Event-level Prediction of Urban Crime Reveals Signature of Enforcement Bias in U.S. Cities

Victor Rotaru[1,3], Yi Huang[1], Timmy Li[1,3], James Evans[2,4,6] and Ishanu Chattopadhyay,[1,4,5★]

[1]Department of Medicine, University of Chicago, Chicago, IL 60637, USA
[2]Department of Sociology, University of Chicago, Chicago, IL 60637, USA
[3]Department of Computer Science, University of Chicago, Chicago, IL 60637, USA
[4]Committee on Quantitative Methods in Social, Behavioral, and Health Sciences, University of Chicago, Chicago, IL 60637, USA
[5]Committee on Genetics, Genomics & Systems Biology, University of Chicago, Chicago, IL 60637, USA
[6]Santa Fe Institute, Santa Fe NM 87501, USA

★To whom correspondence should be addressed: e-mail: ishanu@uchicago.edu.

Policing efforts to thwart crime typically rely on criminal infraction reports, which implicitly manifest a complex relationship between crime, policing and society. As a result, crime prediction and predictive policing have stirred controversy, with the latest AI-based algorithms producing limited insight into the social system of crime. Here we show that while predictive models may enhance state power through criminal surveillance, they also enable surveillance of the state by tracing systemic biases in crime enforcement. We introduce a stochastic inference algorithm that forecasts crime by learning spatio-temporal dependencies from event reports, with mean area under the receiver operating characteristic curve $\approx 90\%$ in Chicago for crimes predicted a week within ≈ 1000 ft. Such predictions enable us to study perturbations of crime patterns that suggest that response to increased crime is biased by neighborhood socio-economic status, draining policy resources from socio-economically disadvantaged areas, as demonstrated in eight major U.S. cities.

The emergence of large-scale data and ubiquitous data-driven modeling has sparked widespread government interest in the possibility of predictive policing[1–5]: predicting crime before it happens to enable anticipatory enforcement. Such efforts, however, do not document the distribution of crime in isolation, but rather its complex relationship with policing and society. In this study, we re-conceptualize the process of crime prediction, build methods to improve upon state of the art, and use it to diagnose both the distribution of reported crime and biases in its enforcement. The history of statistics has co-evolved with the history of criminal prediction, but also with the history of enforcement critique. Siméon Poisson published the Poisson distribution and his theory of probability in an analysis of the number of wrongful convictions in a given country[6]. Andrey Markov introduced Markov processes to show that dependencies between outcomes could still obey the central limit theorem to counter Pavel Nekrasov's argument that because Russian crime reports obeyed the law of large numbers, "decisions made by criminals to commit crimes must all be independent acts of free will"[7].

In this study, we conceptualize the prediction of criminal reports as that of modeling and predicting a system of spatio-temporal point processes unfolding in social context. We report an approach to predict crime in cities at the level of individual events, with predictive accuracy far greater than has been achieved in past. Rather than simply increasing the power of states by predicting the when and where of anticipated crime, our tools allow us to audit them for enforcement biases, and garner deep insight into the nature of the dynamical processes through which policing and crime co-evolve in urban spaces.

Classical investigations into the mechanics of crime[8–10] have recently given way to event-level crime predictions that have enticed police forces to deploy them preemptively and stage interventions targeted at lowering crime rates. These efforts have generated multi-variate models of time-invariant hotspots[11–13], and estimate both long and short term dynamic risks[1–3]. One of the earliest approaches to predictive policing is based on the use of epidemic-type aftershock sequences (ETAS)[4,5], originally developed to model seismic phenomena. While these approaches have suggested the possibility of predictive policing, many achieve only limited out-of-sample performance[4,5]. More recently, deep learning architectures have yielded better results[14]. Machine learning and AI-based systems, however, are often black boxes producing little insight regarding the social system of crime and its rules of organization. Moreover, the issue of how enforcement interacts with, modulates and reinforces crime has been rarely addressed in the context of precise event predictions.

A forecast competition for identifying hotspots prospectively in the City of Portland was recently organized by the National Institute of Justice (NIJ) in 2017 (https://nij.ojp.gov/funding/real-time-crime-forecasting-challenge), which led to the development of multiple effective approaches[15,16] leveraging point processes to model event dynamics, but not accounting for long range and time-delayed emergent interactions between spatial locations. Such approaches, laudable for demonstrating that event-level prediction is possible with actionable accuracy, do not allow for the elucidation of enforcement bias. Informing predictions with the emergent structure of interactions allows us to significantly outperform solutions submitted to the NIJ challenge and simulate realistic enforcement alternatives and consequences.

Results and Discussion

Here we show that crime in cities may be predicted reliably one or more weeks in advance, enabling model-based simulations that reveal both the pattern of reported infractions and the pattern of corresponding police enforcement. We learn from publicly recorded historical event logs, and validate on events in the following year beyond those in the training sample. Using incidence data from the City of Chicago, our spatio-temporal network inference algorithm infers patterns of past event occurrences, and constructs a communicating network (the Granger Network) of local estimators to predict future infractions. In this study, we consider two broad categories of reported criminal infractions: violent crimes consisting of homicides, assault, and battery, and property crimes consisting of burglary, theft and motor-vehicle thefts. The number of individuals arrested during each recorded event is separately modeled and allows us to investigate the possibility and pattern of enforcement bias. We note that while some of these crimes may be more under-reported than others, the relationship between arrests and reports traces police action in response to crime reportage.

We begin by processing event logs to obtain time-series of relevant events, stratified by location and discretized by time, yielding sequential event streams for 1) violent crime (v), 2) property crime (u) and 3) number of arrests (w), as shown in Figure 1, panels a, b and c. To infer the structure of the Granger Net, we learn a finite state probabilistic transducer[17,18] for each possible source-target pair s, r and time lag Δ (Figure 1d), yielding ≈ 2.6 billion modeled associations. Links in the network are retained as they predict events at the target better than the target can predict itself [19]. More details on problem characteristics and performance are provided in Table 1, 2 and Extended Data Table 1 respectively.

For Chicago, we make predictions separately for violent and property crimes, individually within spatial tiles roughly $1000ft$ across and time windows of 1 day approximately a week in advance with area under the receiver operating characteristic curves (AUCs) ranging from $80 - 99\%$ across the city (see below for alternative measures tuned to the concerns of policing policy). We summarize our prediction results in Figure 2, where panels a and b illustrate the geospatial scatter of AUCs obtained for different spatial tiles and types of crime, and c shows the distribution of AUCs. Out-of-sample predictive performance remains stable over time; our predictions on successive years (each using three preceding years for training, and one year for out-of-sample testing, see Extended Data Figure 1) shows little variation in average AUC. Inspecting excerpts of the average daily crime rate for successive years also demonstrates a cose match between actual and predicted behavior (See Extended Data Figure 2, panels a, c and e.) The remaining panels (b, d and f) in the same figure illustrate how the Fourier coefficients match up, showing that we are able to capture crime periodicities at weekly and bi-weekly scales, and beyond.

Unlike previous efforts[1–5], we do not impose pre-defined spatial constraints. In contrast to contiguous diffusion encountered in physical systems, criminal reportage may spread across the complex landscape of a modern city unevenly, with regions hyperlinked by transportation networks, socio-demographic similarity and historical collocation, which cannot be captured with spatial diffusion models[20]. Rather than assuming that distant events across the city will have a weaker influence on prediction compared with those physically near in space or time, we probe the topological structure emergent from inferred dependencies to estimate the shape, size and organization of neighborhoods that best predict events at each location. Results illustrated in Figure 2d and e show that the situation is complex with the locally predictive neighborhoods varying widely in geometry and size, implying that restricting analysis to small local communities within the city is sub-optimal for crime prediction and enforcement analysis. In order to analyze whether the effect of reported criminal infractions diffuse outward in space and time, we simply calculate temporal-spatial distances of predictive dependencies, then average across all neighborhoods in the city, revealing the rapid decay with time delay in diffusion rates shown in Figure 2f. Interestingly we find the property and violent crimes differ in their rates of predictive diffusion (Figure 2f); while signals from property crime decay rapidly within days, violent reported events appear to shape dynamics for weeks in future. These differences in diffusion appear to manifest how people differentially mimic and process exposure to violence[21,22].

Forecasting crime via analyzing historical patterns has been attempted before[23] (See also unpublished manuscript at https://arxiv.org/abs/1806.01486). State of the art approaches use machine deep learning tools based on recurrent and convolutional neural networks (NN). In the first article[23], the authors train a NN model to predict next-day events for $60,348$ sample points in Chicago. The model is trained on crime statistics, demographic makeup, meteorological data, and Google street view images to track graffiti, achieving an out-of-sample AUC of 83.3%. Our AUC is demonstrably higher (see Table 2 and Extended Data Table 1), and we predict with significantly less data (only past events), and 7 days into future (instead of next-day). Additionally, the use of demographic and graffiti is problematic with the possibility of introducing racial and socio-economic bias, with dubious causal value. In the second article[24], the authors combine convolutional and recurrent neural networks with weather, socio-economic, transportation, and crime data, to predict the next-day count of crime in Chicago. As spatial tiles, the authors use standard police beats, which break up Chicago into 274 regions. Police beats reflect the classical notion of neighborhoods, and measure approximately 1 sq. mile on average[25]. In comparison, our spatial times are approximately 0.04 sq. miles, representing a 2500% higher resolution. This model achieves a classification accuracy of 75.6% for Chicago, which compares against our accuracy of $> 90\%$ (See Table 2). While this competing model tracks more crime categories, it is limited to next-day predictions with significantly coarser spatial resolution. We also compare the predictive ability of naive autoregressive baseline models (See Methods and Extended Data Table 2), which perform poorly, but provide a yardstick to meaningfully compare our claimed performance estimates, which underwrite the application of our approach in revealing emergent biases (See Figure 3 and 4) . Apart from AUC and accuracy, we also report other common performance metrics in Table 2, namely specificity obtained at a fixed sensitivity of 80%, and the precision or the Positive Predictive Value (PPV).

We also compute the Predictive Accuracy Index (PAI), and the Prediction Efficiency Index (PEI) achieved for each city we consider. The PAI[16] is defined as the normalized event rate in identified hotspots (tiles predicted to have events), and the PEI[16] is the ratio of PAI achieved to its maximum achievable value by the same algorithm (and is thus bounded between 0 and 1, see Crime Prediction Metrics in Methods). The PAI/PEI have emerged to become metrics of choice for crime models due to the need for maximizing the volume of crime in predicted hotspots to enable law enforcement. Importantly, PAI/PEI comparisons are distinct from AUC calculations; an algorithm can have achieve a high AUC with poor PAI or PEI scores. Our PAI and PEI scores indicate strong performance, with PEIs approaching 1.0 (See Figure 5a).

