How to prevent crimes using earthquakes

Marco Abate

1 *Minority Report*

In 2015, the Fox channel produced a TV series called *Minority Report,* created by Max Borenstein with Steven Spielberg among the executive producers. The series was sort of a sequel to the 2002 movie *Minority Report* by Steven Spielberg; the latter was in turn inspired by a short story written by Philip K. Dick in 1956 whose title was (as you can guess) *Minority Report.*

The main point of the short story, the movie and (somewhat less) the series was to explore the ethical implications of the possibility of predicting violent crimes *before* their actual occurrence. In a not-too-far future two brothers and one sister are born with the uncanny (and disturbing) ability of seeing violent crimes before they happen. The police uses their visions to identify the culprits and arrest them *before* the crimes are actually committed, thus preventing their occurrence and consequently saving lives.

Already here we have an ethical problem, because the alleged culprits are arrested and sentenced *without* having committed the crime they are accused of; but, on the other hand, the potential victims keep living their lives without being harmed, which supposedly is a good thing by itself. Moreover, often enough it was possible to find proofs that the alleged culprits were actually preparing the crimes the three siblings (the *pre-cogs*) predicted, thus justifying the arrests also from a more conventional point of view.

It turns out that the predictions of the pre-cogs are stronger and more accurate when the three siblings are kept together immersed in a futuristic milk bath (don't ask). So they are effectively segregated and imprisoned by the police, forced to stay as much as possible in the milk bath without any personal life outside. It is a small price to pay to have a crime-free city, isn't it?

Marco Abate
Dipartimento di Matematica, Università di Pisa, Largo Pontecorvo 5, 56127 Pisa, Italy, e-mail: marco.abate@unipi.it

And yet, predicting the future is a tricky business, in particular when the act of prediction changes the future thus invalidating those very same predictions. It is then not surprising that in rare cases the three siblings might not agree. Two of them predict a particular culprit for a soon-to-happen violent crime, but the third one does not agree; there is a *minority report.* From a statistical point of view, this is a rare event; the potential error is negligible for the society considered as a whole. But from the point of view of the individuals directly involved, the potential error might be devastating; both acting and not acting can lead to the destruction of somebody's life.

These are the ethical points making the reading of the short story and the vision of the movie compelling (the TV series went in a slightly different direction, ending up cancelled after only one season), and they can be summarised in the old and fundamental conundrum underlying any democracy: up to which point we are ready to go in limiting the freedom of the individual for the sake of the safety of the whole society? Are we willing to limit the freedom of individuals because, according to tools that might be mistaken, we *think* they might commit a crime? If you believe this is just a science-fiction problem and thus it does not matter, start thinking about how religion profiling might be used to allegedly prevent terrorism.

However, this is ***not*** what this note is about. Or is it?

2 Preventing vs. predicting

In 2011, a start-up based in California earned surprising titles in the first pages of many newspapers and online news sites. One site devoted to technological news, *SingularityHub.com*, summarised the hubbub quite efficiently with the following headline:

> *Pre-Cog is here — A new software stops crime before it happens* (SingularityHub.com, August 29, 2011)

The article, written by Peter Murray (see [1]), reported on a software developed by a small company called *PredPol* that was being tested in California by Santa Cruz and Los Angeles Police Departments to prevent crimes from happening.

The journalist of course quoted *Minority Report* (and the TV series *Numb3rs*), but there is a fundamental difference in the way the algorithm devised by PredPol works with respect to the way the pre-cogs worked in the movie: the aim of the algorithm is to *prevent* crimes, not to *predict* crimes. More precisely, the aim of the algorithm is to identify areas where a particular kind of crime is more likely to occur, without giving any information on who will possibly commit a crime there.

This is explained in the web page of the company [2]. After presenting itself as *The Predictive Policing Company,* the company states that "PredPol uses artificial intelligence to help you prevent crime by predicting when and where crime is most likely to occur, allowing you to optimize patrol resources and measure effectiveness.” In another page of the site it is clarified what this algorithm does and what it does not:

PredPol does:

- Increase Law Enforcements odds of stopping crime
- Predict *where* and *when* crime is most likely to occur
- Work for both large and small agencies
- Help new officers on-board quicker
- Make it easy to access predictions anywhere and anytime

PredPol does not:

- Map out past crimes or another ‘hotspot’ tool [we shall discuss hotspots later]
- Predict *who* will commit crimes
- Use PII (Personally identifiable information), eliminating civil liberties issues
- Replace law enforcement veterans or analysts intuition and experience
- Require additional hiring or new hardware

More precisely, the aim of the PredPol algorithm is to suggest to the police force *where and when* to send patrols, with the idea that police presence in the right place at the right time will prevent crimes from happening. Nobody specific is targeted in advance; only the locations are indicated, and updated in (almost) real time.

