

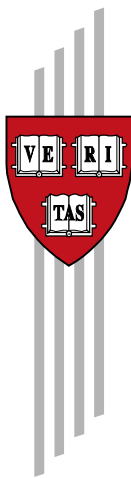
**How and where do criminals operate?**

**Using Google to track Mexican drug  
trafficking organizations**

Michele Coscia/Viridiana Rios

CID Research Fellow and Graduate Student Working  
Paper No. 57, August 2012

© Copyright 2012 Coscia, Michele, Rios, Viridianna, and the  
President and Fellows of Harvard College



**Working Papers**

Center for International Development  
at Harvard University

# How and where do criminals operate? Using Google to track Mexican drug trafficking organizations

Michele Coscia\*

CID, Harvard Kennedy School

Viridiana Rios†

Department of Government, Harvard University

October 23, 2012

## Abstract

We develop a tool that uses Web content to obtain quantitative information about the mobility and *modus operandi* of criminal groups, information that would otherwise require the operation of large scale, expensive intelligence exercises to be obtained. Exploiting indexed reliable sources such as online newspapers and blogs, we use unambiguous query terms and Google’s search engine to identify the areas of operation of criminal organizations, and to extract information about the particularities of their mobility patters. We apply our tool to Mexican criminal organizations to identify their market strategies, their preferred areas of operation, and the way in which these have evolved over the last two decades. By extracting this knowledge, we provide crucial information for academics and policy makers increasingly interested in organized crime. Our findings provide evidence that criminal organizations are more strategic and operate in more differentiated ways than current academic literature had suggested.

## 1 Introduction

Organized crime has become an increasing concern for international security studies<sup>1</sup>. Yet, as much attention as organized crime has recently attracted, very

---

\*michele\_coscia@hks.harvard.edu

†vrios@fas.harvard.edu

<sup>1</sup>Examples abound: Violent criminal groups in Mexico have been attributed 51 thousand homicides in the last five years<sup>2</sup> leading some to conclude that Mexico has become a failed state; the Japanese Mafia controls an increasingly profitable market of methamphetamines in Asia [23, 18] raising concerns over the ability of the state to enforce the law; rackets of human trafficking in Haiti operate mostly untouched by authorities [27], as do rackets of endangered-species smugglers in other developing countries like Singapore, Manila, Indonesia

few is known about its territorial dynamics<sup>3</sup> If until now academics and policy-makers have been unable to fully explain and understand criminal organizations in detail is because of an inherent difficulty posed by the study of illegal actors: lack of data.

Most research has tried to overcome lack of data by relying in interviews [56, 13], or expensive long-term data classification endeavors from wiretaps or newspapers [35, 32]. Yet, the information extracted in this way is site-specific, and vary in quality and extension depending on who collects it. In the practical ground, the high cost of intelligence exercises has left few options for enforcement agencies in developing countries with low budgets to assess the activities of organized crime.

Developing an easy-to-use, inexpensive tool to extract information of where and how criminal organizations operate feels urgent. Criminology literature points to tracking the location and territory of criminal organizations as one of the most basic and useful tools for improving governments' enforcement capabilities and for promoting the rule of law<sup>4</sup>. If data about criminal organizations could be extracted easily and at lower cost our capacity to understand, and thus control and contain criminal operations will be boosted.

This paper does it. We develop a tool that helps academics and low-budget enforcement agencies to assess where and how criminal organizations operate, without requiring costly intelligence exercises. Our algorithm filters the vast amount of knowledge available at the Web and, utilizing some already indexed reliable sources from Google, such as online newspapers and blogs and unambiguous query terms, identifies critical features of criminal organizations. Particularly, we identify the municipalities at which thirteen Mexican drug trafficking organizations have operated over the last two decades, and how these differ in their market strategies. Our results provide the most up-to-date and accurate information of Mexican criminal operations

We use the web because it is an extremely valuable depository of up-to-date, public, and private information. If a decade ago the main constraint faced by investigators was information access, now the constraint is how to manage, filter,

---

and Jakarta [49], and Bosnia's sex trafficking industry [4], and contemporary African pirates [22] are expanding criminal enterprises that pose important security threatens.

<sup>3</sup>Research has pointed out to its increasing degree of fluidity and structural complexity [2, 46, 55] (cited by [7]), but as [53] mentioned in his influential review of our current understanding of crime, "studies tend to lack detail regarding the modus operandi and logistics of transnational criminal activities."

<sup>4</sup>Large reductions in homicide rates in cities like Rio de Janeiro, Bogota, a and various cities in the United States have all been attributed to improved mechanisms for understanding criminal mobility, territoriality and the timing of crime. For example, a reduction of 75% in drug-related homicides happening at Tampa, Florida from 1989 to 1991 has been attributed to the ability of local police departments to predict where criminals will go after one of their bases of operation has been cracked down on

and gather the massive amounts of information that we have access to. Under current circumstances, the huge potential of internet as a tool for information collection remains weakly exploited. We seek to harness this power with our information processing tool. At the web, data that used to be secret or difficult to access is publicly shared, discussed and posted on specialized blogs and forums, and the activities of criminals groups are not the exception. Members of Mexican criminal organizations regularly share information of their operations, activities and rivalries on websites. They communicate with their allies, threat their enemies and brag about their achievement in forums and discussion boards<sup>5</sup>. Not only criminals share information at the web, also journalists and civilians do. Sensitive information which would endanger the lives of those exposing it can be safely shared in digital, local newspapers and information blogs. It has been documented that people share more information on websites than in personal interviews or any other form of human interaction<sup>6</sup>.

Even if the web is the most famous and the largest repository of explicit knowledge, it is not until now that academics have started using it to extract explicit knowledge because of two large difficulties: its size and reliability. Reading all that has ever been posted at the web is almost impossible, and many are the concerns over whether the information contained in the web is to be trusted.