Finally, a head-to-head comparison of the efficacy of our approach over reported tools is obtained for data used in a recent crime forecast challenge hosted by the NIJ. The Portland Police Department provided crime data from March 2012 up to the end of February 2017, and participants were asked to forecast crime hotspots for four types of incidents (burglary, motor vehicle theft, street crime, and all calls for service) over the months of March, April, and May of 2017. In particular, participants were asked to define a grid restricted to Portland boundaries, and predict “hotspot” grid cells for each crime type over several forecasting windows. This challenge was a true prospective forecasting test as the validation time-period was in future, non-existent at time of submission. Forecasts were made for 1-week, 2-week, 1-month, 2-month, and 3-month time windows and scored with PAI and PEI. The two metrics are not equivalent, as illustrated in the NIJ challenge results, with different teams winning in different categories with respect to the different metrics. While a natural equivalency between PAI and PEI has been suggested [16], frameworks have not been reported previously that optimize them both. Our results on the data released for this challenge are enumerated in Figure 5b, where we outperform the best performing team in 119 of 120 categories (under-performing on street crimes at the 3 month horizon).

With the above-discussed predictive performance establishing the validity of our models, we run a series of computational experiments that perturb rates of violent and property crimes, then log the resulting alterations in future event rates across the city. By inspecting the effect of socio-economic status (SES) on perturbation response, we investigate whether enforcement and policy biases modulate outcomes. The inferred stress response of the city suggests the presence of socio-economic enforcement bias (See Figure 3). Wealthier neighborhoods respond to elevated crime rates with increased arrests, while arrest rates in disadvantaged neighborhoods drop, but the converse does not occur (See Figure 3, panels e and f). We argue that resource constraints on law enforcement, combined with biased prioritization to wealthier neighborhoods, result in reduced enforcement across the remainder of the city. Thus, our results align with suspected enforcement bias within U.S. cities that parallels widely discussed notions of suburban bias in high SES suburbs[26,27]. While self-evident at the scale of countries and regions, the existence of unequal resource allocation in cities, where political power and influence concentrates in selective, high SES neighborhoods, has been widely suspected[28–31]. Our analysis corroborates this contention, which shows up robustly for all years analyzed, going back over one and a half decades in Chicago. Extended Data Figs 3, 4 and 5 show that these patterns are stable over the time period we analyze. Additionally, Extended Data Figure 3 shows the effect of perturbations across all variables, suggesting that crime reduction from perturbations seems most effective in regions with high crime rates, acknowledging confounding with SES.

The Granger Net allows for precise simulation of the impact of complex local and global event patterns, and has the potential to emerge as an important tool in policy-making. Thus, empirical validations of model predictions are important. To corroborate claimed disparities in enforcement response without using our inferred models, we identify similar, naturally occurring patterns in crime and arrest rates across the city of Chicago. Without the use of our models, it is difficult to obtain uniform event stimuli across the city. In one approach, we exploit the seasonality of crime, and compare summer months against late winter. Figure 6a(i) shows the increase in violent and property crimes from February to June/August, averaged across rich and poor neighborhoods over 4 years from 2014 to 2017 (95% confidence bounds shown). Here we define rich neighborhoods as communities with hardship index < 20 (results are not sensitive to the choice of threshold). We observe that the average percentage increase in event rate from late winter to summer is broadly comparable across the city, thus approximating a uniform perturbation in crime rate. As shown in Figure 6a(ii), the corresponding deviation of mean percentage change in arrest rate from the city-wide average reflects our conclusions above: wealthier communities see an increase in arrest rate per unit event with the seasonal rise in crime, while others experience a draw-down.

Changes in enforcement response from winter to summer months do not necessarily establish that a uptick in arrests in high SES areas is associated with a down-tick elsewhere in the near future. Thus, we carry out a more granular interrogation of the raw crime data as follows. Aggregating data on the number of daily arrests over Chicago communities (Chicago has 77 community areas[32]), we compute the correlation between daily change in the total number of arrests, and their 1-day delayed versions in neighboring communities with more economic hardship (higher hardship indices). For each community s, we denote as $\mu(s)$ the value of this correlation minimized over all neighboring communities of s. Figure 6b(i) shows the variation of $\mu(s)$ with $h(s)$, the hardship index of the community s. We see that the arrest rate change in wealthier communities are more strongly anti-correlated with the 1-day delayed arrest rate change in neighboring more disadvantaged communities. And Figure 6b(ii) shows the correlation of $\mu(s)$ with the average hardship index of neighboring communities of s, computed separately within community groups of similar economic status. We observe that for wealthier communities, the anti-correlation between daily change in arrests and its delayed version in lower SES neighboring communities is stronger the more economically disadvantaged the neighbors are. The higher the average hardship index of the neighbors, the more negative μ, leading to more negative values in Figure 6b(ii). We also see that this effect vanishes and eventually reverses as the SES of focal communities themselves become lower—as their economic status degrades. These direct observations lend credence to the model-based indication of enforcement bias arising from differential resource allocation.

Beyond Chicago, we analyze criminal event logs available in the public domain for seven additional major US cities: Detroit, Philadelphia, Atlanta, Austin, San Francisco, Los Angeles and Portland. In all these cities we obtain comparably high performance in predicting violent and property crimes, with average AUC ranging between 86-90% (See Figure 4a-f and Supplementary Figure 1. In addition, our observed pattern of perturbation responses in Chicago, which suggests de-allocation of policing resources from disadvantaged neighborhoods to advantaged ones, is replicated in all these cities. While crime rate increases with degrading SES status of local neighborhoods, number of predicted events a week after a positive 5-10% increase in crime rate goes down. Thus increasing the crime rate leads to a smaller number of reported crimes, a pattern holding more often in lower SES neighborhoods.

Our analysis also sheds light on continuing debate over the choice for neighborhood boundaries in modeling crime in cities[33–36]. In Figure 2d-f, we demonstrate that despite apparent natural boundaries, predictive signals are often communicated over large distances and decay slowly, especially for violent crimes. More importantly, this study reveals how the "correct" choice of spatial scale should not be a major issue in sophisticated learning algorithms where optimal scales can be inferred automatically. We find that there exists a skeleton set of spatial tiles, which bound predictive dependencies on overall event patterns (See Extended Data Figure 6). These induce a cellular decomposition of the city that identifies functional neighborhoods, where the cell-size adapts automatically to local event dynamics.

Limitations & Conclusion

Our ability to probe for the extent of enforcement bias is limited by our dataset on criminal reportage, without the use of direct data on the spatial distribution of police. In large US cities, place and race is often synonymous[37,38]; disproportionate police response in communities of color can contribute to biases in event logs, which might propagate into inferred models. This possibility has elicited significant push-back against predictive policing[39]. Our approach is free from manual encoding of features (and thus resistant to implicit biases of the modelers themselves), but bias arising from disproportionate crime reportage and surveillance almost certainly remain. We doubt if any amount of scrubbing or clever statistical controls can reliably erase such ecological patterning of apparent crime. Any policy informed by our results must keep this caveat in mind.

Differences in the extent to which different communities trust law enforcement are important in analyzing crime and enforcement. Diverse communities are often less inclined to call law enforcement for help, or report criminal acts they might witness, thus obfuscating underlying crime rates. To mitigate these effects we only consider events, *e.g.*, homicides, battery, assault, automobile-theft, burglary, that are much less likely to be optionally reported by residents, or those which are directly observed by police officers. This is perhaps more true for the violent crime types considered, and our predictive performance and conclusions replicate for both violent and property crimes. The exception is the City of Portland, where we do consider “street crimes” and “all calls for service” to compare our performance in the NIJ forecast challenge. Our performance holds up in these categories (Figure 5b), suggesting that these differential reporting issues may not significantly affect our results, but we note that we outperform the competition to a lesser degree for these categories. Finally, for the City of Chicago, we consider arrests as a distinct variable in addition to crimes logged. Importantly, we only consider arrests related to the crimes considered, mitigating the effects of potential over/under-reporting if all such events were to be included.

Despite our caution, one of our key concerns in authoring this study is its potential for misuse – an issue which predictive policing strategies have struggled with[40]. More important than making good predictions is how such capability will be used. Because policing is as much “person based” as “place based”[41,42], sending police to an area, regardless of how small that area is, does not dictate the optimal course of action when they arrive, and it is conceivable that good predictions (and intentions) can lead to over-policing or police abuses. For example, our results may be falsely interpreted to mean that there is “too much” policing in low crime (often predominantly White) communities, and too little policing in higher crime (often more racially and ethnically diverse) neighborhoods. A policy based on such a mis-interpretation might ramp up enforcement in Black and Latino neighborhoods, creating a harmful feedback of sending more police to areas that might already feel over-policed but under-protected[43]. Instead our results recommend changes in policy that result in more equitable need-based resource allocation, with reduced impact based on the socio-economic status of individual communities. The tools reported here can then be used to track the extent to which such policies approach this trace of equitable enforcement allocation.

Even with its current limitations, our approach is an addition to the toolbox of computational social science, enabling validation of social theory from observed event incidence, supplementing the use of measurable proxies and potential biases in questionnaire-based data collection strategies. While classical approaches[44–47] broaden our understanding of the societal forces shaping both urban and regional landscapes, these approaches have neither successfully attempted to forecast individual infraction reports, nor reveal how these predictive patterns manifest systematic enforcement bias. In this study, we show how the ability of Granger Networks to predict such events not only allows precise intervention, but also advances the diagnosis and explanation of complex social patterns. We acknowledge the danger that powerful predictive tools place in the hands of over-zealous states in the name of civilian protection, but here we demonstrate their unprecedented ability to audit enforcement biases and hold states accountable in ways inconceivable in the past. We encourage widespread debate regarding how these technologies are used to augment state action in public life, and call for transparency that allows for continuous evaluation, reconsideration and critique.

METHODS

In this study we use historical geolocated incidence data of criminal infractions to model and predict future events in Chicago, Philadelphia, San Francisco, Austin, Los Angeles, Detroit and Atlanta. Each of the cities considered have a specific temporal and spatial resolution, which are optimized to maximize predictive performance (See Table 1). The predictive performance obtained in these cities are enumerated in Table 2 and Extended Data Table 1. The distribution of AUCs obtained in Chicago for earlier years (2014-2017, predicted individually) are shown in Extended Data Figure 1.

Data Source

The sources of crime incidence data used in this study for the different US cities are enumerated in Table 1. Theses logs include spatio-temporal event localization along with the nature, category, and a brief description of the recorded incident. For the City of Chicago, we also have access to the number of arrests made during or as a result of each event. For Chicago, the log is updated daily, keeping current with a lag of 7 days, and we make predictions for each of the years 2014-2017 (using 3 years before the target year for model inference, and 1 year for out-of-sample validation) for the prediction results shown in Figure 1. The evolving nature of the urban scenescape[48] necessitates that we restrict the modeling window to a few years at a time. The length of this window is decided by trading off loss of performance from shorter data streams to ignoring evolution of underlying generative processes with longer streams. The training and testing periods of other cities is tabulated in Table 1. In this study, we consider two broad categories of criminal infractions: violent crimes consisting of homicides, assault, battery etc., and property crimes consisting of burglary, theft, motor vehicle theft etc. Drug crimes are excluded from our consideration due to the possibility of ambiguity in the use of violence and the potential for biased documentation of such events. For the City of Chicago, the number of individuals arrested during each recorded event is considered a separate variable to be modeled and predicted, which allows us to investigate the possibility of enforcement biases in subsequent perturbation analyses.

We also use data on socio-economic variables available at the portal corresponding to Chicago community areas and census tracts, including % of population living in crowded housing, those residing below the poverty line, those unemployed at various age groups, per capita income, and the urban hardship index[49]. Such data is also obtained from the City of Chicago data portal. Additionally, we use data on poverty estimates for the other cities, which are obtained https://www.census.gov.

