Inspira Crea Transforma

UNIVERSIDAD EAFIT®

Ph. D IN MATHEMATICAL ENGINEERING
DOCTORAL SEMINAR II

November 24, 2017

Forecasting of the crime with matching between murders and google trends

ANDRES PEREZ CORONADO
PhD Student

GUSTAVO CANAVIRE
HENRY LANIADO
Thesis advisors

Ph. D., main objective

To design a spatial-temporal econometric model to anticipate how criminal networks must be dismantling.

Ph. D., scientific first paper

Probabilistic model for foresight of organized crime, linked with social media.

Outline

1. Prediction model
   - ✓ Definition Kernel Density Estimation (KDE)
   - ✓ Latent Dirichlet Allocation (LDA)
2. Police Quadrants
   - ✓ National Model of Community Policing by Quadrants (MNVCC)
   - ✓ Murders statistic in Bogota D.C., Medellin and Cali
3. Google Trends and limitations
   - ✓ Queries and trends
   - ✓ Citizen security information

Prediction model

www.sciencedirect.com

Predicting crime using Twitter and kernel density estimation - ScienceDirect

Kernel Density Estimation (KDE)

$$f_1(p) = k\,(p,h) = \frac{1}{Ph}\sum_{j=1}^{P} K\left(\frac{\|p - p_j\|}{h}\right)$$

T= type of crime
p= spatial point
P= total number of crimes T
j= indexes a single crime location
K= standard normal density function
||.||= Euclidean norm
p_j= location of crime j

Kernel Density Estimation (KDE)