Sending patrols in high-risk areas has always been standard police procedure. To explain in which sense the PredPol algorithm is different, let’s first take a look to more standard approaches used by police departments in major cities.

Traditionally, the decisions about where and when to send patrols were taken by senior officers relying on their own experience and intuition. This can be very effective or completely ineffectual depending on the officer; and discontinuities are anyway created by the retirement of a senior officer in charge, with the consequent loss of his/her experience.

In the last forty years or so, using the huge amount of data collected by law enforcements, hotspot maps have been devised and have become a standard tool to help less experienced officers to decide where to send patrols, and that might provide an approach more consistent and less depending on the specific senior officer in charge.

A *hotspot map* is a map of a (given area of) a city showing (usually using colours or shades of grey; see Fig.1) the probability of occurrence of a specific type of crime in a particular place, sometimes also taking into account the specific time of day (or night). In standard hotspot maps the probability is computed in a very simple (and naive) way: it is simply the quotient between the number of crimes of the specified type occurred in the given place at the specified time and the total number of crimes of the specified type occurred in the city at the specified time. This is a computation that can be done as soon as enough data are accumulated, and can be updated any time new data are acquired. In other words, crime hotspot maps are frequency maps based on past occurrences of a specified typology of crime; and are used by sending police patrols more often in areas with higher probability of occurrence of crimes (that is, in areas where that kind of crime has occurred more often).

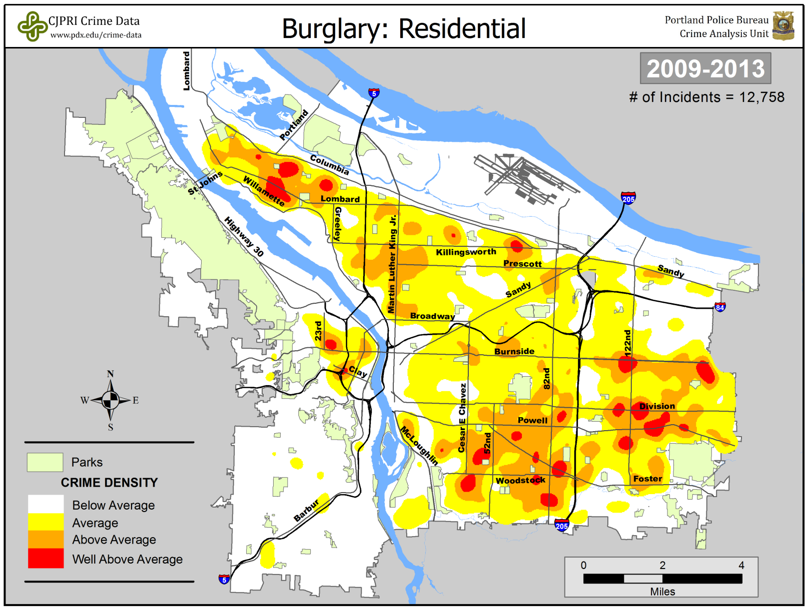

Fig. 1 Hotspot map of residential burglaries in Portland in the years 2009-2013. (Source: Henning, K., (2017, November 1). Residential burglary hotspot. Retrieved from [3])

As a quick google search will show, there exist many kinds of crime hotspot maps (see [4] for hotspot maps of crime in Italy), but most of them can be reduced to two categories: long-term maps, and short-term maps.

Long-term maps are based on data collected on a span of several years, sometimes decades. Being based on large amounts of data, they are fairly stable; this is both a strength and a weakness. It is a strength because they are robust; for instance, occasional mistakes in classifying past occurrences of crimes have a very limited effect on long-time hotspot maps. By the same token, they are slow to change; one needs many new data to significantly change a long-time hotspot maps. As a consequence, they are not able to detect recent or transient phenomena. Moreover, they tend to suggest patterns of police patrolling mostly constant in time, and thus foreseeable by criminals that then adapt their behaviour to escape being spotted by the patrols. Of course such changes in behaviour will eventually be detected by long-term hotspot maps; but this needs time, and thus limits the utility of such maps.