Spatial and Temporal Discretization & Event Quantization

Event logs are processed to obtain time-series of relevant events, stratified by occurrence locations. This is accomplished by choosing a spatial discretization, and focusing on one individual spatial tile at a time, which allows us to represent the event log as a collection of sequential event streams (See Figure 1c). Additionally, we discretize time, and consider the sum total of events recorded within each time window.

Coarseness of these discretizations reflects a trade-off between computational complexity and event localization in space and time. Spatial and temporal discretizations are not independently chosen; a finer spatial discretization dictates a coarser temporal quantization, and visa verse to prevent long no-event stretches and long periods of contiguous event records, both of which reduce our ability to obtain reliable predictions. For the City of Chicago, we fix the temporal quantization to 1 day, and choose a spatial quantization such that we have high empirical entropy rates for the time series obtained. This results in spatial tiles measuring $0.00276^\circ \times 0.0035^\circ$ in latitude and longitude respectively, which is approximately $1000'$ across, roughly corresponding to an area of under 2×2 city blocks. Thus, any two points within our spatial tile are at worst in neighboring city blocks. We dropped from our analysis the tiles that have too low a crime rate ($< 5\%$ of days within the modeling window had any event recorded) to reduce computational complexity, resulting in an $N = 2205$ of spatial tiles in the city of Chicago. The temporal and spatial resolution is adjusted in a similar manner for other cities (See Table 1).

Thus, we end up with three different integer-valued time series at each spatial tile: 1) violent crime (v), 2) property crime (u) and 3) number of arrests (w) in the City of Chicago. For other cities, we have only the first two categories, because information on arrests was not available. We ignore the magnitude of the observations, and treat them as Boolean variables. Thus, our models simply predict the presence or absence of a particular event type in a discrete spatial tile within a neighboring city block and observation window, $i.e.$, within the temporal resolution chosen, which is 1 day except for Atlanta, where is it is chosen to be 2 days (See Table 1).

Inferring Generators of Spatio-temporal Cross-dependence

Let $\mathcal{L} = \{\ell_1, \cdots, \ell_N\}$ be the set of spatial tiles, and $\mathcal{E} = \{u, v, w\}$ be the set of event categories as described in the last section. At location $\ell \in \mathcal{L}$ for variable $e \in \mathcal{E}$, at time t, we have $(\ell, e)_t \in \{0, 1\}$, with 1 indicating the presence of at least one event. The set of all such combined variables (space + event type) is denoted as $\mathcal{S}$, $i.e.$, $\mathcal{S} = \mathcal{L} \times \mathcal{E}$. Let $T = \{0, \cdots, M-1\}$ denote the training period consisting of M time steps. Because for any time t, $(\ell, e)_t$ is a random variable, our goal here is to learn its dependency relationships with its own past, and with other variables in S to accurately estimate its future distribution for $t > T$.

To infer the structure of our predictive model, we learn a finite state probabilistic transducer[18] (referred to as a Crossed Probabilistic Finite State Automata or a XPFSA (a generalization of probabilistic finite state automata models for stochastic processes[17], see unpublished manuscript at http://arxiv.org/abs/1406.6651) for each possible source-target pair $s, r \in \mathcal{S}$. Given a sequence of events at the source, these inferred transducers estimate the distribution of events at target r for some future point in time. Ability to estimate such a non-trivial distribution indicates success in prediction. With too many uncontrollable factors influencing the outcomes, causality cannot be inferred from data for the problem at hand. Here we characterize directional dependency as the source being able to predict events occurring at the target, better than the target can do by itself. This prediction-centered approach has been called Granger-causal influence[50], but while this has been criticized as a weak indicator of causality, it is directly tuned to the challenge of forecasting future events. Importantly, we do not assume that the underlying processes are iid, or that the model has any particular linear structure. Additionally, predictive dependencies are= not restricted to be instantaneous. The source events might impact the target with a time delay, $i.e.$, a specific model between the source and target might predict events delayed by an a priori determined number of steps $\Delta_{max} \geqq \Delta \geqq 0$ specific to the model. Here we model the dependency structure for each integer-valued delay separately. Thus, for source s and target t, we can have $\Delta_{max} + 1$ transducers each modeling dependencies for a specific delay in $\{0, \Delta_{max}\}$. The maximum number of steps in time delay Δ_{max} is chosen a priori, based on the problem at hand.

While these dependencies may differ for different delays, they need not be symmetric between source and target pairs. The complete set, comprising at most $|\mathcal{S}|^2(\Delta_{max} + 1)$ models, represents a predictive framework for asymmetric multi-scale spatio-temporal phenomena. Note that the number of possible models increase quickly. For example, for the City of Chicago, for $\Delta_{max} = 60$ with 2205 spatial tiles and three event categories, the number of inferred models is bounded above by ≈ 2.6 billion.

Our approach consists of inferring XPFSAs in two key steps (See Figure 1d, and discussion later in Supplementary

Methods: First, we infer XPFSA models for all source-target pairs and all delays up to Δ_{max}. In the second step, we learn a linear combination of these transducers to maximize predictive performance. Denoting the observed event sequence in time interval $(\infty, t]$ at source s as $s_t^{-\infty}$, the XPFSA $\mathbb{H}^s_{r,k}$ estimates the distribution of events for target r at time step $t+k$. This is accomplished by learning an equivalence relation on the historical event sequences observed at source s, such that equivalent histories induce an approximately identical future event distribution at target r, k steps in the future. Thus, for example, the XPFSA shown in Figure 1d has four states, indicating that there are 4 such equivalence classes of observations that induce the distinct output probabilities shown from each state. Often this estimate is imprecise due to the possibility for multi-scale and multi-source dependencies, $e.g.$, when target r is predicted by multiple sources with different time delays. In the second step, we employ a standard gradient boosting regressor for each target, to optimize the linear combination of inferred transducers and learn the scalar weights $\omega^s_{r,k}$ for source s, target r and delay k. Detailed pseudocode of the inference algorithms are provided in the Supplementary Methods.

To compare with a standard neural net architecture, these probabilistic transducers may be viewed as local non-linear activation functions. With neural networks we repeatedly compute affine combination of inputs and apply fixed non-linear activation to the combined input and finally optimize affine combination weights via backpropagation, but here we first learn the local non-linear activations, and then optimize the linear or affine combination of weak estimators. Optimizing the weights is a significantly simpler, local operation and may be done with any standard regressor. In contrast to recurrent neural nets (RNN), the role of hidden layer neurons is partially accounted for by states of the XPFSA, which are a priori undetermined both with respect to their multiplicity and their transition connectivity structure.

Computational & Model Complexity

We assume the maximum time delay in prediction propagation to be 60 days for all cities, which for the City of Chicago results in at most $2,669,251,725$ inferred models, of which $61,650,000$ are useful with $\gamma \geqq 0.01$. Model inference in this case consumed approximately $200K$ core-hours on 28 core Intel Broadwell processors, when carried out with incidence data over the period Jan 1, 2014 to December 31, 2016. Computational cost for other time-periods and other cities are comparable and roughly scale with the square of the number of spatial tiles, and linearly with the length of time-quantized data-streams considered as input to the inference algorithm.

Crime Prediction Metrics

For each spatial location, the inferred Granger Net maps event histories to a raw risk score as a function of time. The higher this value, the higher the probability of an event of target type occurring at that location, within the specified time window. To make crisp predictions, however, we must choose a decision threshold for this raw score. Conceptually identical to the notion of Type 1 and Type 2 errors in classical statistical analyses, the choice of a threshold trades off false positives (Type 1 error) for false negatives (Type 2 error). Choosing a small threshold results in predicting a larger fraction of future events correctly, $i.e.$, have a high true positive rate (TPR), while simultaneously suffering from a higher false positive rate (FPR), and vice versa. The receiver operating characteristic curve (ROC) is the plot of the FPR vs the TPR, as we vary this decision threshold. If our predictor is good, we will consistently achieve high TPR with small FPR resulting in a large area under the ROC curve denoted as the AUC. Importantly, AUC measures intrinsic performance, independent of the threshold choice. Thus, the AUC is immune to class imbalance (the fact that crimes are rare events). An AUC of 50% indicates that the predictor does no better than random, and an AUC of 100% implies that we can achieve perfect prediction of future events, with zero false positives.

For evaluating AUC, we treat a positive prediction as correct if there is at least one event recorded in ± 1 time steps in the target spatial tile.

We also evaluate the PAI and PEI achieved in our framework. The PAI is defined as follows: Given a set of k predicted hotspot cells, the PAI is determined by computing the ratio of the proportion of crime captured in the hotspots relative to the proportional area of the city flagged as hotspots. Specifically, defining H to be the union of the hotspot cells (which does not need to be connected) and S the spatial region of interest ($e.g.$, Portland, OR), the PAI is defined

$$PAI(H) = \frac{N(H)|S|}{|H|N(S)} \tag{1}$$

where $N(H)$ is the number of events in H over the forecasting window and $|H|$ is the size of the hotspot region $H \subseteq S$. Letting $\lambda(H) = N(H)/|H|$ be the estimated intensity of events in region H and $\overline{\lambda} = N(S)/\,|S|$ be the total intensity of events in the region of interest, the PAI becomes

$$PAI(H) = \frac{\lambda(H)}{\overline{\lambda}} \propto \lambda(H) \tag{2}$$

which is only a function of $\lambda(H)$ since $\overline{\lambda}$ is independent of H. Thus, PAI is interpreted as the average rate of crime in predicted hotspots relative to the average crime rate in a city. The trends obtained for the PAI and PEI with our approach match those reported in the literature (See Figure 3 in Mohler *et al.*[16]).

Predictability Analysis

In the City of Chicago, we can predict events approximately a week in advance at the spatial resolution of ± 1 city blocks with a temporal resolution of ± 1 day, with a false positive rate of less than 20% and a median true positive rate of 78%. The predictive performance in other cities is enumerated in Table 2. While not directly modeled in the frequency domain, we found that the event forecasts produce very similar signatures in the frequency domain (See Extended Data Figure 2), when compared over the first 150 days of each out-of-sample period (1 yr). We also consider prediction periods of 7, 14, 30, 60 and 100 days to evaluate the variation of PAI/PEI for the cities considered (See Figure 5a).

Spatial Neighborhoods

The degree of directed predictive dependency between one variable (the source stream) on another (the target stream), also called the (Granger-)causal influence, is quantified by the coefficient of dependence (γ, see Supplementary Methods). Identifying the source-target pairs for which the coefficient of dependency (or Granger-causality) is high (See Extended Data Figure 6), we note that there exists a sparse set of spatial tiles which exert nearly all of the directed dependency in the entire set of observed variables. Thus, observing these variables alone would enable us to make good event forecasts. These tiles span the expanse of the city, and a Voronoi decomposition based on the centers of these tiles in shown in Extended Data Figure 6b. Such a decomposition demonstrates an algorithmic approach to choosing optimal neighborhoods for urban analysis.

Perturbation Analysis

We experimented with positive and negative perturbations to both violent and property crime rates ranging from 1 to 10% of observed rates. Response to perturbed crime rates was measured as the relative change from nominal baseline in estimated time-average for the predicted event frequencies 1 week in the future, corresponding to violent and property crimes and number of arrests.