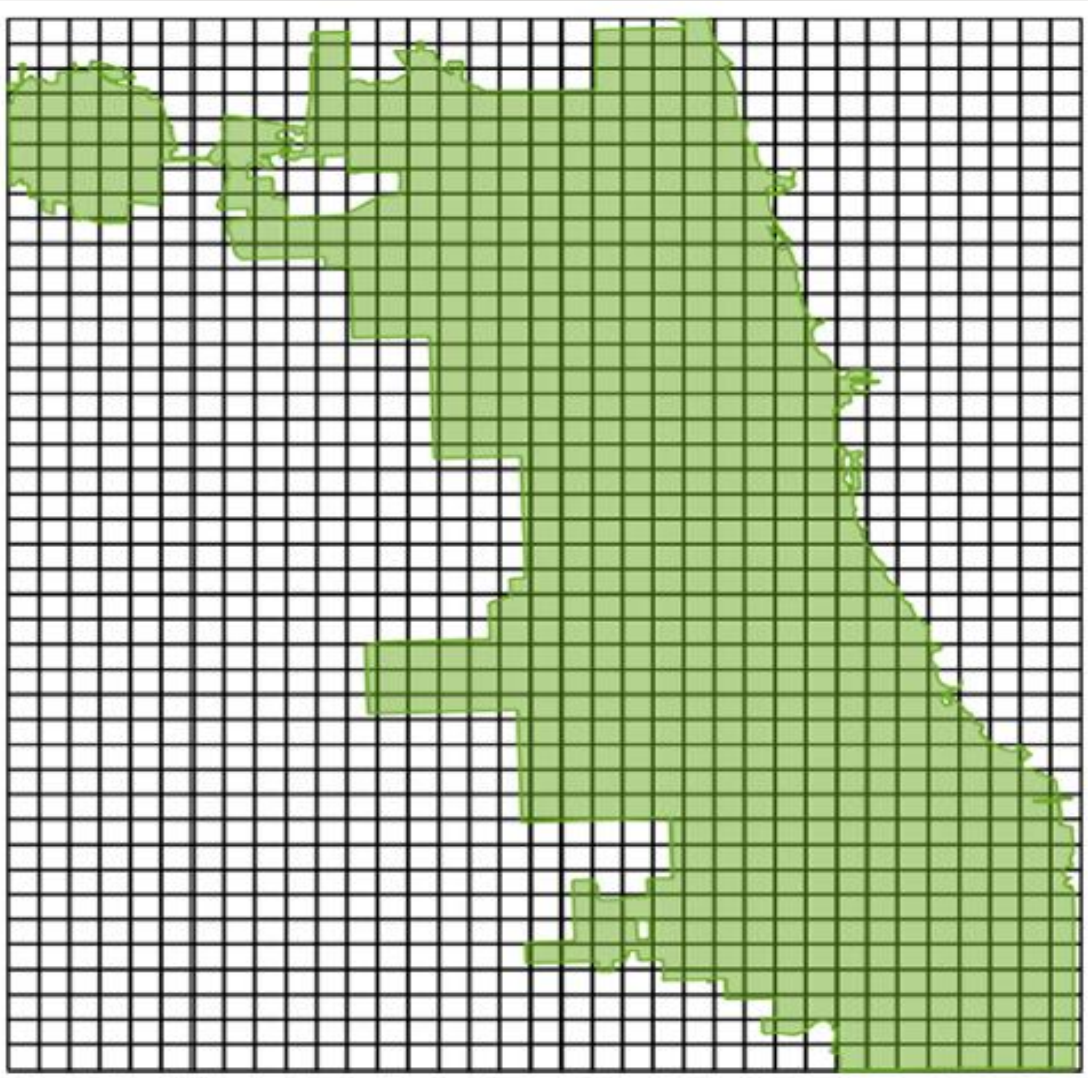
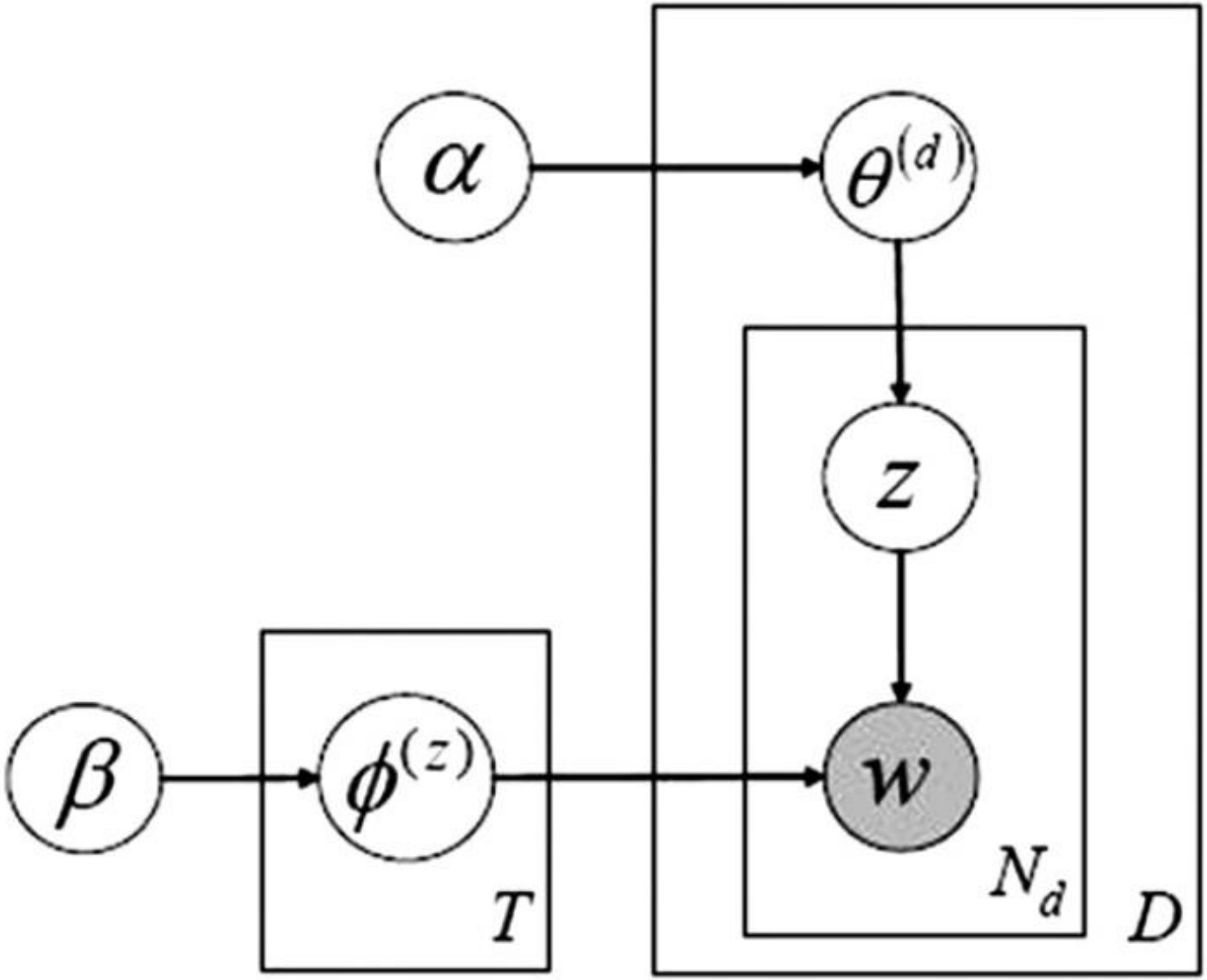


Fig. 3. Neighborhood boundaries for computing tweet-based topics. We only used the green neighborhoods (i.e., those within the city boundary) in our analysis.