Short-term maps are instead based only on more recent data, and are used to try and detect recent trends and behaviours. They complement the information provided by long-term maps, giving visibility to phenomena that would otherwise have remained hidden; but being based on small amounts of data they are unstable, and thus more susceptible to mistakes in classification. Furthermore, their predictions change fast, sometimes too fast to be useful.

Two examples of quickly developing crime trends can help understanding the difference between long-term and short-term hotspot maps.

The first example concerns home burglaries in suburban areas. It is unfortunately not uncommon that usually quiet suburban areas are subjected to an unusual number of home burglaries concentrated in a short amount of time. What has happened is that a burglar has discovered that that area is poorly guarded, and thus (alone or with accomplices) has burglarised as many houses he could in the shortest possible amount of time, before the inevitable increase in police patrolling that would shortly follow. Thus we have a spike (high in magnitude, short in time) of burglaries there.

The second example concerns gang-related crimes. An incident between members of rival gangs may lead to a sequence of retaliations ending only when some sort of (fragile) truce is reached — or enough gang members are injured or dead. Again, we have a spike of incidents, even though this time in an area already indicated as potentially critical by long-term hotspot maps.

Spikes in crimes tend to make headlines more than the usual crimes; and this makes police departments even more interested in trying to prevent them. Short-term hotspot maps are used to try and identify spikes when they are starting to happen; but clearly they are a very rough tool.

So one needs a way to merge long-term maps and short-term maps in a single tool able to keep into account both history and transient trends; and to do so some mathematics can be useful.

3 Earthquakes and crimes

This is the problem a group of researchers at UCLA started to study around the year 2009. It was quite an interdisciplinary group, comprising mathematicians (A. Bertozzi, M.B. Short and G.O. Mohler, who later moved to Santa Clara University where he refined and expanded the models), a statistician (F.P. Schoenberg), an anthropologist (P.J. Brantigham) and a criminologist (G.E. Tita), with the fundamental assistance of experienced police officers, W. J. Bratton from LAPD and Z. Friend from Santa Cruz PD.

The starting point was the idea that to provide an effective model of crime occurrences one needs to keep into account two concurrent aspects:

- a ground state of likelihood of crimes mostly due to the (social, urban and cultural) environment, slowly changing with time;
- spikes of events, short in time but possibly strong in magnitude, often sparked by single incidents.

Possibly because they were living in California, the group realised that there was another type of phenomena presenting the same kind of concurrent aspects: earthquakes.

Indeed, the likelihood of an earthquake in a particular zone is mostly due to the geological structure of the local terrain, and in particular to the characteristics and

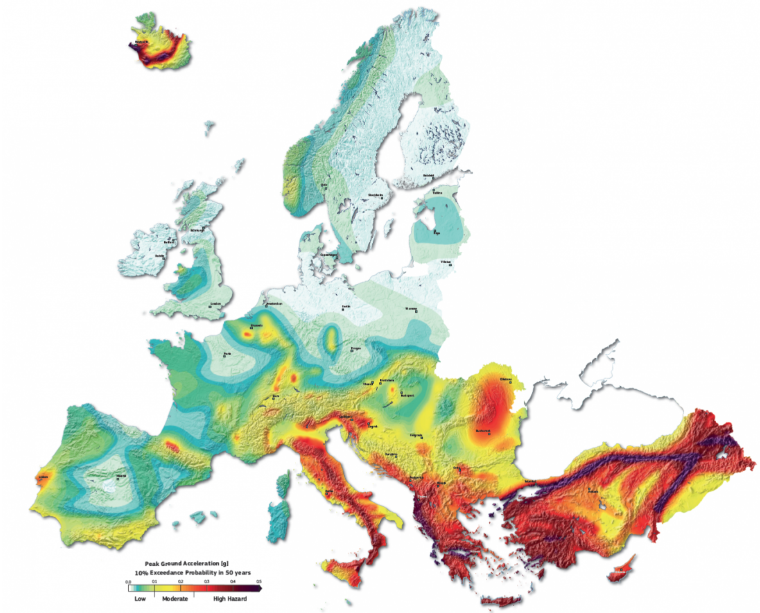

Fig. 2 European seismic hazard map. (Source: SHARE project [5])

positions of faults in the area. This gives a background probability, slowly changing over (geological) time. But it is also very well-known that a single event very often causes swarms of events (aftershocks) concentrated in a (relatively) short time span. Thus we have coexistence of slowly-changing and quickly-happening causes of events, in a way very similar to what happens with crime — and, not surprisingly, a typical way for visualising the likelihood of earthquakes is by hazard maps (see Fig. 2), which is the way earth scientists call hotspot maps.