Results from our perturbation experiments shed light both on the stability characteristics of crime in Chicago, and further allowed us to look for evidence of biased police enforcement responses under stress. Under stress, well-off neighborhoods tend to drain resources disproportionately from disadvantaged locales (See Figure 3). Economically well-off neighborhoods in the bottom 25% of the hardship index are much more likely to see a near -proportional increase ($\approx\ 15\%$) in law enforcement response, measured by the number or predicted arrests on a 10% increase in crime rates (See Figure 3, panels c and d, which show how regions with increased enforcement response are concentrated in well-off neighborhoods), while the rest of the city sees a drop in predicted response of about twice the magnitude ($> 30\%$). Increased crimes causes enforcement resources to be drained from disadvantaged neighborhoods to support their better socioeconomic counterparts. We performed multi-variable linear regression analysis to evaluate the question in another way. Here we regressed violent and property crime rates, independently, on the variables listed in (Figure 3b), including a slope intercept variable in each model. In both models, the hardship index's strong, negative coefficient for changes in arrest rate from perturbations that increase violent and the property crime rates contradicts what might be expected in the absence of bias. Lower SES neighborhoods have more crime and so these socio-economic indicators should contribute positively to the arrest rate with increasing crime. These patterns were replicated in our perturbation experiments for all preceding years we analyzed (2014 through 2017, See Extended Data Figs 4 and 5). Response measured in the property an violent crimes, and in the associated arrests from perturbations is detailed in Extended Data Figure 3.

We also carried out similar perturbation analyses for the other cities, and observed that with increasing poverty we have expected increase of observed crime rates, but an unexpected decrease in violent and property crimes after a 5-10% simulated uptick in either category of crimes (See Figure 4).

Naive Baselines: Autoregressive Integrated Moving Average (ARIMA) Models

To explore the predictive ability of naive baseline models on our datasets, we consider four ARIMA[51] configurations with lag orders $p = 5$ and 10, numbers of differencing $d = 1$ and 2, and the window of moving average $q = 0$. Let y_t be the series we want to model and y'_t be y_t differenced by d times, the ARIMA(p, d, q) models series y'_t by

$$y'_t = c + \phi_1 y'_{t-1} + \cdots + \phi_p y'_{t-p} + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q} + \varepsilon_t \tag{3}$$

where $\phi_1, \ldots, \phi_p$ and $\theta_1, \ldots, \theta_q$ are the coefficients to be fitted. In Eq. (3), y'_{t-k}s are the historical values of y'_t whose inclusion models the influence of past values on the current value (autoregression), and ε_{t-k}s are the white noise terms whose inclusion models the dependence of current value against current and previous (observed) white noise error terms or random shocks (moving average). Specifically, we use the following four models for the earthquake and the crime datasets

$$y_t^{(1)} = c + \phi_1 y'_{t-1} + \cdots + \phi_5 y'_{t-5} \tag{4}$$

$$y_t^{(1)} = c + \phi_1 y'_{t-1} + \cdots + \phi_5 y'_{t-10} \tag{5}$$
$$y_t^{(2)} = c + \phi_1 y'_{t-1} + \cdots + \phi_5 y'_{t-5} \tag{6}$$
$$y_t^{(2)} = c + \phi_1 y'_{t-1} + \cdots + \phi_5 y'_{t-10} \tag{7}$$

where $y_t^{(d)}$ is y_t different by d times ($y_t^{(1)} = y_t - y_{t-1}$ and $y_t^{(2)} = y_t - 2y_{t-1} + y_{t-2}$). For simple benchmarks we apply the ARIMA model to each individual time series, which means the predictive model is trained without exogenous variables. For the implementation, we use the Python `statsmodels` package[52], and the result is shown in Extended Data Table 2. The inadequate performance of ARIMA may be due to 1) the use of a single data stream limits the ability of ARIMA to capture the interplay between co-evolving processes, and 2) a pre-determined lag order fails to capture the possibly varying temporal memory of individual processes.

Data Availability

Crime incident data used in this study is in the public domain. The weblinks for the data sources for seven out of the eight cities considered here are as follows: opendata.atlantapd.org, data.austintexas.gov,data.detroitmi.gov,data.lacity.org,www.opendata.philly.org,data.sfgov.org,data.cityofchicago.org, and for Portland the data along with the leader-board data for the forecasting challenge was obtained from nij.ojp.gov.

Code Availability

Software with source code is available at https://github.com/zeroknowledgediscovery/Cynet, and the current version of the software may be referenced by the https://doi.org/10.5281/zenodo.5730613. Any questions on implementation should be directed to the corresponding author.

Acknowledgments

Our work greatly benefited from discussion of everyone who participated in our workshop series on crime prediction at the Neubauer Collegium for culture and society (https://neubauercollegium.uchicago.edu/events/uc/crimes_of_prediction_workshop/), and with those with whom we had extended conversations to ground and refine our modeling approach.

Data was provided by the City of Chicago Data Portal at https://data.cityofchicago.org. The City of Chicago ("City") voluntarily provides the data on this website as a service to the public. The City makes no warranty, representation, or guaranty as to the content, accuracy, timeliness, or completeness of any of the data provided at this website (https://www.chicago.gov/city/en/narr/foia/data_disclaimer.html), and the authors of this study are solely responsible for the opinions and conclusions expressed in this study. Sources of the crime incidence data for the other cities are tabulated in Table 1. Socio-economic data for metropolitan areas was obtained from https://www.census.gov.

This work is funded in part by the Defense Sciences Office of the Defense Advanced Research Projects Agency projects HR00111890043/P00004 and W911NF2010302, and the Neubauer Collegium for Culture and Society through the Faculty Initiated Research Program 2017. The claims made in this study do not necessarily reflect the position or the policy of the sponsors, and no official endorsement should be inferred.

Author Contributions

YH and VR are co-first authors with equal contribution. YH and IC worked out key mathematical details of the Granger Net framework. VR, TL, YH, and IC implemented the algorithms and generated results. YH, VR and TL contributed equally in realizing the current implementation of the software. IC generated the visualizations in this study. JE provided key insights into modeling and interpreting social dynamics. JE and IC conceived and designed the research, and wrote the paper.

Competing Interests

Authors declare no competing interests.

TABLE 1
Crime Event Log Information for Cities Considered

	Atlanta	**Austin**	**Detroit**	**Los Angeles**	**Philadel-phia**	**San Francisco**	**Chicago**	**Portland**
no. of variables[1]	510	1082	1161	3287	1037	975	3826	9354
temporal resolution	2 days	1 day	1 day	1 day	1 day	1 day	1 day	3 days
bounding box of modeled region	33.65°N, 33.86°N, 84.54°W, 84.31°W	30.14°N, 30.48°N, 97.89°W, 97.63°W	42.30°N, 42.45°N, 83.28°W, 82.91°W	33.71°N, 34.33°N, 118.65°W, 118.16°W	39.88°N, 40.12°N, 75.27°W, 74.96°W	37.71°N, 37.81°N, 122.51°W, 122.36°W	41.64°N, 42.06°N, 87.88°W, 87.52°W	45.23°N, 45.81°N, 123.05°W, 122.22°W
spatial resolution	$983' \times 983'$	$983' \times 983'$	$983' \times 983'$	$983' \times 983'$	$983' \times 983'$	$983' \times 983'$	$951' \times 1006'$	$591' \times 591'$
Spatial exclusion threshold[2]	2.5%	2.5%	2.5%	2.5%	5.0%	2.5%	5.0%	2.0%
training period	14/01/01-18/12/31	16/01/01-18/12/31	12/01/01-14/12/31	16/01/01-18/12/31	16/01/01-18/12/31	14/01/01-16/12/31	14/01/01-16/12/31	12/03/01-17/02/28
test period	19/01/01-19/07/20	19/01/01-19/04/11	15/01/01-15/04/11	19/01/01-19/04/11	19/01/01-19/04/11	17/01/01-17/04/11	17/01/01-17/04/11	17/03/01-17/05/31
prediction horizon	6 days	3 days	3 days	3 days	3 days	3 days	7 days	9 days
violent crime stat.	event count 2649, rate 3.98%	event count 20132, rate 5.45%	event count 20922, rate 3.72%	event count 72355, rate 4.83%	event count 33803, rate 8.11%	event count 23317, rate 7.16%	event count 179274, rate 7.7%	See Table 1
property crime stat.	event count 23522, rate 4.51%	event count 88929, rate 6.22%	event count 39840, rate 3.30%	event count 205435, rate 5.49%	event count 85683, rate 9.02%	event count 197835, rate 12.83%	event count 263661, rate 7.0%	See Table 1
data source	opendata.atlantapd.org	data.austintexas.gov	data.detroitmi.gov	data.lacity.org	www.opendata\philly.org	data.sfgov.org	data.cityofchicago.org	nij.ojp.gov

[1] No. of variables indicates the total number of time series considered for violent and property crimes.

[2] Tiles with less than threshold event-rate were excluded.

TABLE 2
Prediction performance with Granger Net for seven US cities

city	**property crimes**				**violent crimes**			
	specificity†	**AUC**	**acc.**††	**PPV**⋆	**specificity**	**AUC**	**acc.**	**PPV**
Atlanta	0.68	0.90	0.84	0.39	0.71	0.88	0.84	0.38
Austin	0.66	0.87	0.82	0.40	0.66	0.88	0.83	0.38
Detroit	0.72	0.90	0.86	0.37	0.66	0.89	0.84	0.35
Philadelphia	0.64	0.87	0.81	0.48	0.65	0.87	0.81	0.47
Los Angeles	0.66	0.84	0.83	0.39	0.65	0.84	0.83	0.36
San Francisco	0.67	0.86	0.80	0.52	0.65	0.86	0.81	0.42
Chicago	0.68	0.87	0.93	0.43	0.67	0.87	0.94	0.46

† Median specificity at 80% sensitivity

†† Accuracy calculated with $\max$ sensitivity $\times$ frequency $+$ specificity $\times$ $(1 - \text{frequency})$.

⋆ Positive predictive value (PPV) calculated with $\max \frac{\text{sensitivity} \times \text{frequency}}{\text{sensitivity} \times \text{frequency} + (1 - \text{specificity}) \times (1 - \text{frequency})}$.

Fig. 1. Crime Data & Modeling Approach. Panels a and b show the recorded infractions within the 2 week period between April 1 and 15 in 2017. Plate c illustrates our modeling approach: We break city into small spatial tiles approximately 1.5 times the size of an average city block, and compute models that capture multi-scale dependencies between the sequential event streams recorded at distinct tiles. In this paper, we treat violent and property crimes separately, and show that these categories have intriguing cross-dependencies. Plate d illustrates our modeling approach. For example, to predict property crimes at some spatial tile r, we proceed as follows: Step 1) we infer the probabilistic transducers that estimate event sequence at r by using as input the sequences of recorded infractions (of different categories) at potentially all remote locations (s, s', s'' shown), where this predictive influence might transpire over different time delays (a few shown on the edges between s and r). Step 2) Combine these weak estimators linearly to minimize zero-one loss. The inferred transducers can be thought of as inferred local activation rules, which are then linearly composed, reversing the approach of linearly combining input and then passing through fixed activation functions in standard neural net architectures. The connected network of nodes (variables) with probabilistic transducers on the edges comprises the Granger Network.