Latent Dirichlet Allocation



Analytic approach

$$\Pr\left(Label_p = T \middle| f_1(p), f_2(p), \dots, f_n(p)\right) = F\left(f_1(p), f_2(p), \dots, f_n(p)\right)$$

T= type of crime

$f_1(p)$ = KDE

$f_2(p), \dots, f_n(p)$ = topics from LDA

F= logistic function

Predicting crime using Twitter

common noun, pronoun, proper noun, nominal + possessive, proper noun + possessive, verb, adjective, adverb, interjection, hashtag*, emoticon*, nominal + verbal, proper noun + verbal, existential “there” + verbal.

Full model formulation

$$\Pr(t|r)\ 1 \le t \le T = \#\text{topics}$$
$$1 \le t \le T = \#\text{neighborhoods}$$

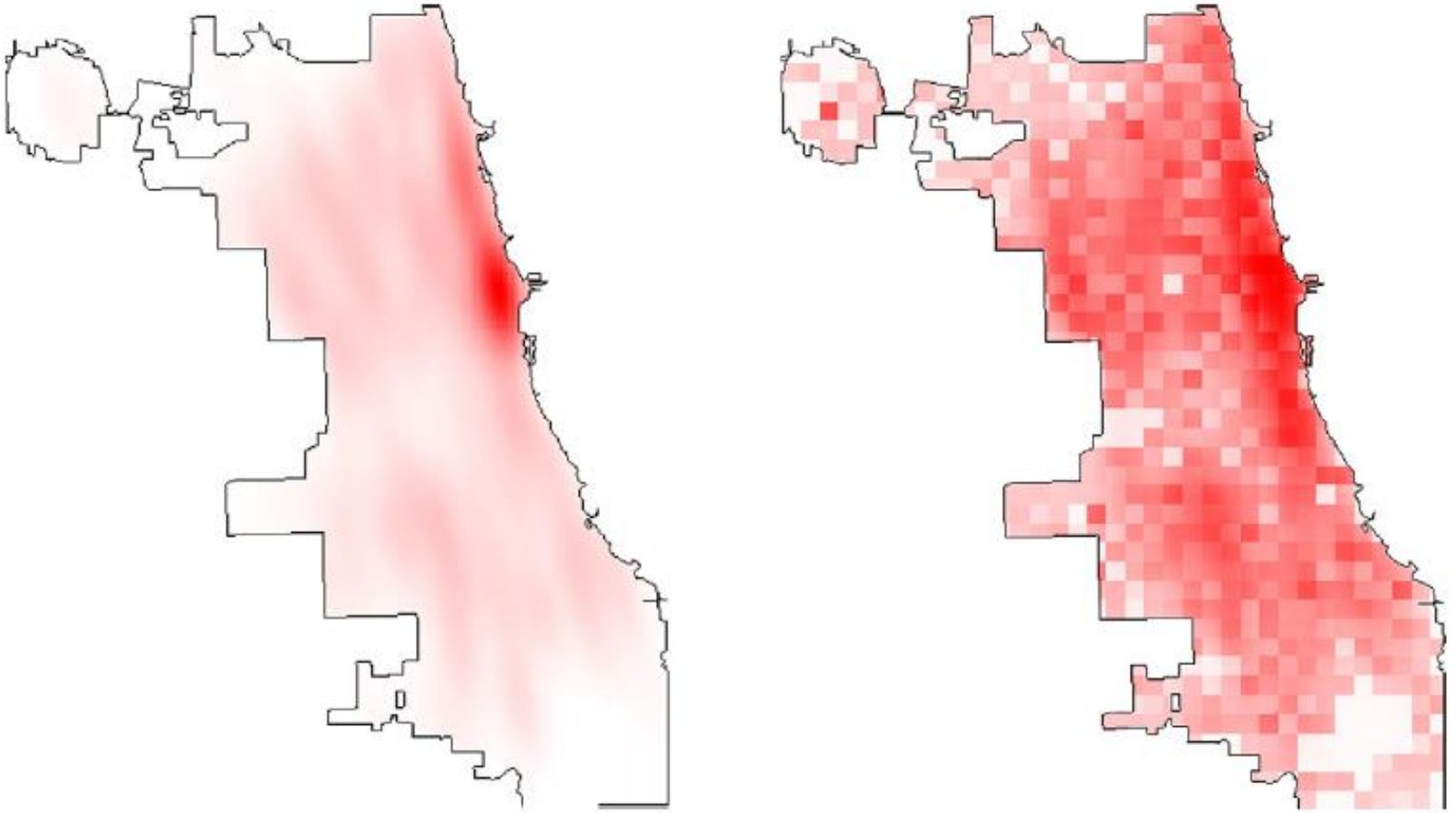
Full model

$$\Pr\left(Label_p = T \middle| f_1(p), f_2(p), \dots, f_n(p)\right) = \frac{1}{1 + e^{-\left(\beta_0 + \prod_{j=1}^{n} \beta_i f_i(p)\right)}}$$

Probability interpolation function

$$Pr_I\left(Label_p = T, W\right) = \sum_{j=1}^{|N(p,W)|} \frac{W - D(p, n_i)}{\sum_{j=1}^{|N(p,W)|} W - D(p, n_i)} * \Pr(Label_{n_i} = T)$$

Evaluation and results



(a) Predicted threat surface using only the KDE feature.