Mostly led by Mohler, the group started to adapt the models used by geologists for computing the probability of earthquakes so that they could be used for computing the probability of crimes. It turned out that from this point of view crimes had a number of advantages over earthquakes: the amount of data available is much larger, the time frames to consider are much shorter and the events present much smaller variations in magnitude, all elements making the construction of probabilistic models easier.

The results of their studies were first published in the paper [6] "Self-Exciting Point Process Modeling of Crime", published online by the Journal of the American Statistical Association on January 01, 2012 (the paper was originally submitted to the journal more than two years before, on September 2009, and accepted after revision on October 2010). Many other papers followed, with different applications and/or generalizations of the original model (see, e.g., [7, 8, 9, 10]). More to the point, the models have been applied to real data provided by the police departments of Los Angeles and Santa Cruz, and then have been used for field tests.

The tests consisted in deciding when and where to dispatch patrols using in some districts the models provided by Mohler's group, while using more traditional methods in other districts, and then comparing the outcomes. The tests gave promising results, and this convinced Mohler and his collaborators to start the spin-off company PredPol to commercialise a software based on the mathematical models they developed.

4 The model

From a mathematical point of view, the model is not too complicated. The probability $p(x,y,t)$ that a crime (of a given kind) will occur at the position of coordinates (x,y) at the time t is represented by a function of the form

$$p(x,y,t) = G(x,y) + \sum_{t_j<t} g(x-x_j, y-y_j, t-t_j) \; . \tag{1}$$

In this formula $G(x,y)$ represents the background probability of occurrence of a crime at (x,y), and corresponds to the contribute given by long-term hotspot maps. The sum takes instead into account the probability that a given crime will spark a swarm of similar crimes; indeed, $g(x-x_j, y-y_j, t-t_j)$ represents the probability that a crime occurred at (x_j, y_j) at the time t_j will cause another similar crime at (x,y) at the time $t > t_j$, and the sum ranges over all past crimes.

Formula (1) is very general; suitably choosing the functions G and g it can be adapted to model many different kinds of phenomena. It turned out that for modelling crimes (as well as for modelling earthquakes) it is not necessary to be very creative in the choice of G and g, and it is sufficient to start from one of the most used functions in probability theory, the Gaussian function

$$G_\sigma(x,y) = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{x^2+y^2}{2\sigma^2}\right) \; ,$$

where σ is a normalization factor. Fig. 3 contains the graph of G_σ for two different values of σ. The smaller is σ the faster G_σ decreases away from the origin $(0,0)$.

The function G should take into account the history of crimes in the area, and in particular the fact that a crime occurred at (x_j, y_j) might influence the later occurrence of crimes in (x,y). Clearly, the influence should decrease when (x,y) is far from (x_j, y_j). This suggested the following choice for G:

$$G(x,y) = \frac{a}{2\pi\sigma^2} \sum \exp\left(-\frac{(x-x_j)^2 + (y-y_j)^2}{2\sigma^2}\right) \; , \tag{2}$$

where a and σ are parameters to be chosen, and the sum ranges over all past occurrences of crimes.