In this paper, we prove that these two difficulties can be overcome in inexpensive and relatively easy ways. We show that we can create accurate yet complex maps of the areas of operation of criminal organizations, a feature that would otherwise remain concealed to the academic world. By doing so, we open the possibility of quantitatively tackling puzzles that have remain largely unsolved by political scientists, such as why actors antagonizing the state decide to operate in some areas, and not in others. Little effort has been made to understand location decisions of criminal and hostile groups [7, 33, 51]. Some research has shown the importance of topography, climate and population size due to strategic concerns [24, 11, 34], access to local knowledge and information [15, 10], or operational costs [29]. Other, particularly within the criminal literature [53], have pointed that actors that antagonize the state tend to operate along legal trade routes to access resources [3, 26, 52], embedded in diaspora communities [50, 6, 8, 31, 54, 39], or in areas where the state has weak presence [14, 45] (cited by [7, 54]). These contending explanations will now be able to be

---

<sup>5</sup>Actually, large criminal organizations have been discovered and broken down precisely by tracking their online communications' forums. The key of this type of law enforcement operation lies in figuring out which forums and blogs to search within the web. Our tool will also be helpful for the execution of these operations by allowing law enforcement agencies to perform larger searches that would otherwise require an unfeasible number of individuals reading online posts.

<sup>6</sup>Academics and entrepreneurs have just begun to take advantage of this behavior with quite promising results. The success of new business models like online-dating services lies precisely on their ability to extract extremely personal information from a large set of internet users - information that people would normally refuse to share personally - and use it to feed complex compatibility-matching algorithms.

tested, relying on our tool.

This paper is organized as follows. In Section 2 we review previous works similar to our approach, we refer to studies about organized crime in Mexico, and we present our algorithm. Section 3 reports the raw data that we obtained using our algorithm. We then explain how we used this data to create actual information about criminal organizations in Section 4. In the following section we present the results of our exercise, explaining where and how different drug trafficking organizations operate. Section 7 concludes the paper.

## 2 Identification strategy

There are several works that try to use Web information to reconstruct complex phenomena. In [28], social relations among politicians, baseball players and physicists are tracked by co-googling them in the well know online search engine, thus building a map of their pairwise correlations. Also some references about the approximations which are hidden behind the Google search form is given. Co-occurrences in the abstracts of papers are also used in the context or music [43], in bio-informatics to disambiguate names of genes and proteins [9], to discover word meanings [17], to rank entities [47], to evaluate the sentiment of people writing opinions [37, 30]. An interesting example about networks of co-occurrences of classifications in classical archaeology publications is [44].

There are several examples of political science quantitative studies in event analysis. An example of such a system is provided in [25]. Other political studies range from the analyses of presidential, legislator, and party statements [19], to treaty-making strategies [48], to disaster relief organization through social media responses [1]. In general, a good review work of political science applications of techniques similar to the one presented in this paper can be found in [20], which also provide information about the general organization of works in those category, that also apply to this paper. None of the reported methods take advantages of the freely available information present in the web, from reliable sources like the newspapers indexed by Google News.

In the case of Mexico’s criminal organizations, to the extent of our knowledge, there is no other data set privately or publicly compiled that contains the level detail and length as the one we collected. Private efforts like Stratfor<sup>7</sup> and Guerrero [21] had provided information on the territories of operation of drug trafficking organizations but only at the state level and without time variation. Mexican secret intelligence office (CISEN) has information at the municipal level but is not available for research purposes and does not provide information for years before 2006<sup>8</sup>. Our tool provides information, detailed to the subnational

---

<sup>7</sup>Link needed.

<sup>8</sup>We estimate that about 63% of CISEN’s data had at some point been covered by Google

level, of where and when one of thirteen different drug trafficking organizations were present in each of the 2,457 municipalities of Mexico, yearly from 1991 to 2010.

We extracted our data using an algorithm that we developed and named MOGO<sup>9</sup>. MOGO works in three steps. First, it defines the actors that will be identified, in this case an extensive list of criminal organizations and criminals and a list of locations. Then, we create set of queries to search for matches between criminal organizations and locations in the text of notes at Google News. We decided to use Google News as it organizes sources that are supposedly reliable (official newspapers and blogs). Queries use key words to make sure that searches are non-ambiguous. Finally, we store the hits that we obtained and identify the date in which the note was created<sup>10</sup>.

### 3 Extracted Raw Data

In this section we provide some statistics about the raw data retrieved using MOGO.

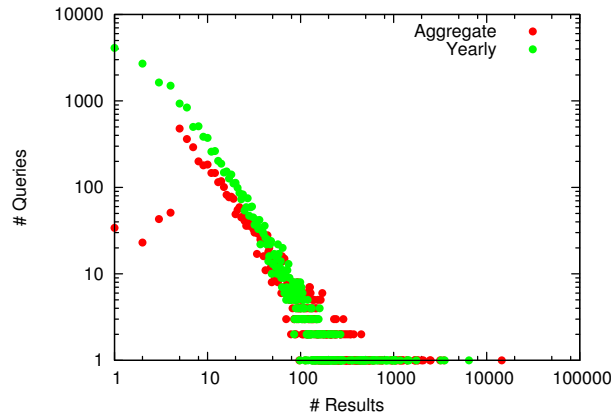
First, in 1 we present the distribution of hits per query. We depicted the results per couple both aggregate over the entire time window and disaggregate per year. We see a fat tail distribution in both cases because each year, there are many municipalities with a very low drug-related criminal activity and some “hubs” where criminal activity is concentrated that record between 1,000 and 10,000 hits. In the aggregate distribution, we observe also that the number of couples with few hits are less than expected. This indicates that the pattern of activity is not constant. Municipalities that do not show any activity in one year often start to experience activities in our 20-year interval.

---

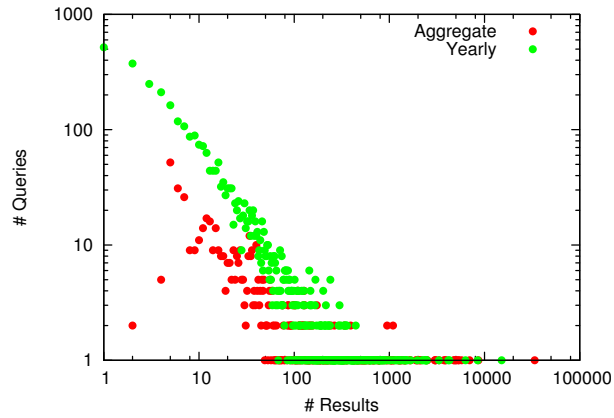
news. This estimate comes from comparing a dataset of personal communications between traffickers that we collected from the web to the same dataset collected by CISEN. Out of a total of 1421 communications collected by CISEN, 888 were reported at Google News. We took this as a reference of the amount of CISEN’s information that is available at the web

