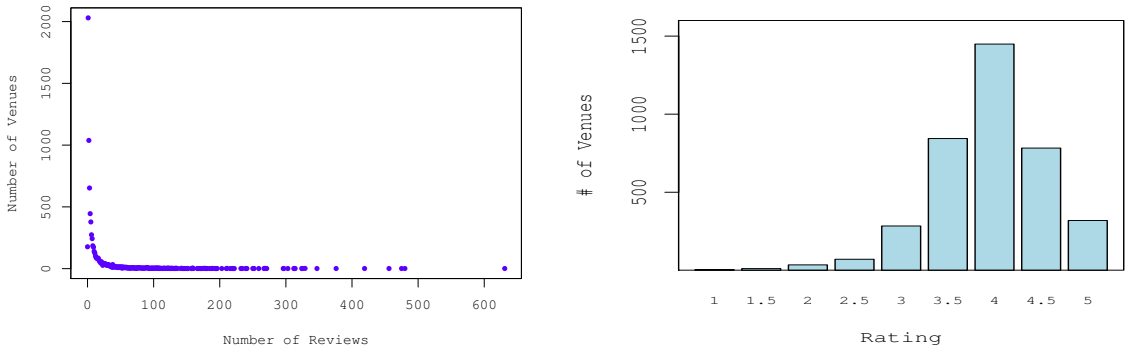


Fig. 3. Miami-Dade county: geographical distribution of population. Polygons represent Census Block Groups.

We use Census data sets [4], reporting population counts and demographic information. The data is divided into geographical extents e.g. polygons, called *census block groups*. Each block contains information about the population within (e.g., population count, various statistics). According to the data, Miami Dade county has a population of $2,496,435$. Figure 3 shows the geographical distribution of the population in the Miami Dade county.



(a) (b)

Fig. 1. Statistics of Miami venues: (a) Distribution of number of reviews per venue. (b) Distribution of venue ratings. Venues recording less than 4 reviews were filtered out.

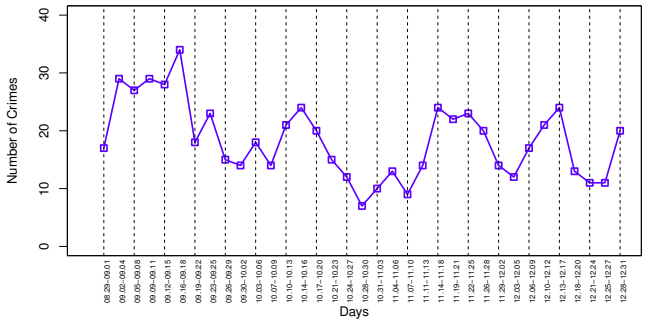


Fig. 4. 18 week (August 29-December 31, 2011) evolution of the number of crimes reported within one Miami-Dade block.

III. Location Based Safety

Figure 4 shows the evolution of the number of crimes reported in a Miami-Dade block during an 18 week interval (August 29-December 31, 2011). The dotted lines contain points representing the number of crimes reported during the work days (Monday-Thursday), The other points denote crimes reported during the week-end (Friday-Sunday). It shows that the number of crimes can evolve significantly between consecutive groups of days.

IV. Forecasting Tools

ARIMA Model. ARIMA(p, d, q) incorporates autoregressive (p), integration (d) and moving average terms (q) to provide higher fitting and forecasting accuracy. It uses the input data (p, d, and q) to determine the appropriate model form. The ARIMA forecasting procedure consists of four steps [5], (1) identifying the ARIMA(p, d, q) structure, (2) estimating the unknown parameters, (3) fitting tests on the estimated residuals and (4) forecasting future outcomes based on historical data.