(b) Predicted threat surface using the KDE feature and the Twitter features.

Quadrants of police

Quadrant

CAI PLAZA
Cuadrante 1
Cuadrante 2
Cuadrante 3
Cuadrante 4
Cuadrante 5
Cuadrante 6
Cuadrante 7
Cuadrante 8
Cuadrante 9
C 10
C 11

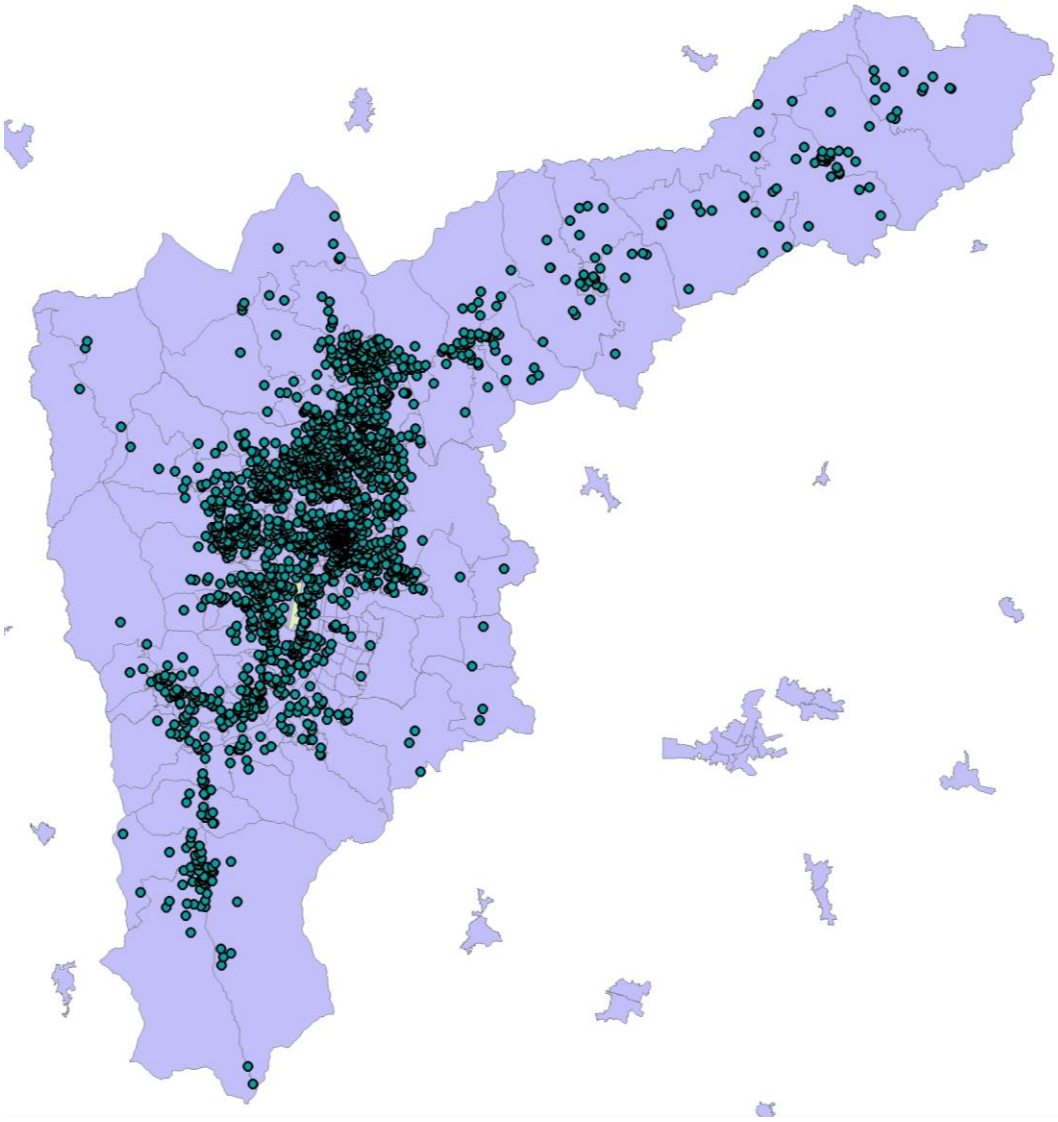
Quadrant definition

A quadrant is a fixed geographical sector that, based on its criminal, contraventional characteristics, social, demographic, geographical and economic services receives different types of police service, with orientation to the solution of problems of citizen coexistence and citizen security in the urban and rural areas.