Fig. 2. Predictive Performance of Granger Nets. Panels a and b illustrate the out-of-sample area under the receiver operating characteristics curve (AUC) for predicting violent and property crimes respectively. The prediction is made a week in advance, and the event is registered as a successful prediction if we get a hit within ± 1 day of the predicted date. Panel c illustrates the distribution of AUC on average, individually for violent and property crimes. Our mean AUC is close to 90%. Panels d-f show the influence Diffusion & Perturbation Space. If we are able to infer a model that is predicts event dynamics at a specific spatial tile (the target) using observations from a source tile Δ days in future, then we say the source tile is within the influencing neighborhood for the target location with a delay of Δ. Panel d illustrates the spatial radius of influence for 0.5, 1, 2 and 3 weeks, for violent (upper panel) and property crimes (lower panel). Note that the influencing neighborhoods, as defined by our model, are large and approach a radius of 6 miles. Given the geometry of the City of Chicago, this maps to a substantial percentage of the total area of urban space under consideration, demonstrating that crime manifests demonstrable long-range and almost city-wide influence. Panel e illustrates the extent of a few inferred neighborhoods at time delay of at most 3 days. Panel f illustrates the average rate of influence diffusion measured by number of predictive models inferred that transduce influence as we consider longer and longer time delays. Note that the rate of influence diffusion falls rapidly for property crimes, dropping to zero in about a week, whereas for violent crimes, the influence continues to diffuse even after three weeks.

Fig. 3. Estimating Bias. Panel a illustrates the distribution of economic hardship index [53]. Panels c, d, e and f suggest biased response to perturbations in crime rates. With a 10% increase in violent or property crime rates, we see an approximately a 30% decrease in arrests when averaged over the city. The spatial distribution of locations that experience a positive vs. negative change in arrest rate reveals a strong preference favoring high SES locations. If neighborhoods are doing better socio-economically, increased crime predicts increased arrests. A strong converse trend is observed in predictions for lower SES poor and disadvantaged neighborhoods, suggesting that under stress, wealthier neighborhoods drain resources from their disadvantaged counterparts. Panel b illustrates this more directly via a multi-variable regression, where hardship index is seen to make a strong negative contribution.

Fig. 4. Prediction of property and violent crimes across major US cities and dependence of perturbation response on socio-economic status of local neighborhoods. Panels a-f illustrate the AUCs achieved in six major US cities. These cities were chosen on the basis of the availability of detailed event logs in the public domain. All of these cities show comparably high predictive performance. Panel g illustrates the results obtained by regressing crime rate and perturbation response against SES variables (shown here for poverty, as estimated by the 2018 US census). We note that while crime rate typically goes up with increasing poverty, the number of events observed one week after a positive perturbation of 5-10% increase in crime rate is predicted to fall with increasing poverty. We suggest that this decrease is explainable by reallocation of enforcement resources disproportionately, away from disadvantaged neighborhoods in response to increased event rates, which leads to smaller number of reported crimes.

Fig. 5. Panel a shows the Predictive Accuracy Index (PAI) and the Prediction Efficiency Index (PEI) calculated for seven metropolitan cities. Panel b shows the comparison of PAI/PEI achieved by our approach (Granger Net) against the best performing teams in a recent crime forecast challenge hosted by the National Institute of Justice (NIJ) in 2017 (https://nij.ojp.gov/funding/real-time-crime-forecasting-challenge), where teams attempted to predict hotspots for five different crime categories over different horizons prospectively. Our approach outperforms the teams in all 120 but one category (highlighted).

Fig. 6. Direct observation of differential response of arrest rate changes with SES variables. Panel a(i) shows the increase in violent and property crimes from February to June/August, averaged over the rich (hardship index < 20) and poor neighborhoods (hardship index > 20), over 4 years from 2014 to 2017 (95% confidence bounds shown). While the average percentage increase in event rate from late winter to summer is more or less comparable across the city, panel a(ii) shows that the deviation of mean percentage change in arrest rate from the city-wide average varies with the average SES of the communities. The wealthier communities see an increase in arrest rate per unit event, while others experience a draw-down. Panel b(i) shows the correlation between the daily change in the number of arrests, and their 1-day delayed versions in neighboring communities with higher hardship indices (μ), vs the hardship index of the communities themselves. Panel b(ii) shows the correlation of μ with the average hardship index of neighboring communities, computed within community groups of similar SES. These results illustrate that in wealthier communities, higher the average hardship index of the neighbors, more negative the μ, whereas this effect vanishes and eventually reverses as communities themselves become poorer. The locations of the top two community clusters as per their average hardship indices is shown on the Chicago map.

References

[1] Bowers, K. J., Johnson, S. D. & Pease, K. Prospective hot-spotting: The future of crime mapping? *The British Journal of Criminology* **44**, 641–658 (2004).

[2] Chainey, S., Tompson, L. & Uhlig, S. The utility of hotspot mapping for predicting spatial patterns of crime. *Security Journal* **21**, 4–28 (2008).

[3] Fielding, M. & Jones, V. 'disrupting the optimal forager': Predictive risk mapping and domestic burglary reduction in trafford, greater manchester. *International Journal of Police Science & Management* **14**, 30–41 (2012).

[4] Mohler, G. O., Short, M. B., Brantingham, P. J., Schoenberg, F. P. & Tita, G. E. Self-exciting point process modeling of crime. *Journal of the American Statistical Association* **106**, 100–108 (2011).
[5] Mohler, G. O. *et al.* Randomized controlled field trials of predictive policing. *Journal of the American Statistical Association* **110**, 1399–1411 (2015).
[6] Poisson, S. D. *Probabilité des jugements en matière criminelle et en matière civile, précédées des règles générales du calcul des probabilitiés* (Bachelier, 1837).
[7] Du Sautoy, M. *The Creativity Code: Art and Innovation in the Age of AI* (Harvard University Press, 2020).
[8] Ferdinand, T. N. Demographic shifts and criminality: An inquiry. *The British Journal of Criminology* **10**, 169–175 (1970).
[9] Cohen, L. & Felson, M. Social change and crime rate trends: A routine activity approach. *American Sociological Review* **44**, 588–608 (1979). Cited By 4102.
[10] Cohen, L. E. Modeling crime trends: a criminal opportunity perspective. *Journal of Research in Crime and Delinquency* **18**, 138–164 (1981).
[11] Wang, X. & Brown, D. E. The spatio-temporal modeling for criminal incidents. *Security Informatics* **1**, 2 (2012).
[12] Liu, H. & Brown, D. E. Criminal incident prediction using a point-pattern-based density model. *International Journal of Forecasting* **19**, 603 – 622 (2003).
[13] Caplan, J. M., Kennedy, L. W., Barnum, J. D. & Piza, E. L. Crime in context: Utilizing risk terrain modeling and conjunctive analysis of case configurations to explore the dynamics of criminogenic behavior settings. *Journal of Contemporary Criminal Justice* **33**, 133–151 (2017).
[14] Kang, H. W. & Kang, H. B. Prediction of crime occurrence from multi-modal data using deep learning. *PLoS ONE* **12**, e0176244 (2017).
[15] Flaxman, S., Chirico, M., Pereira, P. & Loeffler, C. Scalable high-resolution forecasting of sparse spatiotemporal events with kernel methods: A winning solution to the nij “real-time crime forecasting challenge”. *The Annals of Applied Statistics* **13**, 2564–2585 (2019).
[16] Mohler, G. & Porter, M. D. Rotational grid, pai-maximizing crime forecasts. *Statistical Analysis and Data Mining: The ASA Data Science Journal* **11**, 227–236 (2018).
[17] Chattopadhyay, I. & Lipson, H. Abductive learning of quantized stochastic processes with probabilistic finite automata. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* **371**, 20110543 (2013).
[18] Mohri, M. *Weighted Finite-State Transducer Algorithms. An Overview*, 551–563 (Springer Berlin Heidelberg, Berlin, Heidelberg, 2004).
[19] Granger, C. W. J. Testing for causality: A personal viewpoint. *Journal of Economic Dynamics and Control* **2**, 329 – 352 (1980).
[20] Papachristos, A. V. & Bastomski, S. Connected in crime: the enduring effect of neighborhood networks on the spatial patterning of violence. *American Journal of Sociology* **124**, 517–568 (2018).
[21] Papachristos, A. V., Wildeman, C. & Roberto, E. Tragic, but not random: The social contagion of nonfatal gunshot injuries. *Social Science & Medicine* **125**, 139–150 (2015).
[22] Green, B., Horel, T. & Papachristos, A. V. Modeling contagion through social networks to explain and predict gunshot violence in chicago, 2006 to 2014. *JAMA internal medicine* **177**, 326–333 (2017).
[23] Kang, H.-W. & Kang, H.-B. Prediction of crime occurrence from multi-modal data using deep learning. *PloS one* **12**, e0176244 (2017).
[24] Stec, A. & Klabjan, D. Forecasting crime with deep learning. *arXiv preprint arXiv:1806.01486* (2018).
[25] Hannon, L. Neighborhood residence and assessments of racial profiling using census data. *Socius* **5**, 2378023118818746 (2019).
[26] Meyer, W. B. & Graybill, J. K. The suburban bias of american society? *Urban Geography* **37**, 863–882 (2016).
[27] Lipton, M. *et al. Why poor people stay poor: a study of urban bias in world development* (London: Canberra, ACT: Temple Smith; Australian National University Press, 1977).
[28] Sternlieb, G. & Jackson, K. T. Crabgrass frontier: The suburbanization of the united states. *Political Science Quarterly* **101**, 493 (1986).
[29] Duany, A., Plater-Zyberk, E. & Speck, J. Suburban nation: the rise of sprawl and the decline of the american dream. *Choice Reviews Online* **38**, 38–1251–38–1251 (2000).
[30] Lazare, D. *America's Undeclared War: What's Killing Our Cities and how to Stop it* (Harcourt, 2001).
[31] Young, I. M. *Inclusion and democracy* (Oxford University press on demand, 2002).
[32] Kaplan, M. S., Crespo, C. J., Huguet, N. & Marks, G. Ethnic/racial homogeneity and sexually transmitted disease: A study of 77 chicago community areas. *Sexually Transmitted Diseases* **36**, 108–111 (2009). URL https://doi.org/10.1097%2Folq.0b013e31818b20fa.
[33] SHERMAN, L. W., GARTIN, P. R. & BUERGER, M. E. Hot spots of predatory crime: Routine activities and the criminology of place*. *Criminology* **27**, 27–56 (1989).
[34] WOOLDREDGE, J. Examining the (ir)relevance of aggregation bias for multilevel studies of neighborhoods and crime with an example comparing census tracts to official neighborhoods in cincinnati*. *Criminology* **40**, 681–710 (2002).

[35] MEARS, D. P. & BHATI, A. S. No community is an island: The effects of resource deprivation on urban violence in spatially and socially proximate communities*. *Criminology* **44**, 509–548 (2006).

[36] Weisburd, D., Groff, E. R., Yang, S.-M. & Telep, C. W. *Criminology of Place*, 848–857 (Springer New York, New York, NY, 2014).

[37] Small, M. L. Four reasons to abandon the idea of "the ghetto". *City & community* **7**, 389–398 (2008).