<sup>9</sup>We named our tool MOGO after a famous Australian aborigine tracker who in 1834 found a missing ve-year-old boy, lost for over ten hours in the wilderness, using indigenous tracking methods. British settlers - who since then deemed these skills as remarkable and miraculous. They soon started enlisting the so called “aboriginal trackers” to assist law enforcement authorities in the capture of bush-rangers. Their unquestionable success led to the development of regular tracking operations that remain at the very top of Australian security methods. Actually, ever since, indigenous patrolling skills have a proud record in assisting colonial and modern-day law enforcement and military operations. The spirit of aboriginal trackers has now been embedded by the Regional Force Surveillance Units, a specialized Australian infantry unit which patrols the less-populated Northern territories, providing valuable intelligence to customs, state and Federal police forces, and the intelligence community. Mogo also stands for **M**aking **O**rders using **G**oogle as an **O**racle

<sup>10</sup>A detailed description of the algorithm is described in Coscia and Rios (2012)



(a)



(b)

Trafficker	Total Hits
Sinaloa	57,988
Zetas	37,992
Golfo	28,837
Beltrán Leyva	27,824
Tijuana	17,987
Familia	15,448
Juarez	11,817
Teo	4,505
Azul	2,656
Barbie	2,418
Others	8,961

(c)

Figure 1: (a) The hits distributions aggregated as (trafficker, municipality) and dis-aggregated as (trafficker, municipality, year); (b) Hits distribution per municipality and dis-aggregated as (municipality, year); (c) Hit distribution per trafficker.

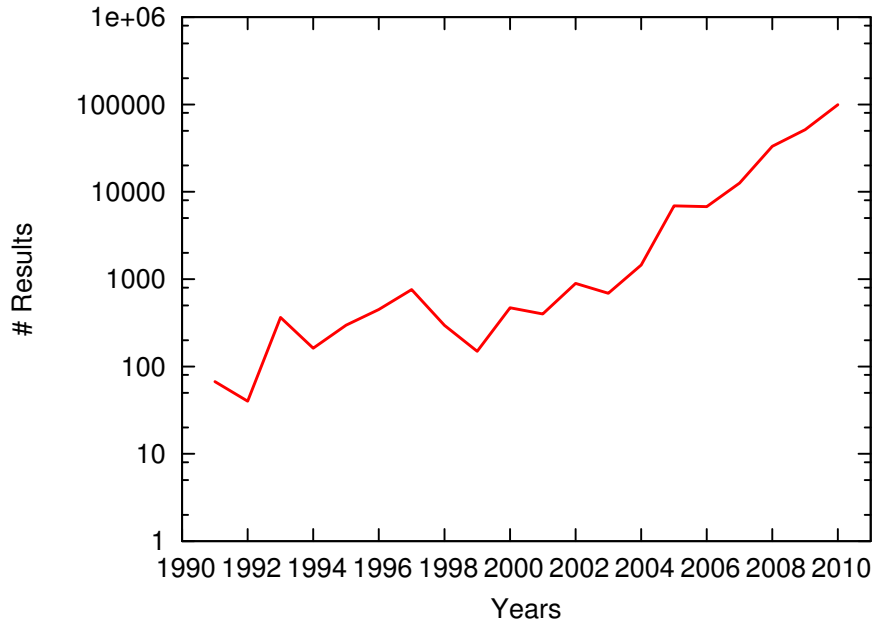


Figure 2: The results distributions per year.

To confirm this hypothesis, we plot in Figure 1b the same hits distributions aggregating over drug trafficking organization. Each dot is now either a municipality (red dots) or a municipality in a year (green dots). The same pattern emerges.

In Table 1c we report trafficking organizations sorted by popularity (we report only the top 10 over the 13 trafficking organizations, aggregating the remaining as “Others”). Given the few data points we do not observe a fat tail distribution, but the number of hits is heavily unbalanced. The top 4 trafficking organizations (out of 13) account for more than 70% of the total amount of hits.

These pictures suggest that we are observing a complex system<sup>11</sup> with an exponential growth, with typical complex system characteristics such as the presence of uneven (power-law) distributions in the number of hits [36].

These findings suggest that to understand Mexico’s drug trafficking industry is not useful to study each single traffic organization, territory and law enforcement organization, i.e. the parts of the system, but rather how these parts interact as a whole, one with each other.

---

<sup>11</sup>A complex system is a system composed of different parts that expresses at the global level properties that are not present in any single part taken alone.



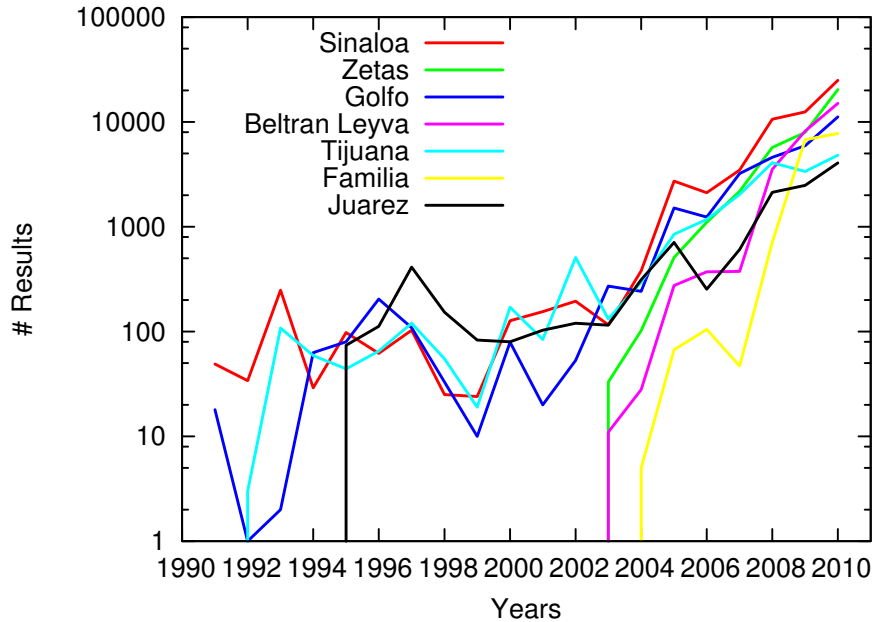


Figure 3: The results distributions per trafficker per year, for the most popular seven traffickers.