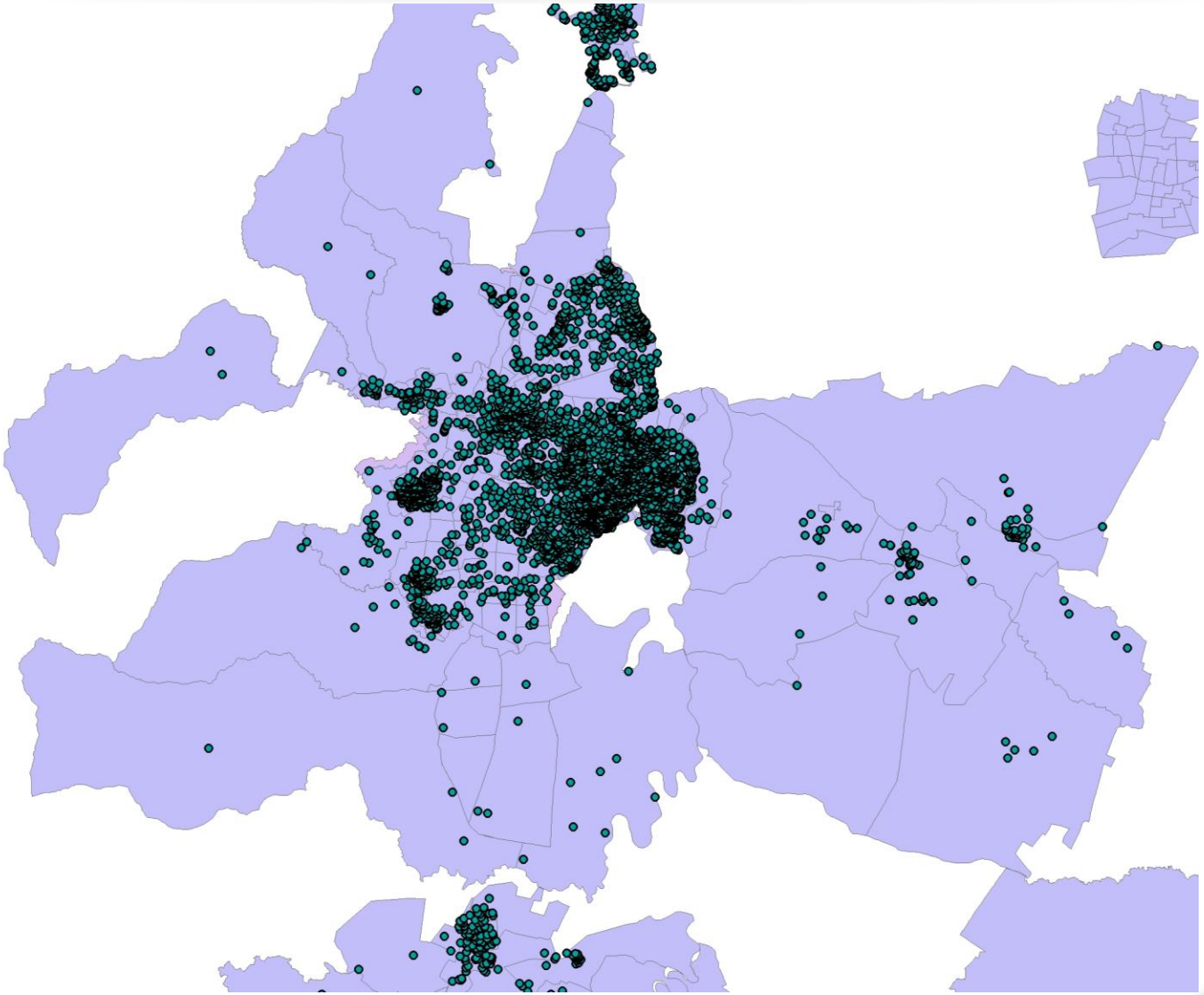
The quadrant is the scenario of tactical and operational deployment in which all specialties converge of the police service.

According to the criminal, contraventional and geographical characteristics, the following may be defined: types of quadrants: urban, rural, road and river.

