Inspire Create Transform

UNIVERSIDAD
EAFIT®

PH.D IN MATHEMATICAL ENGINEERING

DOCTORAL SEMINAR III

JUNE 08, 2018

Inspire Create Transform | Vigilada Mineducación

NON-PARAMETRIC SPACE-TIME PREDICTION FOR ORGANIZED CRIME, BASED ON SOCIAL MEDIA, POLICE DATA AND OPEN SOURCES

Andrés Pérez-Coronado

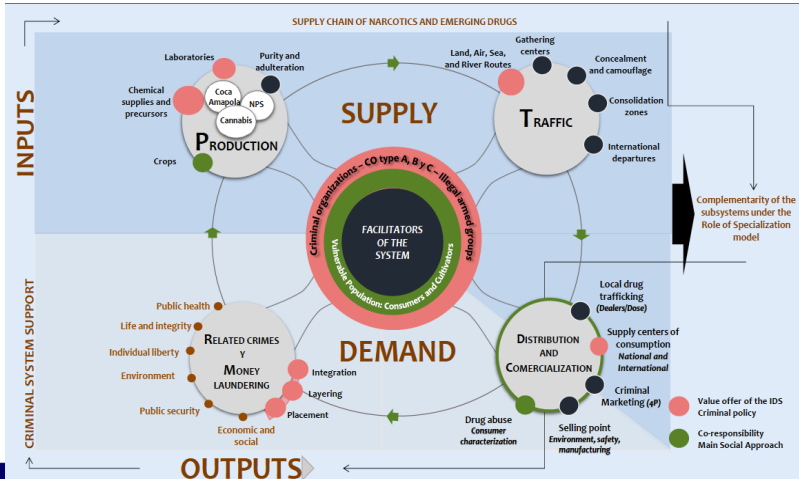**Thesis Advisor:**
Henry Laniado and Gustavo Canavire

EAFIT University
School of Sciences - Department of Mathematics Sciences
Ph.D in Mathematical Engineering
June 14, 2018

Inspire Create Transform | Vigilada Mineducación
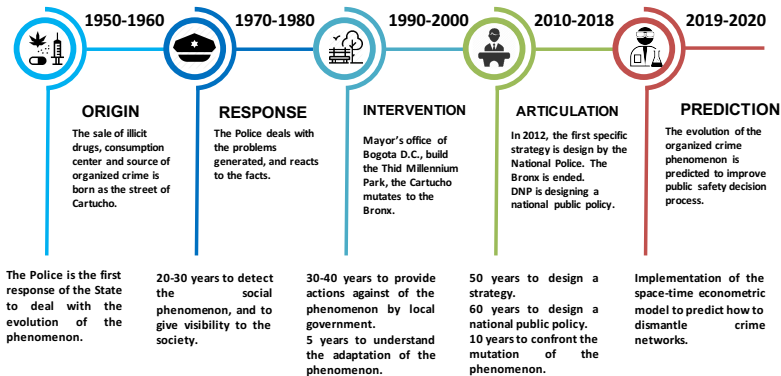
UNIVERSIDAD
EAFIT

System of Illicit Drugs

Source: National Police of Colombia

Longitudinal analysis: drug trafficking

Click here

Research problem:
from Calle del Cartucho to Bronx

1950-1960   **1970-1980**   **1990-2000**   **2010-2018**   **2019-2020**

ORIGIN

The sale of illicit drugs, consumption center and source of organized crime is born as the street of Cartucho.

RESPONSE

The Police deals with the problems generated, and reacts to the facts.

INTERVENTION

Mayor's office of Bogota D.C., build the Thid Millennium Park, the Cartucho mutates to the Bronx.

ARTICULATION

In 2012, the first specific strategy is design by the National Police. The Bronx is ended. DNP is designing a national public policy.

PREDICTION

The evolution of the organized crime phenomenon is predicted to improve public safety decision process.

The Police is the first response of the State to deal with the evolution of the phenomenon.

20-30 years to detect the social phenomenon, and to give visibility to the society.

30-40 years to provide actions against of the phenomenon by local government.
5 years to understand the adaptation of the phenomenon.

50 years to design a strategy.
60 years to design a national public policy.
10 years to confront the mutation of the phenomenon.

Implementation of the space-time econometric model to predict how to dismantle crime networks.

Outline

Research introduction

Work in progress
    Twitter and Kernel Density Estimation (KDE)
    Thin Plate Spline (TPS)
    Shrinkage
    Google trends

Future research line
    Social network analysis framework
    Process of making predictive decision

Inspire Create Transform | **Vigilada Mineducación**

UNIVERSIDAD
EAFIT

Research introduction

Criminal approach

- **Crime:** an action or omission which constitutes an offence and it is punishable by law.
- **Criminal:** an individual who has committed a crime.
- **Organized crime:** a structured network (criminals) whose primary objective is to obtain money through illegal activities (crimes).

Research question

Can a statistical model predict the organized crime felonies, and breaking up its criminal networks?

Research objectives

Main objective:

To design a space-time statistical model that predicts crime events and how it could be useful to break up its criminal networks.