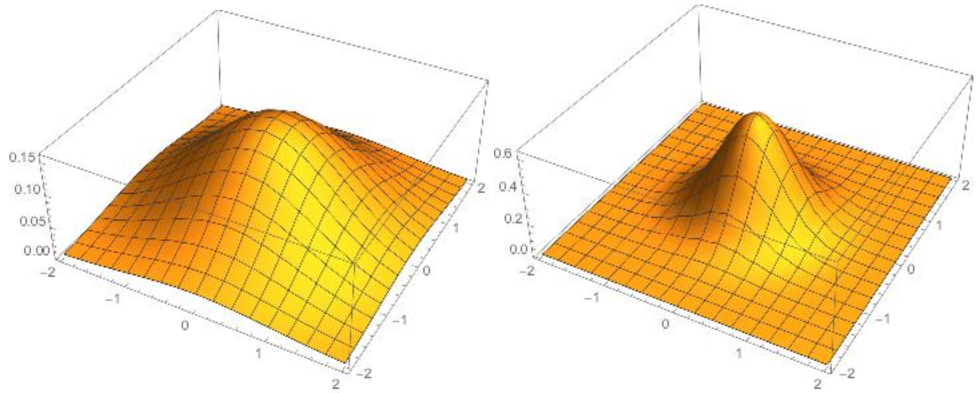

Fig. 3 Graph of G_σ for $\sigma = 1$ (left) and $\sigma = 1/2$ (right).

The function g should also take into account the time passed from the occurrence of the given crime, decreasing as the interval of time increases. Since one would like to model spikes concentrated in time without long-lasting effects, the time dependence should be represented by a quickly decreasing function, possibly exponentially decreasing. This suggested the following form for the function g:

$$g(x-x_j, y-y_j, t-t_j) = b\gamma \exp\bigl(-\gamma(t-t_j)\bigr)\frac{1}{2\pi\tau^2}\exp\left(-\frac{(x-x_j)^2+(y-y_j)^2}{2\tau^2}\right), \tag{3}$$

where b, γ and τ are parameters to be chosen.

The model then depends on five parameters: a, b, σ, τ and γ. The parameters σ and τ control the spatial influence of past crimes, the parameter γ controls the time influence of past crimes, and the parameters a and b can be used to control the relative importance of the background crime probability (corresponding to long-term maps) with respect to the effects of recent crimes (corresponding to short-term maps).

Up to here the mathematics is relatively elementary. The difficult part is the *calibration* of the model, that is the choice of the parameters so that at the time t_0 when the computation is done the probability distribution p given by equation (1) is as close as possible to the actual probability distribution computed by using the frequency of past crimes as we described above for hotspot maps. This is done by using non trivial techniques coming from statistics and numerical analysis.

Once the model is calibrated (that is, the parameters are chosen), $p(x,y,t)$ will give the probability of occurrence of a crime at the position (x,y) at the time $t > t_0$, thus suggesting to send patrols at time t in the places where $p(x,y,t)$ is higher. Of course, one expects the accuracy of the prediction to decrease with time; thus one needs to repeat every so often the calibration by including the new crimes occurred since the previous calibration. And this introduces another difficulty: the calibration algorithms should be fast enough to run as close as possible to real time.

Not surprisingly, the scientific papers do not describe in detail the calibration algorithms used (and as a consequence we shall not talk about them here, referring to [6, 7, 8, 9, 10] for some information). Indeed, it is the efficiency of the algorithms to determine whether a model is commercially viable; and to be too public about the details will just help the competition. Furthermore, the scientific papers describe the general approach only; the actual implementation might include tweaks requested by a single customer and valid only for that specific kind of crimes in that specific kind of area. Indeed, a specific model might need to take into account, e.g., geographical features; for instance, the probability of a home burglary in the middle of a lake is zero independently of the distribution of past home burglaries along the shores of the lake.

Right now there are several companies selling predictive policing algorithms (beside PredPol, one can quote at least HunchLab [11] and CivicScape [12]). The competition is strong, and this explains why it is difficult to find details of the actual implementation of such models (there is a notable exception: CivicScape recently released publicly its software; see below). Furthermore, the use of such algorithms is becoming popular; the example set by Mohler and his collaborators is spreading fast.

5 Final thougths

A lot of research is currently going on in this area. For instance, up to now we assumed that the calibration of the model is based on the past occurrences of the *same* kind of crime one would like to prevent. But this is a limitation: for instance, a street assault with knives can lead to retaliations with guns, and possibly to murders — or, conversely, a murder can cause less severe repercussions in the same area. Thus a more accurate model should take into account several kind of crimes, but with appropriate weights and possibly using different probability functions. See, e.g., [13] for a discussion.

One natural question now is: do these models actually work? As often happens, there is not a clear cut answer. According to most of their users, they do. Here is a typical quote (taken from the PredPol web site [14]):

> The Santa Cruz, CA Police Department saw assaults drop by 9%, burglaries decrease by 11%, and robberies down 27% in its first year using the software (2011-2012). Crime overall dropped 25% in June 2013 and 29% in July 2013 compared with those same months the previous year.