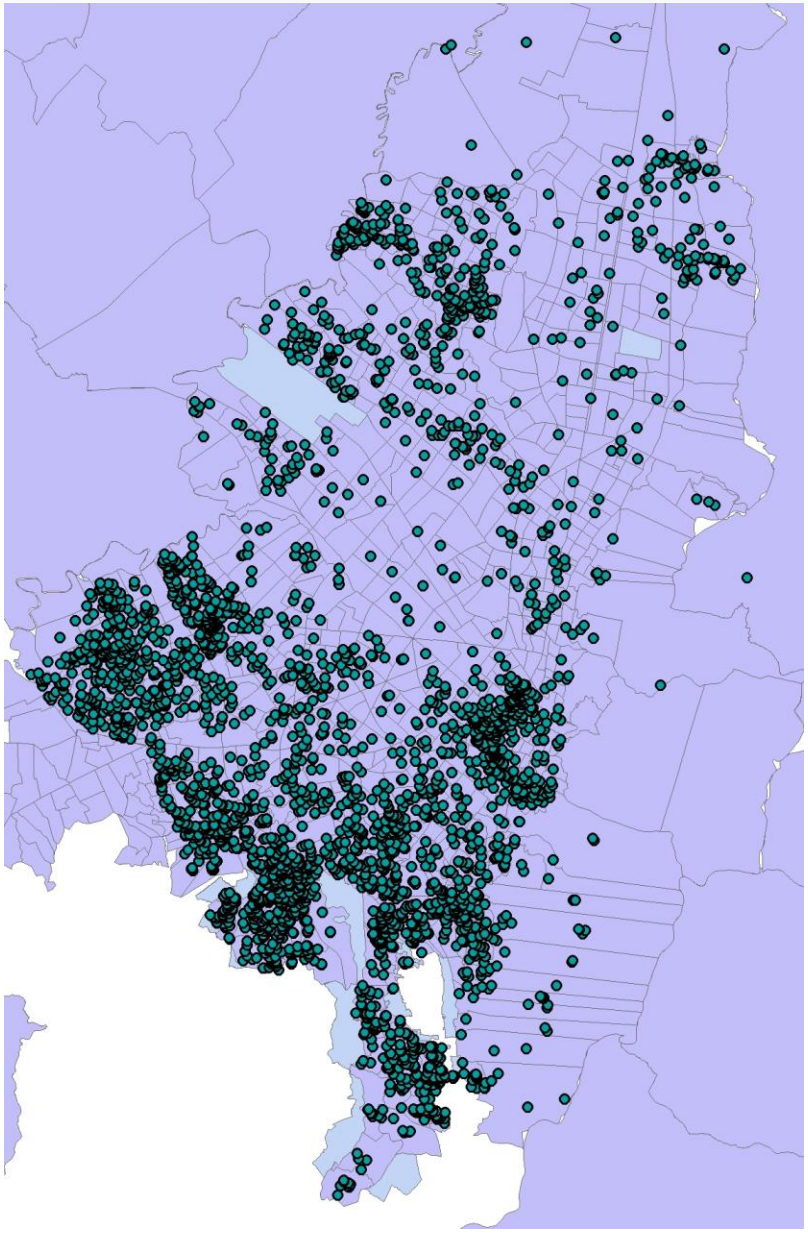
Medellin’s murders



Cali's murders

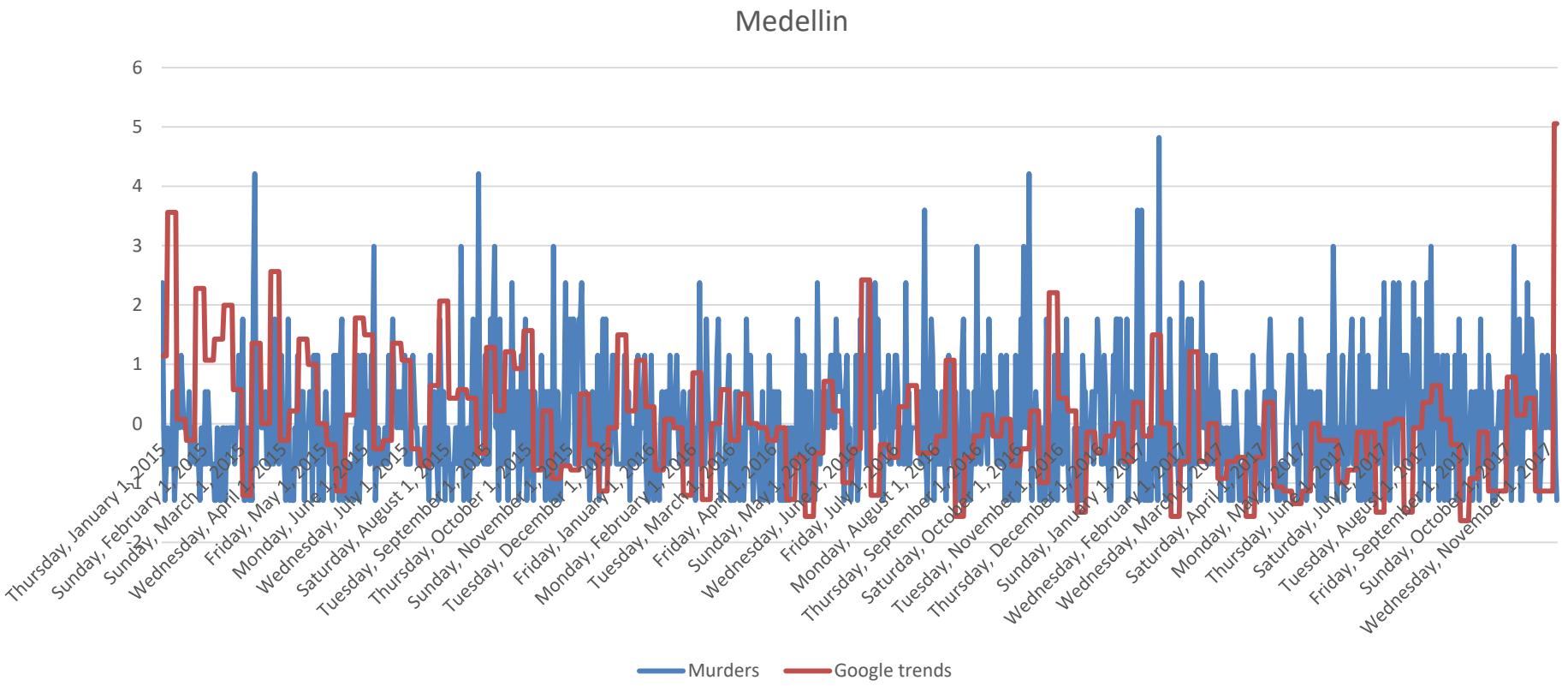


Bogota D.C.’s murders

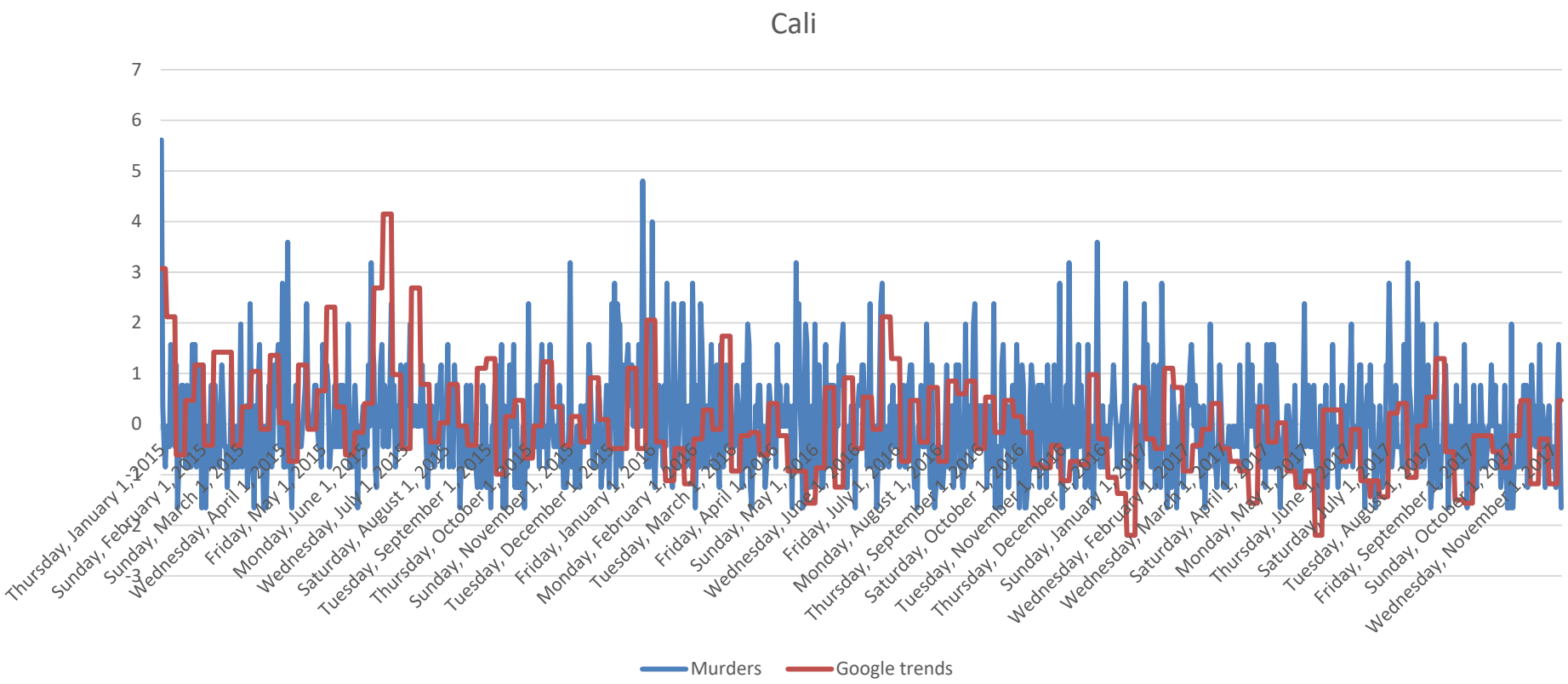


Google Trends and limitations

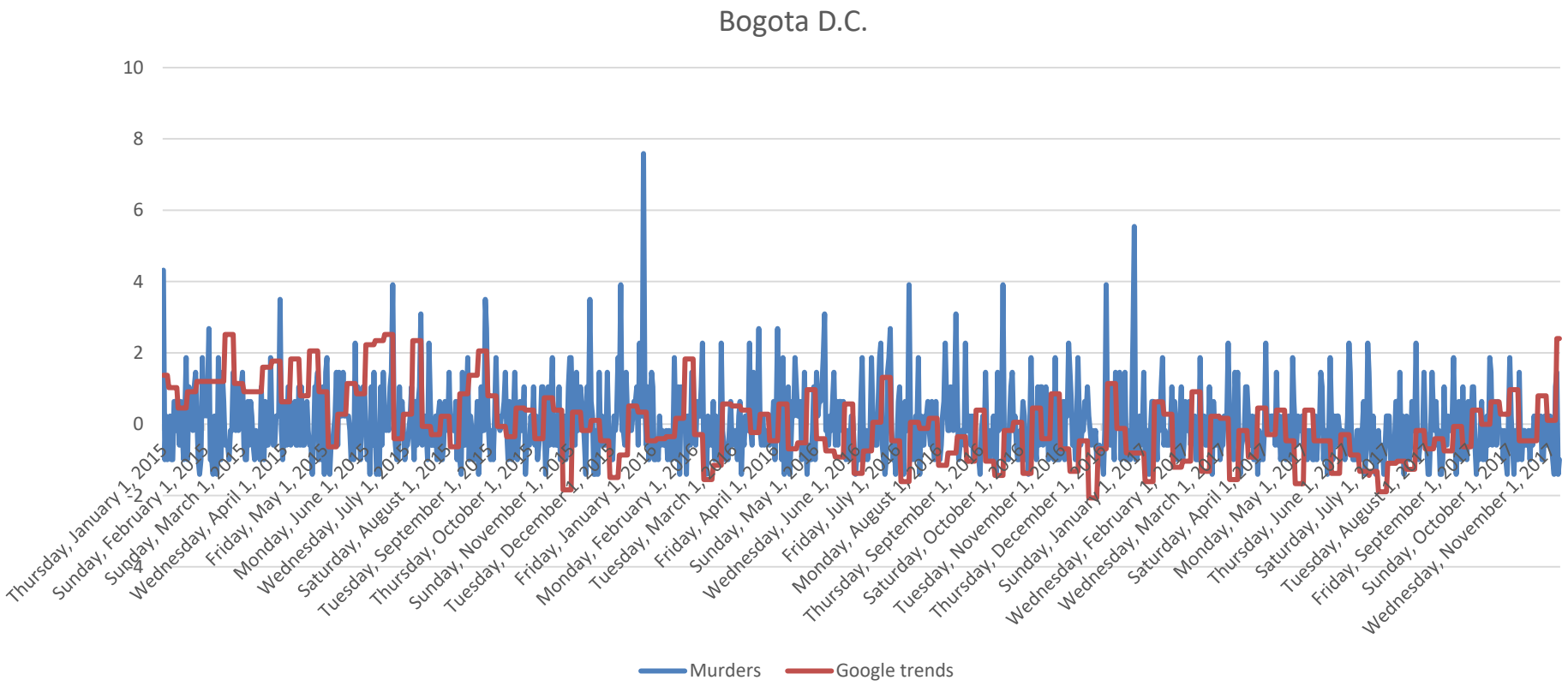
Statistics vs Trends



Statistics vs Trends



Statistics vs Trends



Limitations of information

- Not too much georeferenced information, in databases and Twitter.
- The Colombian population doesn't habituate to use location data.
- Few access in rural areas to internet services.

Conclusions

- Police intelligence reports are the best way to complement social media information.
- Integral Analysis of Citizen Security and Coexistence (AISEC), to get information about criminal organization, and the social and policing context.
- Monitoring System for the Detection of Illicit Crops (SIIMA), to understand georeferenced information.