First specific objective:

To propose a non-parametric space-time model to predict the occurrence of the crimes, based on social media, police data and open sources.

Work in progress

Work in progress: Twitter and Kernel Density Estimation (KDE)

Analytic approach

$$Pr(Label_p = T | f_1(p), f_2(p), ..., f_n(p)) = F(f_1(p), f_2(p), ..., f_n(p))$$

$T =$ type of crime
$f_1(p) =$ density function
$f_2(p), ..., f_n(p) =$ topic analysis

Kernel Density Estimation (KDE)

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

P = total number of crime T
p = spatial point
h = bandwidth
K = standard normal density function
$||.||$ = euclidean norm
p_j = location of crime j

Longitudinal analysis: KDE
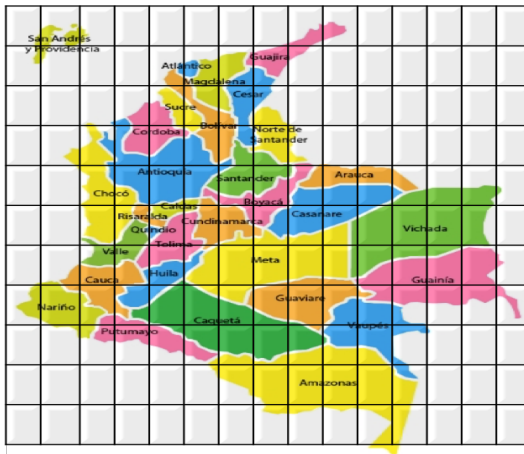
Longitudinal analysis: corruption

Click here

Topic analysis

Source: https://pointwest.com.ph -April 24 Philipinas Debate 2016

TOPIC ANALYSIS
TOPICS THAT EMERGED FROM THE EXTRACTED TWEETS GRAPHED AGAINST TIME

Before the debate started, tweets of Binay supporters flocked Twitter

Tweets supporting Santiago during her opening statement

Netizens started describing their dream President when the candidates answered various issues

Complaints to ABS-CBN for being biased in favor of Mar Roxas during the fast talk segment

Support for Miriam Defensor-Santiago poured out when she delivered her closing statement

Topic 1: Tweets supporting Jejomar Binay

Topic 2: Tweets supporting Miriam Santiago

Topic 3: Netizens' standpoint on the candidates' answers regarding the issues tackled

Topic 4: Alleged bias of ABS-CBN and Karen Davila in favor of Mar Roxas

Topic 5: Netizens' description of their ideal president borrowing each candidate's best personality

Topic 6: Opinions on the candidates' promises upon appointment as President

Inspire Create Transform | Vigilada Mineducación

UNIVERSIDAD EAFIT®

Full model formulation

(Binary logistic regression model)

$$Pr(Label_p = T | f_1(p), f_2(p), ..., f_n(p) = F(f_1(p), f_2(p), ..., f_n(p)) =$$
$$\frac{1}{1 + e^{-(\beta_0 + \prod_{i=1}^{n} \beta_i f_i(p))}}$$

$i = 1$, f_i equals the KDE
$i > 1$, f_i equals $Pr(i - 1 | r)$
$r =$ is the unique topic neighborhood that spatially contains p
$\beta_i =$ coefficients

Inspire Create Transform | **Vigilada Mineducación**

Probability interpolation function

(Distance-weighted spatial interpolation (IDW))

$$Pr_1(Label_p = T, W) =$$

$$\sum_{i=1}^{|N(p,W)|} \frac{W - D(p, n_i)}{\sum_{j=1}^{|N(p,W)|} W - D(p, n_j)} * Pr(Label_{ni} = T)$$

$Pr_1 =$ probability interpolation function

$W =$ is a windowing parameter of, for example 500 meters

$N(p, W) =$ is the set of p's neighbors within a distance of W
(this set includes p itself)

$D(p, n_i) =$ is the straight-line distance between p and one of
its neighbors n_i

$(Label_{ni}) = T$ is the non-interpolated probability

Distance-weighted spatial interpolation (IDW)
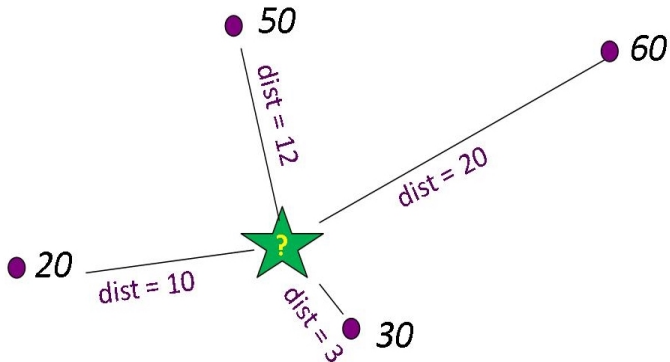
Work in progress: Thin Plate Spline (TPS)