In Figure 2 for each year we report the number of hits for all queries. This picture is backing the the fact that the phenotype of Mexico drug traffic is growing beyond control (please notice the logarithmic y axis)<sup>12</sup> However, if there would be no superlinear growth, one would expect the line to be stable after 2006. Instead, it is exactly from 2006 onward that we witness the most incredible growth, jumping one order of magnitude (from 10,000 to 100,000 articles) in just four years.

In Figure 3 we disaggregate the previous figure by drug trafficking organization (for clarity purposes, this time we select the top 7 organizations reported in the majority of documents). Some organizations appear slightly before their foundation. Zetas, for example, have a low initial popularity (below ten hits) but grow impressively, catching up with older trafficking organizations in just a few years. In general, also old trafficking organizations such as Sinaloa and Golfo, grow superlinearly, It is this growth rate what makes themselves complex systems beyond any description that can be made for each of their leaders and main components.

<sup>12</sup>We expect some downward bias for years before 2006, while Google News was still in beta, before the collection of articles in years previous to 2006 may not have been complete.

## 4 Creating Knowledge from Raw Data

The raw data showed in the previous section, as any information retrieved from online search engines, yields information that is incomplete, noisy and potentially affected by the size and extension of online news coverage or web popularity. For example, retrieving hits connecting two different actors does not necessarily mean that the two actors are involved in an actual relationship in the real world.

Thus, we developed a way to filter, clean and validate our raw data in order to transform them into meaningful results that actually describe the reality of Mexico’s criminal organizations. We normalized the total number of hits we are getting using a hyper-geometric cumulative distribution function. Our goal is to overlap a trafficking organization  $t_i$  and a municipality  $m_i$  in order to identify municipalities where drug trafficking organizations operate for real, cutting out noise. We can draw a parallel between the null model we are trying to identify and the extraction of a labelled ball from a bin. The municipalities would be the bins, and the trafficking organizations would be the labelled balls. Thus, the number of balls we have for each  $t_i$  is equal to the total number of hits that all the queries related to  $t_i$  returned, a figure that we represent as  $\mathcal{T}_i$ . The question we are asking is whether the number of times a ball is seen in a bin is larger or smaller than what it would be expected by chance, given both the number of balls and bins<sup>13</sup>. Since the total number of balls we have is limited and equal to the number of times a  $t_i$  appears overall, the null model we are looking for is a hyper-geometric probability distribution.

The hyper-geometric distribution is a probability distribution describing the probability of extracting  $\bar{t}_i$  times in  $\bar{m}_i$  attempts a ball labelled with  $t_i$ , given that the total number of such balls is  $\mathcal{T}_i$ , from a total set of balls equal to  $\mathcal{M}$  [41]. The corresponding probability mass function  $PMF$  for  $t_i$  is defined as follows:

$$PMF(t_i = \bar{t}_i) = \frac{\binom{\mathcal{T}_i}{\bar{t}_i} \binom{\mathcal{M} - \mathcal{T}_i}{\bar{m}_i - \bar{t}_i}}{\binom{\mathcal{M}}{\bar{m}_i}}$$

As an example, assume  $\mathcal{T}_i = 5$ , i.e. we obtained 5 hits for  $t_i$ , that is recording a total of 50 total hits ( $\mathcal{M} = 50$ ). Considering municipality  $m_i$ , with a total number of 10 hits ( $\bar{m}_i = 10$ ), we find that among those 10 hits, 4 were related to  $t_i$  ( $\bar{t}_i = 4$ ). The corresponding probability of this happening is equal to  $\frac{\binom{5}{4} \binom{45}{6}}{\binom{50}{10}}$  or approximately 0.00396.

Our final results, what we actually use as our final real-world information, is not the exact probability of obtaining  $\bar{t}_i$  results for  $t_i$  in  $m_i$  but rather the

---

<sup>13</sup>We assume each  $t_i$  is independent

probability of obtaining  $\bar{t}_i$  or less results, that is the cumulative distribution function (*CDF*) is defined as:

$$CDF(t_i = \bar{t}_i) = \sum_{a=0}^{\bar{t}_i} PMF(t_i = a),$$

which takes values from 0 to 1.

The reason is that the probability mass function is not monotonic, while the cumulative distribution function is. If  $t_i$  is very popular, it is difficult to find few documents referring to it in a popular municipality. The probability grows and peaks to the expected value of pure chance. Then it starts becoming lower, as it is difficult that all documents related to  $t_i$  appear in the same municipality. With a non monotonic function, we cannot have a simple rule stating “If the function value is low, the relation is significant”, because low values can be generated both by particularly strong or particularly weak relations. In other words,  $PMF(t_i = n) < PMF(t_i = n + 1)$  does not hold. Instead, being the cumulative distribution function monotonic (i.e. the following relationship holds:  $CDF(t_i = n) < CDF(t_i = n + 1)$ ), we can say that the higher the value, the stronger the relation (until the theoretical maximum of 1).

## 5 Robustness Tests

We validate the results obtained by MOGO and our filter to create knowledge by using it to identify the areas of operation of Mexican politicians, i.e. we substituted drug trafficking organizations with Mexican state governors. If MOGO is returning real relationships between actors and locations, each state governor should be mainly found to operate in the municipalities of the state she is governing.

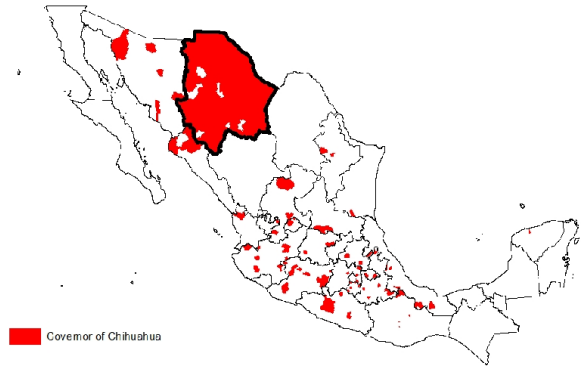
In Figure 4 we depict the relations between the municipalities and three governors<sup>14</sup>. We highlighted in red all the municipalities for which the chosen governor has a *CDF* equal to or higher than 0.95. We also highlighted the state governed by the politician by enlarging the thickness of its borders. In Figure 4(a) we highlighted the relationships between municipalities and the governor of Chiapas, in Figure 4(b) the governor of Chihuahua and finally in Figure 4(c) the governor of Durango. We can see that in all the three cases the governors are related mostly, if not entirely, to the municipalities of their state. Furthermore, the amount of municipalities they are related which are outside their state is very low.