The formulation of the ARIMA model depends on the characteristics of the series. Generally, it is originated from the autoregressive model AR (p), the moving average model MA (q) and the combination of AR (p) and MA (q), the ARMA (p, q) model [6]. Like most time series, ours is non-stationary. Hence we cannot apply stationary ARIMA processes directly. One way of handling non-stationary series is to apply *differencing* (d) so as to make them stationary. Then, to find the best ARIMA model, we used the autocorrelation (ACF) and partial autocorrelation (PACF) functions for preliminary estimations of the AR(p) and MA(q) components. The ACF function is a set of correlation coefficients between the series and lags of itself over time while the PACF function is the partial correlation coefficients between the series and lags of itself over time. We use the Corrected Akaike Information Criterion (AICc) [5] as the primary criterion in selecting the orders of a fitted ARIMA model which acts as an estimator of the expected discrepancy between the true model and a fitted candidate model. We choose the ARIMA model that has the minimum AICs value. We use T-statistics with 95% confidence interval to test the significance of the parameters in the fitted ARIMA model.

Linear (Double) Exponential Smoothing (LES) Model. Brown's linear (double) exponential smoothing [7] includes trend variations of the time series without a significant seasonal component. The process is controlled by a smoothing parameter α whose value ranges between 0 and 1. α decides the weight placed on the most recent observations during the forecast process. We determine the value of α by minimizing the root mean squared error (RMSE) [8] from one step-ahead forecasts and repeating the process for all forecast values.

Artificial Neural Network (ANN). ANNs are data-driven self-adaptive methods that learn and generalize from experience and capture subtle functional relationships among the empirical data even if the inherent relationships are unknown or difficult to describe. In this paper we focus on the multilayer perceptrons (MLP) ANN model, which is particularly suitable for forecasting, due to its ability for input-output mapping. The ANN we consider consists of an input layer (of the same size as the input vector), two layers of hidden nodes and an output layer providing the forecast value. Before the training phase, we normalize the input data to a $(-1, 1)$

range; following the prediction step we map the output back to the initial range. For the training phase we use a multilayer feedforward network trained using back propagation and the Levenberg-Marquardt algorithm to perform function fitting (nonlinear regression).

V. iSafe

We now present the proofs of theorems that we omitted from the main document.

Theorem 1: An adversary $\mathcal{A}$ controlling $k - a$ out of k participants in the iSafe algorithm, can only find the sum of the input values (BWC or Tblk) of the remaining a honest participants.

Proof: Secret splitting is information theoretical secure: Without knowing *all* the shares of a secret, no information can be inferred about the secret. The adversary $\mathcal{A}$ has access to all intermediate values multicast in Algorithm 1, as well as $k - h$ shares of the secret of each of the remaining h honest participants. Let R_i denotes the random value of the i-th (honest) participant and let $s_{1i}, s_{2i}, .., s_{ki}$ be the shares received by that participant from all the other participants. Then, the sum $R_i + s_{1i} + s_{2i} + .. + s_{ki}$ is random and cannot be predicted by $\mathcal{A}$: $\mathcal{A}$ only controls $k - h$ shares of R_i (out of $k - 1$ shares), but not R_i. Thus, the remaining h values in the sum are random and not under the control of $\mathcal{A}$. Thus, $\mathcal{A}$ cannot infer the value (BWC or TBlk) of user i by comparing the value of S before and after user i's multicast. ■

Theorem 2: iSafe provides location privacy.

Proof: (Summary) The adversary $\mathcal{A}$ can only access user location information from (i) user trajectory traces, (ii) queries made by iSafe (Algorithm 1 line 12) and (iii) during computations of the aggregate super user crime and safety indexes (the $multiPartySum$ operation).

For the first point, we observe that user trajectories are only stored on the the user's mobile devices and are never shared with other participants. For the second point, the queries made by iSafe clients to $\mathcal{A}$ are private, e.g., through the use of PIR. Thus, $\mathcal{A}$ cannot learn the location of the user with a probability non-negligible higher than $1/n$, where n is the number of census blocks, without breaking the security of the PIR solution employed. The third point's implicit requirement is that the provider colludes with users in order to learn information about their neighbors. The use of random, frequently changing MAC (or physical device) addresses by participating devices prevents however even such a powerful adversary from linking a device identifier to a user, thus linking a user to a location. Moreover, Theorem 1 shows that if $\mathcal{A}$ controls at most $NThr - h$ clients at any location where at least $NThr + 1$ clients are located, $\mathcal{A}$ can only learn the sum of the secret values of the remaining (at least h+1, $h > 1$) honest clients. ■