TPS concept
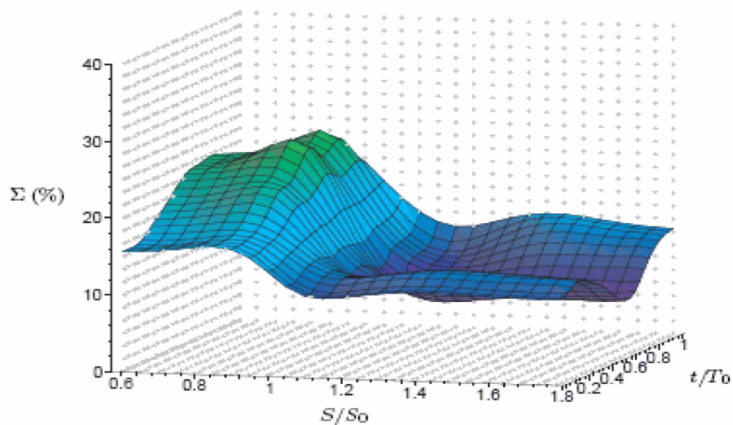
TPS Vs IDW

Thin Plate Spline (TPS)
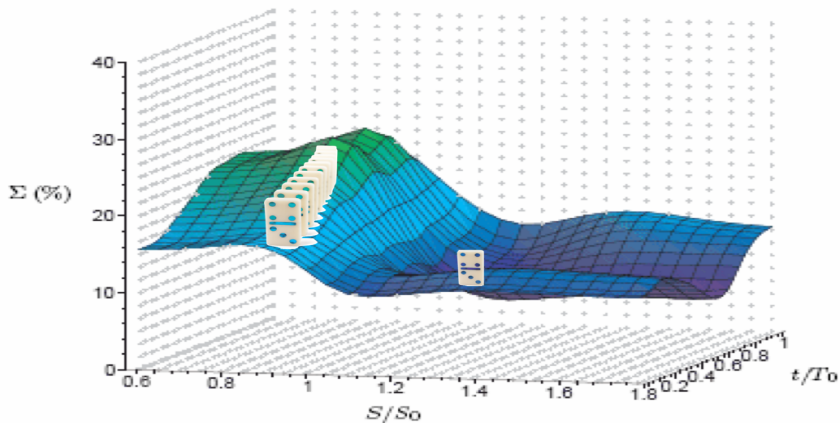
Work in progress: Shrinkage

Shrinkage

Source: M.J.D. Powell. Some algorithms for thin plate spline interpolation to functions of two variables

$$\begin{pmatrix} \Phi & P \\ P^T & 0 \end{pmatrix} \begin{pmatrix} \lambda \\ c \end{pmatrix} = \begin{pmatrix} f \\ 0 \end{pmatrix}$$

Work in progress: Google trends

Limitations

- Colombians don't use usually the location in Google search.
- Internet have few coverage in rural areas.
- The Google trends are grouped by capital cities, and we cannot get information about others towns.

Future research line

Research objectives

Second specific objective:

To develop a predictive framework based on social network analysis that be useful on breaking up criminal organizations.

Third specific objective:

To propose a making decision process based on artificial intelligence to policing against organized crime.

Future research line: Social network analysis framework

Social network analysis to break up organized crime

SUBSYSTEM OF PRODUCTION

SUBSYSTEM OF DRUG TRAFFICKING

SUBSYSTEM OF MONEY LAUNDRY AND RELATED CRIMES

SUBSYSTEM OF ILLICIT DRUG MARKETS

To understand the criminal links, to find the most important vertex, and to support policing on breaking up current networks and to predict the future networks.

To use social media to understand the space-time crime context, and to perform the process of making predictive decisions.

Inspire Create Transform | Vigilada Mineducación

UNIVERSIDAD EAFIT®

Social network analysis to break up organized crime
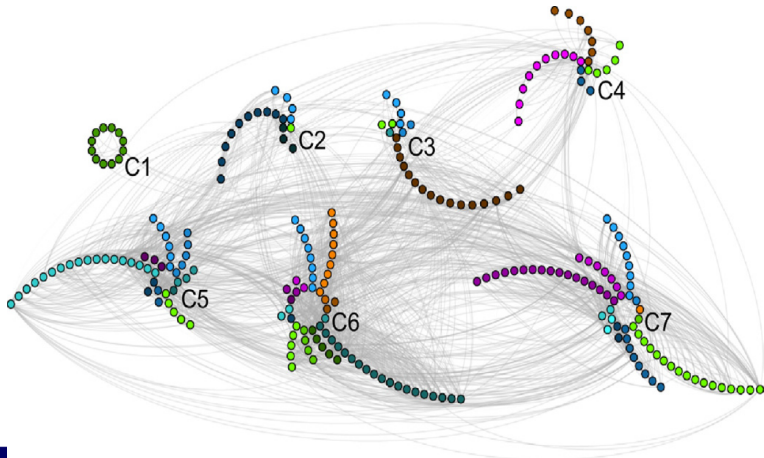
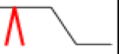# Future research line: Process of making predictive decision
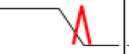
Artificial intelligence for making decision process

Source: Lesca, H. et Lesca, N. (2011)

Artificial intelligence for making decision process

Thanks