As Figure 4 shows, MOGO indeed yields an accurate depiction whether someone operates in a particular location. In the next section we provide some

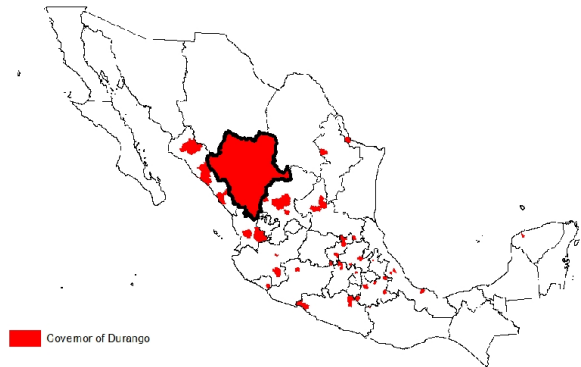
<sup>14</sup>At the time in which we performed the validation searches, i.e. November 2011



(a) Governor of Chiapas.

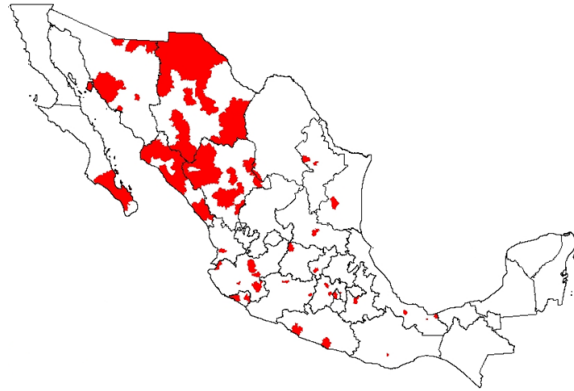


(b) Governor of Chihuahua.

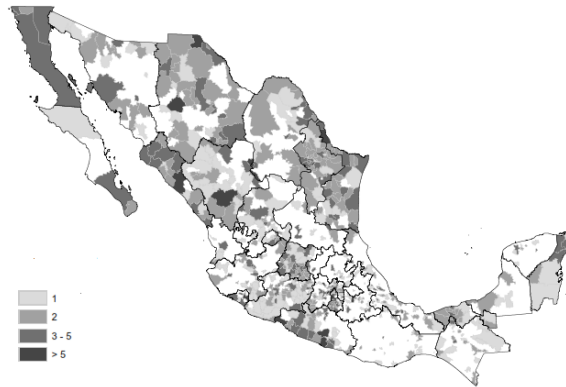


(c) Governor of Durango.

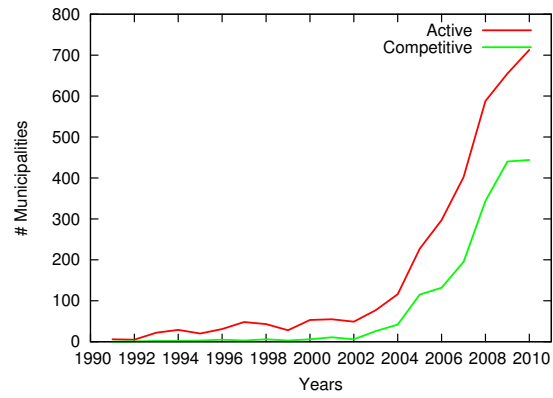
Figure 4: The politician-municipality significant relations. The corresponding state has been highlighted with a thicker black border.



(a) Sinaloa-active municipalities.



(b) The total active municipalities in Mexico.



(c) Active and Competitive municipalities.

Figure 5: Some evidences about the traffickers' activity patterns.

examples of the usefulness of the knowledge extracted in the description of the Mexico's criminal organizations.

## 6 Results

### 6.1 Where do Mexican drug trafficking organizations operate?

The results of the search allowed us to provide the research community with information on the behaviour of 13 trafficking organizations in Mexico, particularly about their municipalities of operation, their migration patterns and their market strategies for a period of 19 years (1991 - 2010).

The disaggregation up to the municipal level allowed us to challenge the widespread assumption that drug traffickers control vast regions of Mexico's territory dividing the country in oligopolistic markets [42]. Instead, we show that traffickers select their areas of operation with finer detail (Figure 5a). The Sinaloa Cartel, a drug trafficking organization which was previously thought to operate in all the state of Sinaloa [38], only operates in 14 of the 18 municipalities of the state, the same case goes to Juarez Cartel that only operates in 28 of 66 municipalities at its home state (Chihuahua), and to La Familia that only operates in 69 of 115 Michoacan.

Actually, according to our results, drug trafficking organizations only operate in 713 of 2,441 municipalities in Mexico. Large areas the country completely lack of the presence of a drug trafficking organizations. Our data changes our understanding of criminal territoriality, showing that drug trafficking organizations pick their areas of operation quite selectively. They concentrate in areas that are closer to ports of entry to the US, large cities within Mexico, and highways that connect cultivation areas or maritime ports to the US-Mexico border (Figure 5b).

This insight matches evidence of criminal activity tracked by the Mexican government in the form of drug-related violence. The Trans-Border Institute analyzed official figures of homicides caused directly or indirectly by the activities of drug trafficking organizations in Mexico<sup>15</sup> and concluded that less than 44% of the municipalities in 2011 had experienced drug-related homicides, a large increase from the 15% municipalities presenting this type of violence in 2007 but still away from arguing that criminal organizations operate in all of the 2,500 municipalities of Mexico.