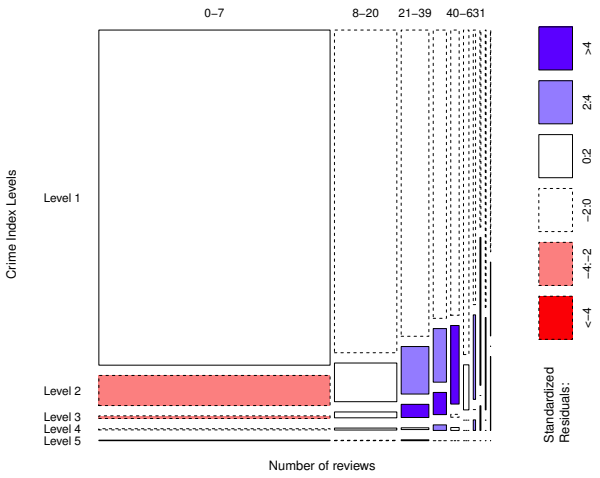


Fig. 5. Mosaic plot showing the relation between the number of reviews received by a venue and the crime index (CI) level of its block.

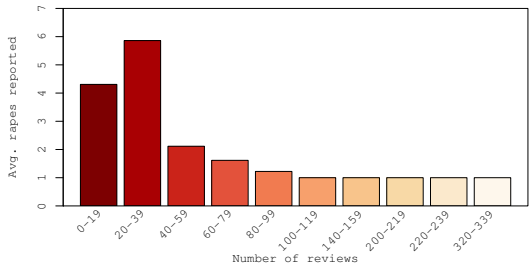


Fig. 6. Number of rapes per number of venue's reviews. Locals and visitors.

VI. Geosocial Network Extensions

In the supplemental material we investigate a second question, of whether there exists a relation between the number of reviews a venue receives and the safety of the venue's location. Once again, even though the number of reviews is not a categorical variable, it is discrete. Therefore, we tested their association with CI values using the χ^2 test. We created review count interval buckets and we assigned each venue to one bucket according to its number of reviews. We computed the range of the intervals using the 1-dimensional k-means algorithm with k set to 10. The χ^2 test produced a corresponding p-value very close to zero, thus answering our question in the affirmative. Figure 5 shows the corresponding mosaic plot of this experiment. It confirms that most Yelp venues are located in safe areas.

In order to identify the sources of the dependencies, we studied a specialized view of this data - the relationship between review counts and crime types. One finding is depicted in Figure 6, showing the relationship between reported rapes and review counts: rapes occur more frequently in places with low number of reviews. Furthermore, we study the relation

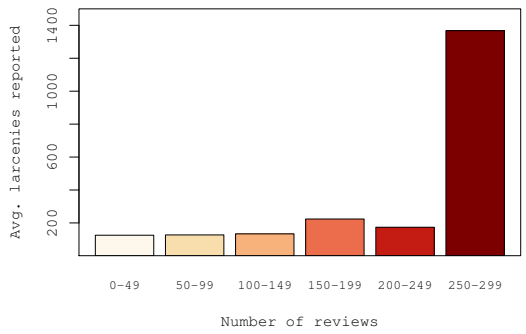


Fig. 7. Number of larcenies/thefts per number of venue's reviews.

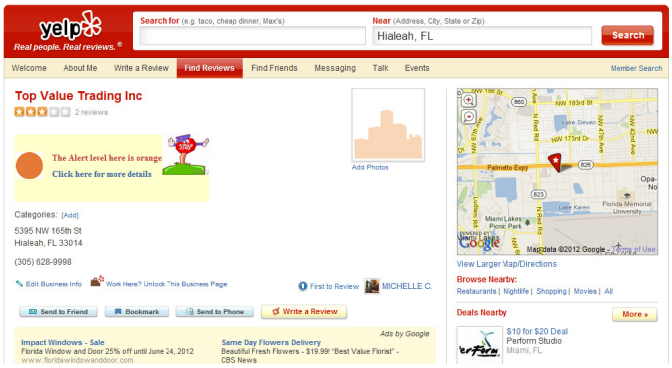


Fig. 8. Snapshot of iSafe's plugin functionality for a Yelp venue. The orange circle indicates the venue's safety level.

between crime types and the number of reviews received from visitors vs. locals. This information is publicly available, as Yelp users need to specify a home city/state. Figure 7 shows that the number of larcenies is high around venues with many local reviews.