[38] Baumgarten, M. Ghetto: The invention of a place, the history of an idea. *Jewish Quarterly* **63**, 62–63 (2016).

[39] Heaven, W. D. Predictive policing algorithms are racist. they need to be dismantled. *MIT ZTechnology Review* **17**, 2020 (2020).

[40] Brayne, S. & Christin, A. Technologies of crime prediction: The reception of algorithms in policing and criminal courts. *Social Problems* (2020).

[41] St. Louis, S. & Greene, J. R. Social context in police legitimacy: giving meaning to police/community contacts. *Policing and Society* **30**, 656–673 (2020).

[42] Weisburd, D. Place-based policing. *Ideas in American Policing* 1–16 (2008).

[43] Kushnick, L. 'over policed and under protected': Stephen lawrence, institutional and police practices. *Sociological research online* **4**, 156–166 (1999).

[44] Sutherland, E. H. Juvenile delinquency and urban areas: A study of rates of delinquents in relation to differential characteristics of local communities in american cities. clifford r. shaw , henry d. mckay , norman s. hayner , paul g. cressey , clarence w. schroeder , t. earl sullenger , earl r. moses , calvin f. schmid. *American Journal of Sociology* **49**, 100–101 (1943).

[45] Sampson, R. J., Raudenbush, S. W. & Earls, F. Neighborhoods and violent crime: A multilevel study of collective efficacy. *Science* **277**, 918–924 (1997).

[46] Miethe, T. D., Hughes, M. & McDowall, D. Social Change and Crime Rates: An Evaluation of Alternative Theoretical Approaches*. *Social Forces* **70**, 165–185 (1991).

[47] Braga, A. A. & Clarke, R. V. Explaining high-risk concentrations of crime in the city: Social disorganization, crime opportunities, and important next steps. *Journal of Research in Crime and Delinquency* **51**, 480–498 (2014).

[48] Silver, D. & Clark, T. *Scenescapes: How Qualities of Place Shape Social Life* (University of Chicago Press, 2016).

[49] Nathan, R. P. & Adams, C. F. Four perspectives on urban hardship. *Political Science Quarterly* **104**, 483–508 (1989).

[50] Granger, C. W. J. Testing For Causality. *Journal of Economic Dynamics and Control* **2**, 329–352 (1980).

[51] Montero-Manso, P. & Hyndman, R. J. Principles and algorithms for forecasting groups of time series: Locality and globality. *International Journal of Forecasting* **37**, 1632–1653 (2021).

[52] Seabold, S. & Perktold, J. Statsmodels: Econometric and statistical modeling with python. In *Proceedings of the 9th Python in Science Conference*, vol. 57, 61 (Austin, TX, 2010).

[53] Laxy, M., Malecki, K. C., Givens, M. L., Walsh, M. C. & Nieto, F. J. The association between neighborhood economic hardship, the retail food environment, fast food intake, and obesity: findings from the survey of the health of wisconsin. *BMC Public Health* **15**, 1–10 (2015).

Inventory of Supporting Information

Manuscript #: NATHUMBEHAV-210114143A

Corresponding author name(s): Ishanu Chattopadhyay

1. Extended Data

Figure #	Figure title	Filename	Figure Legend
Extended Data Fig. 1	Out of Sample Predictive Performance over the Years	ED_Fig1.pdf	We show that the predictive performance is very stable, and variation in mean AUC is limited to the third place of decimal, at least when analyzing the last few years (4 years shown).
Extended Data Fig. 2	Comparison of Predicted vs Actual Sample Paths in Time and Frequency Domains	ED_Fig2.pdf	Panels a, c and e show that the predicted and actual sample paths are pretty close for different years, when compared over the first 150 days of each year. Panels b, d and f show that the Fourier coefficients match up pretty well as well. More importantly, while our models do not explicitly incorporate any periodic elements that are being tuned, we still manage to capture the weekly, (approximately) biweekly and longer periodic regularities.
Extended Data Fig. 3	Perturbation Effects	ED_Fig3.pdf	We see that the decrease of violent crimes from increase of property crimes are localized in disadvantaged neighborhoods

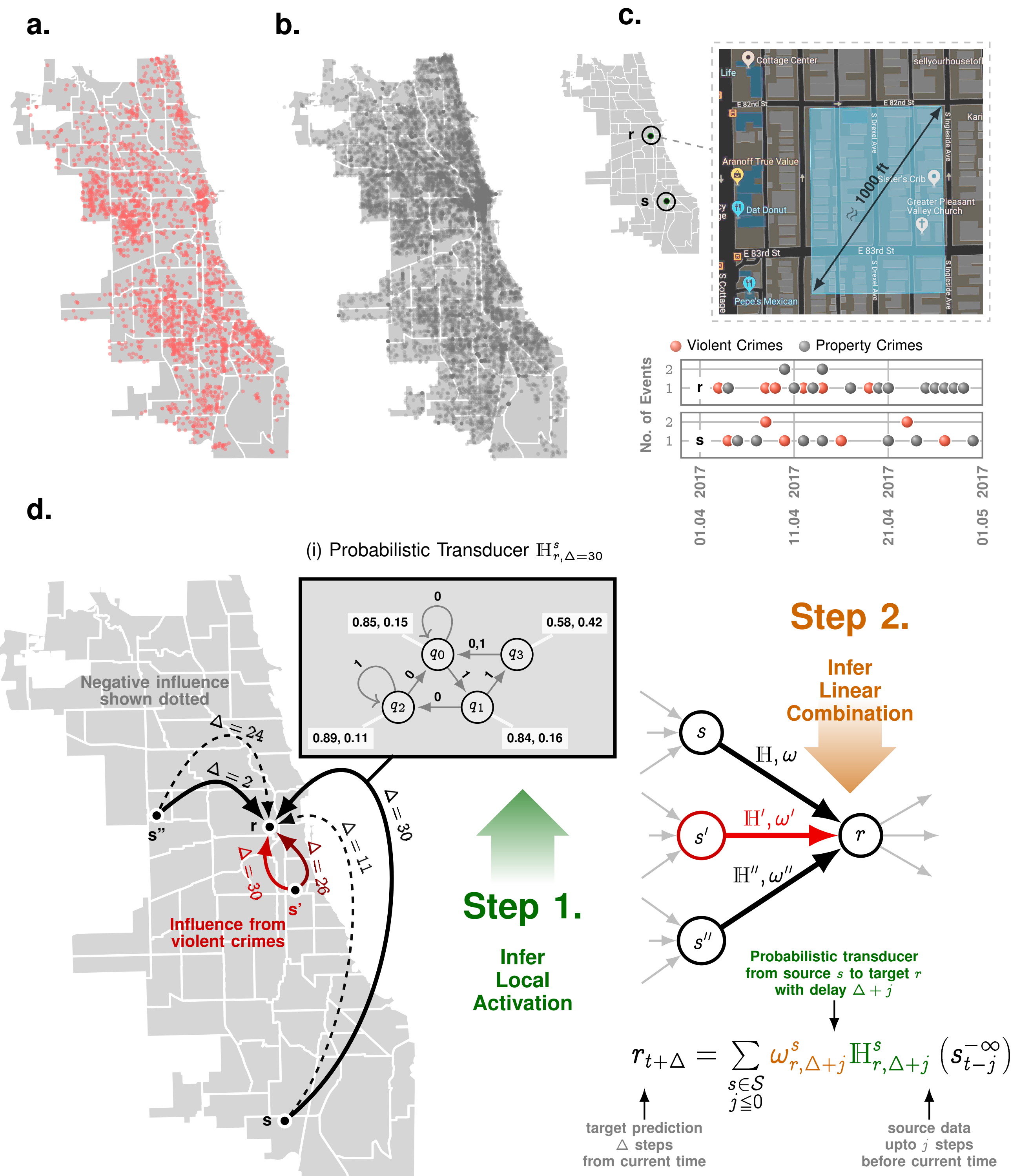
	Across Variables.		(panel g). Similarly, the decrease of property crimes from increase of violent crimes is also localized to disadvantaged neighborhoods (panel a), as well as the decreased violent crimes from increased arrests (panel k). We see a weaker localization for the corresponding increases in crime rates under similar perturbations. Looking at other pairs of variables under perturbation (rest of the panels), we generally do not see a very prominent correspondence with the distribution of socio-economic indicators. It seems crimes (and particularly violent crimes) are easier to dampen in locales with high existing crime rates, which is desirable result. But such conclusions are currently confounded by SES variables, and further work is needed to investigate these effects more thoroughly.
Extended Data Fig. 4	Stability of Suburban Bias over Years (Violent Crimes).	ED_Fig4.pdf	We show that the nature of the perturbation response shown in Fig. 3 holds true for earlier years as well: panels a and b correspond to year 2014, c and d correspond to 2015 and e and f correspond to year 2016, all of which follow the same pattern shown in Fig. 3.
Extended Data Fig. 5	Stability of Suburban Bias over Years (Property Crimes)	ED_Fig5.pdf	We show that the nature of the perturbation response shown in Fig. 3 holds true for earlier years as well: panels a and b correspond to year 2014, c and d correspond to 2015 and e and f correspond to year 2016, all of which follow the same pattern shown in Fig. 3.
Extended Data Fig. 6	Automatic Neighborhood Decomposition Using Event Predictability	ED_Fig6.pdf	Using Event Predictability Computing a bi-clustering on the source-vs-target influence matrix (panel A) isolates a set of spatial tiles that are, on average, good predictors for all other tiles. Using this set, we use a Voronoi decomposition of the city (Panel B), which realizes an automatic spatial

			decomposition of the urban space, driven by event predictability.
Extended Data Table 1	Prediction Statistics for Portland	ED_table1.tiff	Prediction Statistics for the City of Portland, USA
Extended Data Table 2	Naive baseline results: mean AUC achieved with ARIMA models	ED_table2.tiff	Naive baseline results: mean AUC achieved with ARIMA models