Our results also provide the first systematic evidence of changes in the territoriality of criminal organizations over time. Rather than provide cross-sectional

---

<sup>15</sup><http://justiceinmexico.files.wordpress.com/2010/07/2012-tbi-drugviolence.pdf>

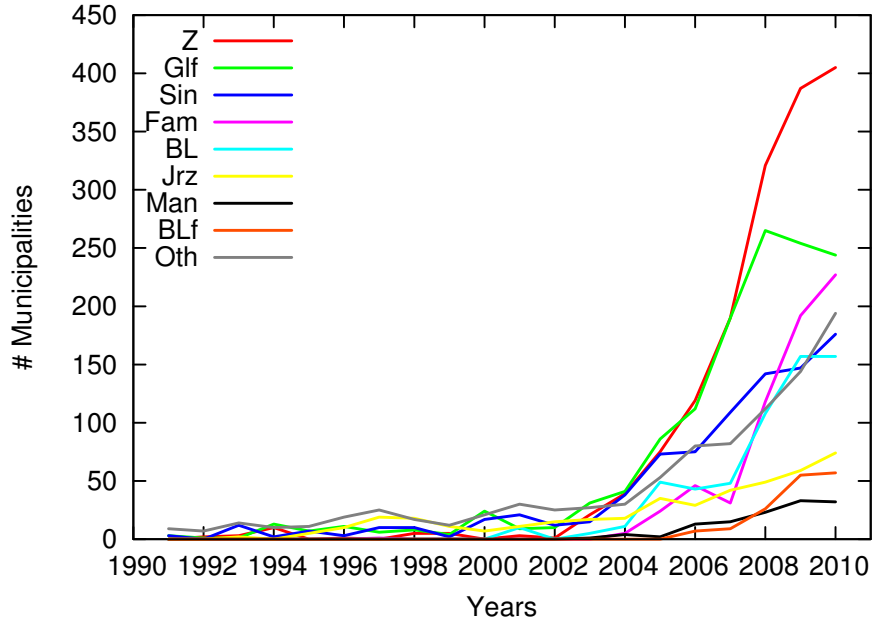


Figure 6: Different activity patterns per trafficker.

data at a particular point of time as [21], we show panel data for almost two decades (Figure 5c). The increase in the number of municipalities with drug traffic activity is evident. Furthermore, our information provides the first portrayal of the market structure of the illegal drug trafficking within Mexico and of its changes over time. Mexico’s organized crime is not an oligopoly as the theoretical literature of organized crime and private protection rackets assumes, rather drug trafficking organizations share territories frequently. As of 2010, 444 (62%) of all municipalities with trafficking operations had more than one criminal organization operating simultaneously a significant increase from a decade ago when only 6 (11%) were competitive (Figure 5c). The market structure of organized crime in Mexico is increasingly competitive; many criminal groups operate in the same territories sharing accesses to highways and ports of entry to the US.

Because our results were designed to provide information about criminal operations by organization, we are able to show the inner behavior of criminal organizations at a level of detail that remained previously unknown. We depart from considering that all organized groups are the same, and that their motivations can be understood under the same logic, and instead provide evidence of important operational differences between them. Criminal groups differ quite substantially in their territorial extension, the strategies that follow to expand, and their patterns of migration.

Org	2010 Terr.	Start Year	Terr.	Abandoned	Years operated	Competitive	Exploratory	$k$
Juarez	74	1997	13.85	10.15	2.78	-0.67	-0.03	1
Tijuana	39	1997	10.1	8.15	2.74	-0.96	-0.21	1
Sinaloa	176	1993	25.6	16.95	2.84	0.18	0.64	1
Barbie	66	2006	5.75	2.45	1.63	-0.48	-0.72	2
Mana	32	2006	3.8	2.2	2.15	-0.82	-0.73	2
Sinaloa faction	53	2008	5.15	2.5	1.96	-0.67	-0.70	2
BL faction	57	2008	5	2.15	1.79	-0.52	-0.75	2
Other	24	2008	2.15	0.95	1.38	-0.99	-0.75	2
BL	157	2004	18.65	10.8	2.08	0.81	-0.36	3
Fam	227	2005	18.75	7.4	2.09	0.95	0.01	3
Golfo	244	1994	35.55	23.5	3.01	1.25	1.07	4
Zetas	405	2003	42.2	21.95	2.71	1.94	2.55	4

Table 1: The main features extracted for each drug trafficking organization.

The timing, velocity and degree of expansion are significantly different between different organizations (Figure 6a, please note that this figure is different from Figure 2 as the latter included raw results before the cleaning stage). The most expansionary drug trafficking organizations started their spread in 2003 (Golfo, Zetas), a second group took off in 2005 (Sinaloa, Beltran-Leyva (BL) and Familia), and a final one, which included most of the criminal organizations in Mexico took place in 2007<sup>16</sup>. The most rapid expanders, organizations like La Familia, three-folded the municipalities at which they operated in just two years; slower expanders, like Sinaloa, took six years to expand similarly. Before 2004, all trafficking organizations operated in less than 50 municipalities. By 2010, Zetas had expanded their presence to more than 400, while others like Juarez Cartel have never operated in more than 80 municipalities. The number of municipalities in which each trafficking organization operated as of 2010 and the year in which they stated operating is shown in Table 1 (first two columns).

## 6.2 How do Mexican drug trafficking organizations operate?

Using the information provided by MOGO we can classify Mexico’s drug trafficking organizations according to their market strategies.

Table 1 reports identified characteristics of the Mexican trafficking organizations. Two of them (number of municipalities in 2010 and year of appearances, first and second columns) were already discussed in the previous subsection. The remaining five are: average number of municipalities in which the trafficking organization starts to operate in each year (third column), average number of municipalities abandoned by the organization in each year (fourth column), average number of years in which the organization operates in a municipality (fifth column), competitive and exploratory indexes (sixth and seventh column). We present each of them and we explain how we use them to cluster organizations in different homogeneous cartel types ( $k$  column).

While some organizations operate in many municipalities simultaneously

<sup>16</sup>These last expansion coincides with increases in prosecution launched after the arrival of a new political administration in December of 2006 at the national level.



others have significantly smaller areas of operation. Zetas appear in an average of 42.2 municipalities every year, while Mana only in three. Second, the average number of years in which an organization operates consecutively in a municipality goes from 3.01 for Golfo, to 1.63 for Barbie. Third, trafficking organizations also tend to abandon markets with quite high variance. While organizations like Sinaloa abandon about 16.95 municipalities on average, the BL faction only abandons 2.15. Finally, we also identified two more dimensions: the propensity to explore new municipalities, and their preferences towards engaging in competitive behaviour. We measured propensity to explore as the number of standardized municipalities in which a given drug trafficking organization was the first to ever operate, and preferences towards engaging in competitive behaviour as the standardized number of municipalities in which a given trafficking organization was sharing a municipality with another one. While organizations like Zetas have a high tendency to engage in competitive behaviour and in exploration of new territories, others like Tijuana and Juarez tend to avoid competition and exploration. The case of BL is particularly interesting, showing a tendency to explore new territories without engaging in competition.