VII. Attacks and Defenses

Safety profiles of co-located users are aggregated to compute the safety image of locations. Since that image impacts user decisions, it can become the target of malicious attacks. For instance, malicious users may attempt to incorrectly (i) improve the safety of desired locations, for instance to attract unsuspecting users to unsafe locations or to (ii) decrease the safety image of target locations. We now describe several mechanisms that could be exploited to perform these attacks, and suggest defenses.

Reporting incorrect locations. Malicious users may report incorrect locations, corresponding to safe areas. Even with GPS verification mechanisms in place, committing location fraud has been largely simplified by the recent emergence of specialized applications for the most popular mobile ecosystems (LocationSpoofer [9] for iPhone and GPSCheat [10] for Android). To prevent this attack, location verification mechanisms can be used [11], [12], [13]. For instance, in previous work [11], one of the authors has developed venue-centric location verification techniques, that rely on devices installed by venue owners within their venues. In the scenario considered in this paper, the owners' incentive for participation is to prevent the tampering of the safety image of their neighborhood.

Turning off devices in unsafe areas. Users could turn off their iSafe application when entering bad areas. While we cannot prevent this behavior, we propose to use rewards and game mechanics to encourage people to report their location. For instance, users gain points for each reported location, perhaps more for the occasional unsafe location. Points are used to acquire badges, similar in principle to those used by geosocial networks like Foursquare [14] or Yelp [15].

VIII. iSafe Implementation

Figure 8 shows iSafe's extension to the Yelp page of the venue "Top Value Trading Inc." in Hialeah, FL (central left yellow rectangle containing iSafe's safety recommendations).

IX. Experiments

References

[1] Terrafly Project. Crimes and Incidents Reported by Miami-Dade County and Municipal Police Departments. http://vn4.cs.fiu.edu/cgi-bin/arquery.cgi?lat=25.81&long=-80.12&category=crime_dade.
[2] Pang-Ning Tan, Michael Steinbach, and Vipin Kumar. *Introduction to Data Mining*. Addison Wesley, 1 edition, May 2005.
[3] NLTK Project. Natural Language Toolkit. http://nltk.org/.
[4] United States Census. 2010 census. http://2010.census.gov/2010census/, 2010.
[5] Fang-Mei Tseng and Gwo-Hshiung Tzeng. A fuzzy seasonal arima model for forecasting. *Security Journal*, 126:367 – 376, 2002.
[6] George E.P. Box, Gwilym M. Jenkins, and Gregory C. Reinsel. *Time series analysis forecasting and control*. John Wiley and sons, 1994.
[7] Robert F. Nau. Decision 411 forecasting. http://www.duke.edu/ rnau/411avg.htm.
[8] H.Brian Hwarng and H.T Ang. A simple neural network for arma(p,q) time series. *Omega*, 29(4):319 – 333, 2001.
[9] Big Boss. Location spoofer. http://goo.gl/59HMk, 2011.
[10] Gpscheat! http://www.gpscheat.com/.
[11] Bogdan Carbunar and Rahul Potharaju. You unlocked the Mt. Everest Badge on Foursquare! Countering Location Fraud in GeoSocial Networks. In *To appear in Proceedings of the 9th IEEE International Conference on Mobile Ad hoc and Sensor Systems (MASS)*, 2012.
[12] Stefan Saroiu and Alec Wolman. Enabling New Mobile Applications with Location Proofs. In *Proceedings of HotMobile*, 2009.
[13] Z. Zhu and G. Cao. APPLAUS: A Privacy-Preserving Location Proof Updating System for Location-based Services. In *Proceedings of IEEE INFOCOM*, 2011.
[14] Foursquare. https://foursquare.com/.
[15] Yelp. http://www.yelp.com.