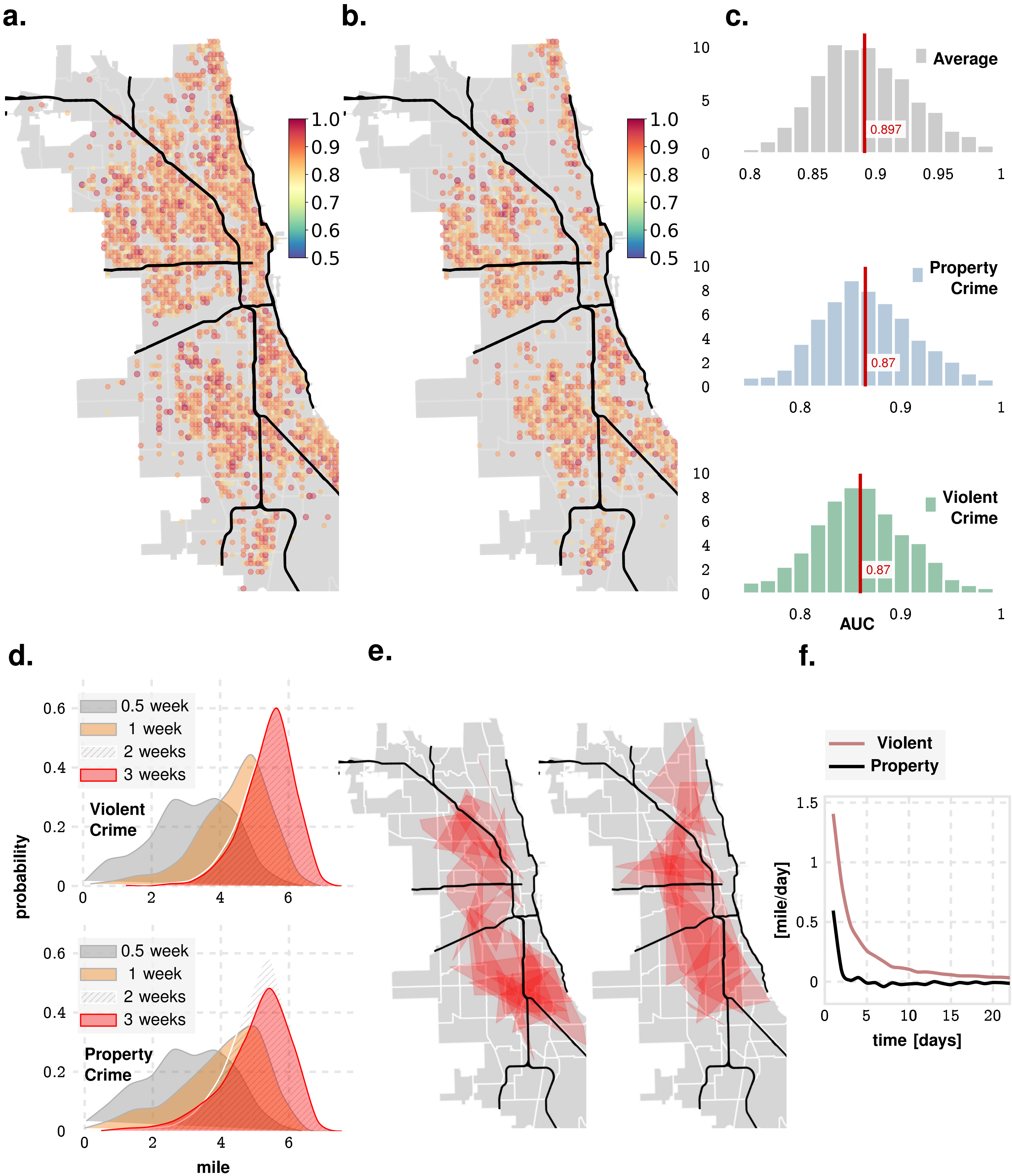
2. Supplementary Information:

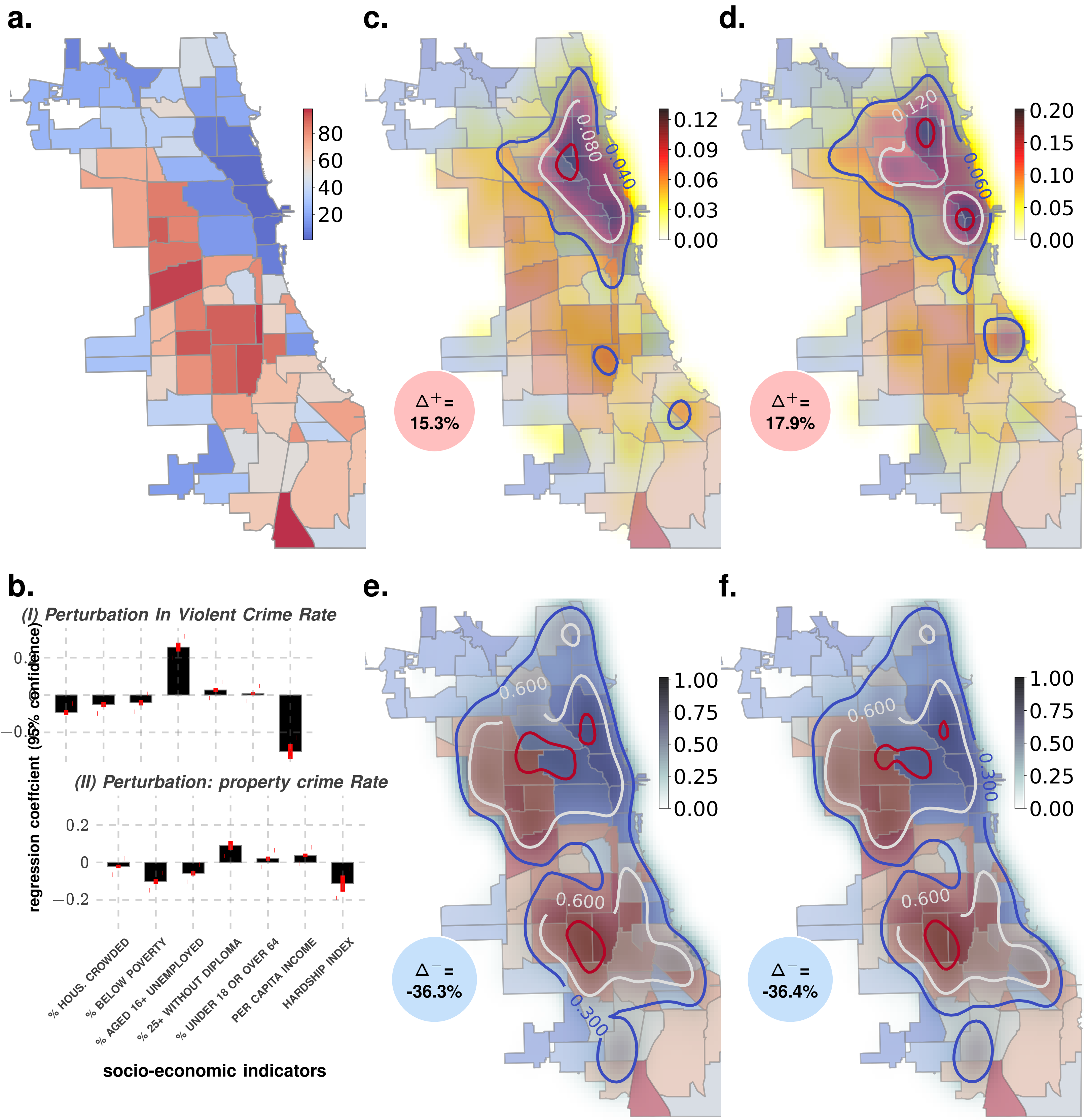
A. Flat Files

Item	Present?	Filename	A brief, numerical description of file contents.
Supplementary Information	Yes	SI_NHB.pdf	*Supplementary Figure 1, Supplementary Methods*
Reporting Summary	Yes	nr-reporting-summary.pdf	
Peer Review Information	No		

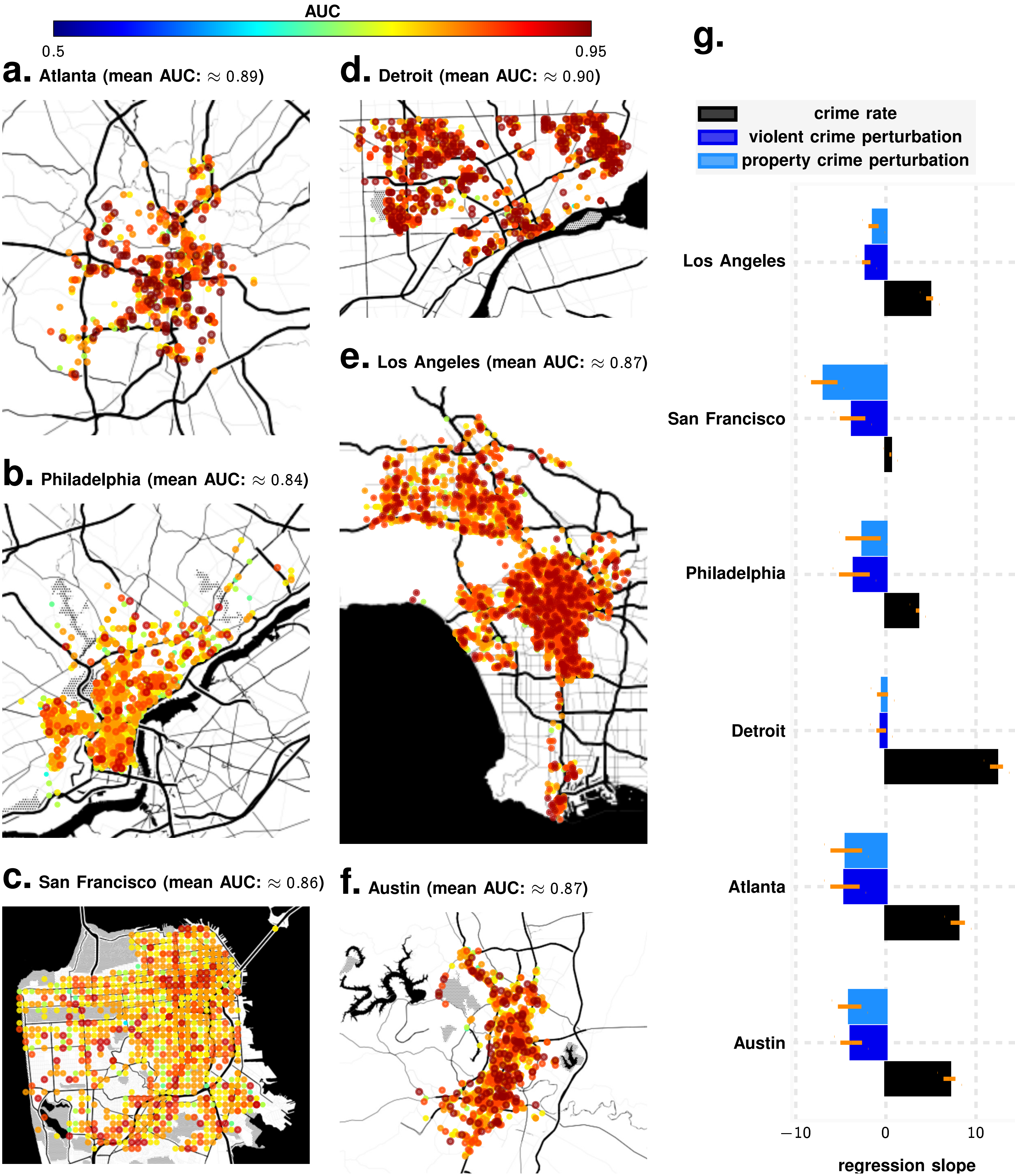


a.
b.
c.
r
s
Cottage Center
sellyourhousetof
Life
E 82nd St
Aranoff True Value
Sister's Crib
Greater Pleasant Valley Church
Dat Donut
1000 ft
S Drexel Ave
S Ingleside Ave
E 83rd St
Pepe's Mexican
Violent Crimes
Property Crimes
No. of Events
2
1
01.04 2017
11.04 2017
21.04 2017
01.05 2017
d.
(i) Probabilistic Transducer $\mathbb{H}^s_{r,\Delta=30}$
0
0.85, 0.15
0.58, 0.42
0,1
q_0
q_3
q_2
q_1
1
0
0.89, 0.11
0.84, 0.16
Negative influence shown dotted
$\Delta = 24$
$\Delta = 2$
$\Delta = 30$
$\Delta = 11$
$\Delta = 26$
s''
s'
Influence from violent crimes
Step 1.
Infer Local Activation
Step 2.
Infer Linear Combination
s
s'
s''
r
$\mathbb{H}, \omega$
$\mathbb{H}', \omega'$
$\mathbb{H}'', \omega''$
Probabilistic transducer from source s to target r with delay $\Delta + j$
$r_{t+\Delta} = \sum_{s \in \mathcal{S}, j \leq 0} \omega^s_{r,\Delta+j} \mathbb{H}^s_{r,\Delta+j} \left(s^{-\infty}_{t-j} \right)$
target prediction Δ steps from current time
source data upto j steps before current time