We used all the different characteristics of drug trafficking organizations to classify them according to their *modus operandi* in four clusters. We create a matrix with the seven features from Table 1. We scale and normalize each feature to avoid that one or some of the features may dominate over the others. Then, we perform several runs of the k-means algorithm for varying  $k$ s. For each  $k$  we calculate the within groups sum of squares, to determine the appropriate number of clusters, that in our case is  $k = 4$ . We then return the four main classes of drug cartels according to our data.

The four classes are: 1 = “Traditional”, 2 = “New”, 3 = “Competitive” and 4 = “Expansionary Competitive”. The results show quite differentiated strategies (Table 1) between each of them. A first cluster, integrated by Juarez, Tijuana, and Sinaloa is integrated by “Traditional” trafficking organizations that have operated in Mexico for the longest time. These organizations have a tendency towards being not competitive, being most of the time the first to operate in a particular territory. They operate in a large number of municipalities but also have a high turn over.

A second cluster, integrated by la Mana, fractions of Sinaloa and BL, Barbie and a residual category, are the “New” organizations. On average, they emerged in 2007, more than ten years after the first cluster. They operate in a very reduced number of municipalities. They are not competitive and do not have a very developed tendency towards exploring new territories. This means their market strategy consists in operating in municipalities that had once been controlled by other criminal organizations but had been abandoned.

A third category is integrated by BL and Familia, two criminal organizations that are relatively new (created on average in 2004), operate in many municipal-

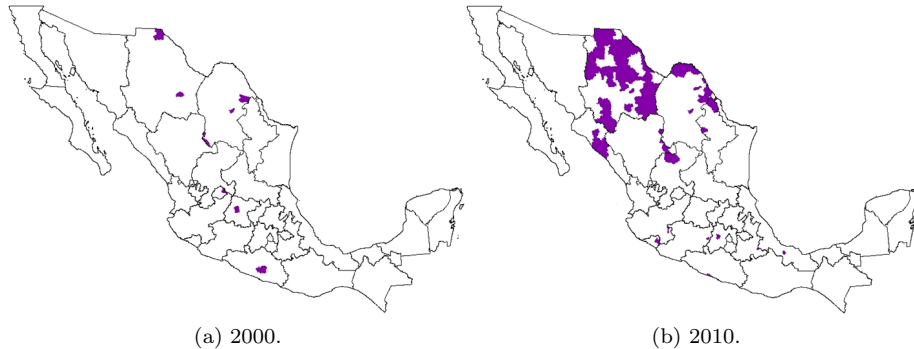


Figure 7: The migration pattern for Juarez cartel.

ities, and have strongly competitive tendencies. We called these organizations “Competitive” because they do not explore new territories but rather operate in places where another organizations is already operating.

This tendency towards invading territories that are already taken is even stronger for the fourth cluster, integrated by Zetas and Golfo organizations. We called these organizations “Expansionary competitive” because they are not only the most competitive but also the ones with the largest tendencies to explore new territories. In other words, they do not only try to invade others’ territories but also are the first to colonize new markets and to operate in areas where drug trafficking organizations had never been present before. In general, this last cluster is the one with the largest criminal organizations, operating on average on 324 municipalities (as of 2012) and spreading to an average of 38.87 new municipalities every year. Yet, it is also important to mention that their mobility is also the largest, they abandon an average of 22 municipalities per year, lasting only an average of 2.86 years in each one of them.

These differences match qualitative analysis done by ethnographers and specialist in Mexican criminal organizations. Journalistic accounts of Tijuana Cartel [5] and Juarez Cartel [12] had previously identified these organizations as localized, with few interests in moving towards other territories and being proactively competitive. Sinaloa cartel had been portrayed as a cartel with larger incentives to invade others although still mostly confined to the north of the country [38]. Both Golfo [40] and Zetas [16] cartels had recurrently being described as the most invasive and aggressive criminal organizations in terms of market expansion. Our findings also match qualitative evidence collected by experts on La Familia, BL and Barbie organizations.

Finally, patterns of migration also differ. While some tend to expand in their neighbouring territories, others expand more broadly around the country (Figures 7 and 8). Notoriously different patterns can be seen in Zetas (Fig-



Figure 8: The migration pattern for Zetas cartel.

ure 8), a drug trafficking organization that tends to migrate more randomly, extending over the whole territory and the a group of more localized criminal organizations like Juarez Cartel (Figure 7).

Overall, our exercise extracted significant knowledge about Mexico’s drug trafficking industry, and about the inner behaviours of different criminal organizations. The collected data further advances our knowledge of criminal organizations and provides quantitative evidence of criminal behaviour that was previously only qualitatively described.

## 7 Conclusion

In this paper we created a tool called MOGO to generate intelligence about where and how criminal organizations operate without large-scale, without expensive intelligence data-gathering exercises. Based on a simple three step process (list definition, query generation, and crawling), MOGO is able to create a knowledge by exploiting indexed reliable sources such as online newspapers and blogs.

We tested MOGO’s power in Mexico’s drug trafficking industry. After testing and cleaning the data extracted by MOGO, we were able to identify differentiated market strategies and areas of operation of thirteen Mexican drug cartels, providing the first map of criminal organizations available yearly (1991 - 2010) and at the municipal level. Information at these level of detail had never been available to academics before. Our tool represent an important advancement for political studies studying the location of criminal and hostile groups, as well as for security policymakers.

We showed that criminal organizations, rather than being similar and oper-

ate under identical mechanics, differ significantly in their market orientations. We identified four types of Mexican criminal organizations: traditional, new, competitive and expansionary competitive. Traditional organizations operate in municipalities that they control since long time ago, on average since 1995. New organizations have only being in operation since 2007 on average, and tend to operate in municipalities where other criminal organizations had at some time being present but were abandoned. Competitive organizations are those that operate in territories are controlled by other organizations. Finally, expansionary competitive are those not only operate in territories that were already taken but also explore new territories, expanding their operations to areas that were drug trafficking organizations had never operated before. Overall, our findings provide evidence that criminal organizations are more strategic and operate in more differentiated ways than current academic literature thought. Criminals act, move and operate according to the environment and resources that they possess.