On the other hand, not everybody is convinced (see, e.g., [15] and [16]). A criticism often raised is that these algorithms seem to target minorities too often, and this has led to advance doubts about racial bias included in the algorithms themselves. It is the need to dispel such doubts that has led CivicScape to the decision of making its software public [17], so that anybody interested could check whether this allegation would be true or not (of course, they claim that it is not true).

To better put these models in context there are two underlying assumptions that I think deserve to be made explicit.

The first one can be summarised as “crime begets crime”. The prediction of crimes in the future is based on the distribution of crimes in the past. Said more bluntly, these models do not believe in the possibility of reform: bad neighbourhoods will stay bad. This has nothing to do with the possibility of reforming single individuals; and indeed these algorithms will not predict the behaviour of an individual, they are not interested in who will commit a crime, but only in where and when. In reality, neighbourhoods can get better; but it is a slow process, and a process usually (but not always) started by an external influence, or anyway by an influence not represented by the distribution of past crimes. To circumvent this problem, more recent models have begun to include in the computations other factors too, ranging from community inputs to weather forecasts.

The second assumption can be summarised as ”crime is boring”. Being statistical models, they are geared toward predicting standard, average, repetitive (boring. . .) behaviours; they might miss anomalies or unprecedented incidents. In particular, they cannot be used to predict crimes of passion, or crimes due to mental illnesses — unless such crimes were more frequent in particular areas up to the point of becoming statistical meaningful.

An inevitable question in present day Europe is: can these models be used to predict terrorist attacks? At present (and hopefully forever) the answer is necessarily negative. Media exposure notwithstanding, terrorist attacks in Europe are quite rare (compared to home burglaries or car thefts, for instance); there are not enough data to meaningfully calibrate a statistical model. Terrorist attacks in Europe are isolated incidents, with a distribution closer to the one of crimes due to mental illnesses than to the one of more standard crimes, and with an even lower number of occurrences. Other models might in principle be used to prevent them, but they must be models based on different assumptions.

However, this applies to terrorist attacks in Europe. Terrorist attacks in other areas of the world, e.g., Iraq, Pakistan or the Middle East, are much more common, and they follow behaviours that are more standardised. Thus they might be amenable to being modelled with tools similar to the ones described here; if so we might rightfully say that mathematics can save lives.

References

[1] http://www4.stat.ncsu.edu/ reich/st733/Journal/Logan3.pdf
[2] http://www.predpol.com
[3] http://www.pdx.edu/crime-data/
[4] http://www.transcrime.it
[5] http://www.share-eu.org
[6] G.O. Mohler, M.B. Short, P.J. Brantigham, F.P. Schoenberg, G.E. Tita: Self-Exciting Point Process Modeling of Crime. *Journal of the American Statistical Association* **106**, 100–108 (2011)
[7] G.O. Mohler: Modeling and estimation of multi-source clustering in crime and security data. *The Annals of Applied Statistics* **7**, 1525–1539 (2013)
[8] M.B. Short, G.O. Mohler, P.J. Brantigham, G.E. Tita: Gang rivalry dynamics via coupled point process networks. *Discrete and Continuous Dynamical Systems, Series B* **19**, 1459–1477 (2014)
[9] J.T. Woodworth, G.O. Mohler, A.L. Bertozzi, P.J. Brantigham: Non-local crime density estimation incorporating housing information. *Philosophical Transactions of the Royal Society A* **372**:20130403 (2014)
[10] G.O. Mohler, M.B. Short, S. Malinowski, M. Johnson, G.E. Tita, A.L. Bertozzi, P.J. Brantigham: Randomized Controlled Field Trials of Predictive Policing. *Journal of the American Statistical Association* **110**, 1399–1411 (2015)
[11] http://www.hunchlab.com
[12] http://www.civicscape.com
[13] G.O. Mohler: Marked point process hotspot maps for homicide and gun crime prediction in Chicago. *International Journal of Forecasting* **30**, 491–497 (2014)
[14] http://www.predpol.com/results/
[15] http://www.nature.com/news/reform-predictive-policing-1.21338
[16] http://singularityhub.com/2017/02/02/could-predictive-policing-lead-to-a-real-life-minority-report/
[17] http://github.com/CivicScape/CivicScape