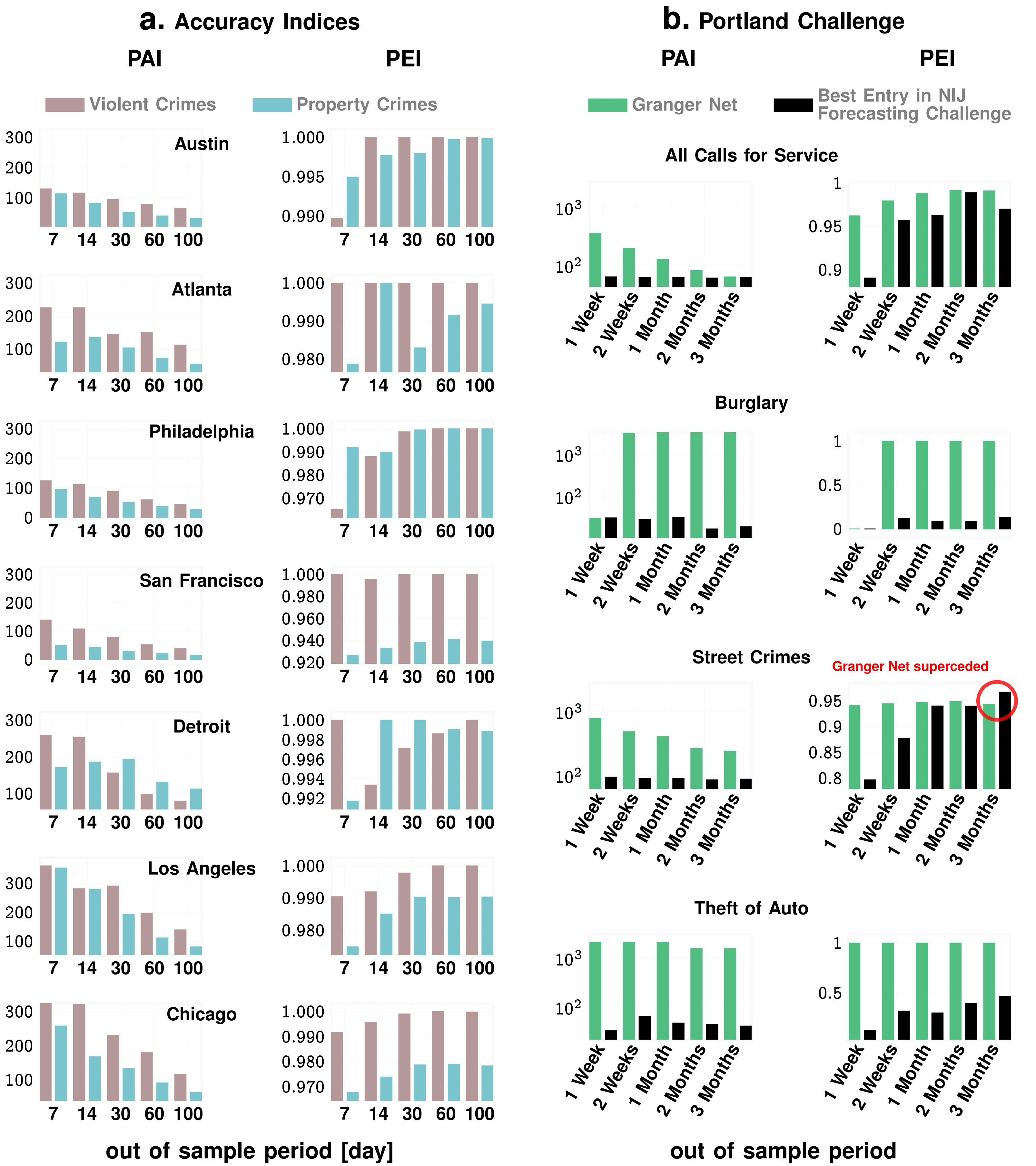

a.
b.
c.
1.0
0.9
0.8
0.7
0.6
0.5
Average
0.897
Property Crime
0.87
Violent Crime
0.87
AUC
d.
e.
f.
0.5 week
1 week
2 weeks
3 weeks
Violent Crime
Property Crime
probability
mile
Violent
Property
[mile/day]
time [days]



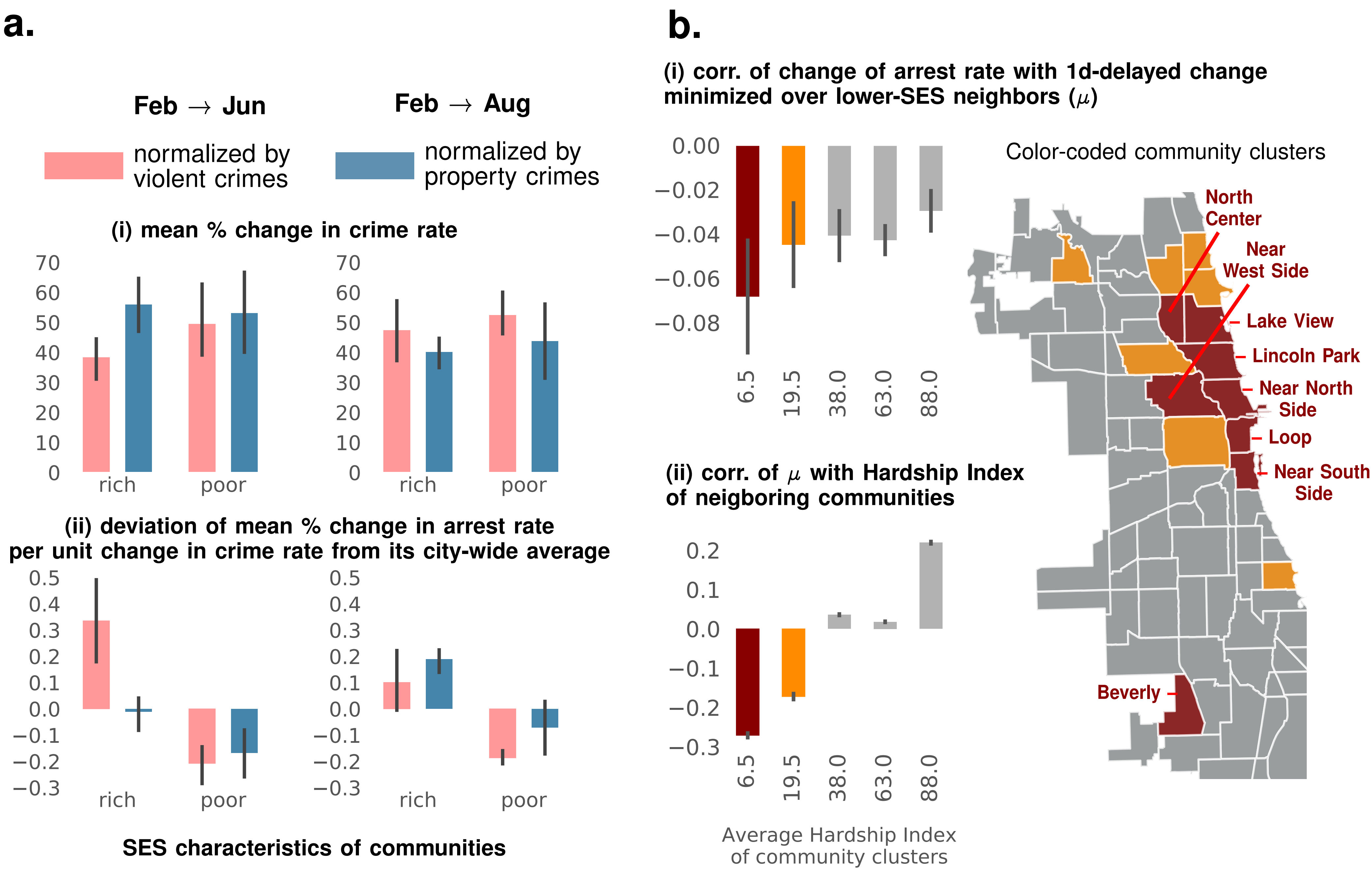
a.
80
60
40
20
b.
(I) Perturbation In Violent Crime Rate
(II) Perturbation: property crime Rate
regression coeffcient (95% confidence)
0.2
0
−0.2
% HOUS. CROWDED
% BELOW POVERTY
% AGED 16+ UNEMPLOYED
% 25+ WITHOUT DIPLOMA
% UNDER 18 OR OVER 64
PER CAPITA INCOME
HARDSHIP INDEX
socio-economic indicators
c.
0.12
0.09
0.06
0.03
0.00
0.080
0.040
Δ+=
15.3%
d.
0.20
0.15
0.10
0.05
0.00
0.120
0.060
Δ+=
17.9%
e.
1.00
0.75
0.50
0.25
0.00
0.600
0.300
Δ−=
-36.3%
f.
1.00
0.75
0.50
0.25
0.00
0.600
0.300
Δ−=
-36.4%



AUC
0.5
0.95
a. Atlanta (mean AUC: ≈ 0.89)
b. Philadelphia (mean AUC: ≈ 0.84)
c. San Francisco (mean AUC: ≈ 0.86)
d. Detroit (mean AUC: ≈ 0.90)
e. Los Angeles (mean AUC: ≈ 0.87)
f. Austin (mean AUC: ≈ 0.87)
g.
crime rate
violent crime perturbation
property crime perturbation
Los Angeles
San Francisco
Philadelphia
Detroit
Atlanta
Austin
−10
0
10
regression slope



a. Accuracy Indices
PAI
PEI
Violent Crimes
Property Crimes
Austin
Atlanta
Philadelphia
San Francisco
Detroit
Los Angeles
Chicago
out of sample period [day]
b. Portland Challenge
PAI
PEI
Granger Net
Best Entry in NIJ Forecasting Challenge
All Calls for Service
Burglary
Street Crimes
Granger Net superceded
Theft of Auto
1 Week
2 Weeks
1 Month
2 Months
3 Months
out of sample period


a.
Feb → Jun
Feb → Aug
normalized by violent crimes
normalized by property crimes
(i) mean % change in crime rate
(ii) deviation of mean % change in arrest rate per unit change in crime rate from its city-wide average
rich
poor
SES characteristics of communities
b.
(i) corr. of change of arrest rate with 1d-delayed change minimized over lower-SES neighbors (μ)
(ii) corr. of μ with Hardship Index of neigboring communities
Average Hardship Index of community clusters
6.5
19.5
38.0
63.0
88.0
Color-coded community clusters
North Center
Near West Side
Lake View
Lincoln Park
Near North Side
Loop
Near South Side
Beverly