This paper opens the way to much future work. Most immediately, the knowledge extracted by MOGO will be used by to identify patterns of criminal mobility within Mexico by linking different types of drug trafficking organization with theories of criminal mobility [33, 51]. Yet, in the near future we will apply MOGO to extract information about different criminal groups in other countries to provide a first worldwide comparable data set of criminal mobility, at the subnational level.

#### **Acknowledgements.**

Michele Coscia is a recipient of the Google Europe Fellowship in Social Computing, and this research is supported in part by this Google Fellowship.

Viridiana Rios was supported by Mexico's Office of the President, the Center for US-Mexico Studies at the University of California in san Diego, and the Program in Inequality and Criminal Justice at Harvard Kennedy School.

Special thanks to Cesar Hidalgo, Gary King, Peter Bol and Ricardo Hausmann for valuable feedback. Andrea Fernandez, Stephanie Kuhl and Scott Reed were crucial for compilation and acquisition of data, and Laura Kelly for editing. Brad Holland, Abigail Martinez and Hugo Rios gave invaluable support and encouragement in Mexico City, during the Spring of 2011, when this data gathering endeavour first started. Theirs is this paper.

## **References**

- [1] Mohammad Ali Abbasi, Shamanth Kumar, Jose Augusto Andrade Filho, and Huan Liu. Lessons learned in using social media for disaster relief - asu crisis response game. In *SBP*, pages 282–289, 2012.
- [2] Jay Albanese. Organized crime in america. 1989.

- [3] P. Andreas. Smuggling wars: Law enforcement and law evasion in a changing world. *Transnational crime in the Americas: An inter-American dialogue book*, pages 85–98, 1999.
- [4] P. Andreas and J. Wallman. Illicit markets and violence: what is the relationship? *Crime Law Soc Change*, 52:225–229, 2009.
- [5] Jesus Blancornelas. El cartel: los arellano félix. In *Plaza y Janés*, 2002.
- [6] G. Bruinsma and W. Bernasco. Criminal groups and transnational illegal markets: a more detailed examination on the basis of social network theory. *Crime, Law and Social Change*, 41(1):79–94, 2004.
- [7] Jharna Chatterjee. The changing structure of organized crime groups. 2005.
- [8] R. V. Clarke and R. Brown. International trafficking in stolen vehicles. *Crime and justice: A review of research*, pages 197–227, 2003.
- [9] A. M. Cohen, W. R. Hersh, C. Dubay, and K. Spackman. Using co-occurrence network structure to extract synonymous gene and protein names from MEDLINE abstracts. *BMC bioinformatics [electronic resource]*, 6(1), April 2005.
- [10] Paul Collier and Anke Hoeffler. Greed and grievance in civil war. 2001.
- [11] John M. Collins. Military geography for professionals and the public. 1998.
- [12] Francisco Cruz. El cartel de juarez. In *Editorial Planeta*, 2009.
- [13] F. J. Desroches. The crime that pays: Drug trafficking and organized crime in canada. 2005.
- [14] B. John Evans. Criminal networks, criminal enterprises. 1994.
- [15] J.D. Fearon and D.D. Laitin. Ethnicity, insurgency, and civil war. *American Political Science Review*, 97(1):75–90, 2003.
- [16] Jorge Fernandez Menendez and Victor Ronquillo. De los maras a los zetas: los secretos del narcotráfico, de colombia a chicago. In *Ediorial Grijalbo*, 2006.
- [17] Olivier Ferret. Discovering word senses from a network of lexical cooccurrences. In *Proceedings of the 20th international conference on Computational Linguistics*, COLING, Stroudsburg, PA, USA, 2004. Association for Computational Linguistics.
- [18] H.R. Friman. Drug markets and the selective use of violence. *Crime, law and social change*, 52(3):285–295, 2009.
- [19] Justin Grimmer. A bayesian hierarchical topic model for political texts: Measuring expressed agendas in senate press releases. *Political Analysis*, 18(1):1–35, 2010.

- [20] Justin Grimmer and Brandon M. Stewart. Text as data: The promise and pitfalls of automatic content analysis methods for political texts. *Open Scholar*, 2012, to appear.
- [21] Eduardo Guerrero. Security, drugs and violence in mexico: A survey. In *7th North American Forum*, 2011.
- [22] R. Hebert-Burns. Compound piracy at sea in the early twenty-century. In Laurence Whitehead, editor, *Violence at Sea: Piracy in the Age of Global Terrorism*. Routledge, 2002.
- [23] D.E. Kaplan and A. Dubro. *Yakuza: Japan's criminal underworld*. University of California Press, 2003.
- [24] John Keegan. A history of warfare. 1993.
- [25] Gary King and Will Lowe. An automated information extraction tool for international conflict data with performance as good as human coders: A rare events evaluation design. *International Organization*, 57:617–642, 2003.
- [26] E. R. Kleemans and H. G. van de Bunt. The social embeddedness of organized crime. *Transnational Organized Crime*, 5(1):19–36, 2002.
- [27] D. Kyle and M. Scarcelli. Migrant smuggling and the violence question: evolving illicit migration markets for cuban and haitian refugees. *Crime, law and social change*, 52(3):297–311, 2009.
- [28] Sang Hoon Lee, Pan-Jun Kim, Yong-Yeol Ahn, and Hawoong Jeong. Googling social interactions: Web search engine based social network construction. *PLoS ONE*, 5(7):e11233, 07 2010.
- [29] Douglas Lemke. The tyranny of distance: Redefining relevant dyads. *International Interactions*, 21(1):23–38, 1995.
- [30] Bing Liu and Lei Zhang. A survey of opinion mining and sentiment analysis. In *Mining Text Data*, pages 415–463. 2012.
- [31] S. Morrison. The dynamics of illicit drugs production: future sources and threats. *Crime, Law and Social Change*, 27(2):121–138, 1997.
- [32] C. Morselli and K. Petit. Law-enforcement disruption of a drug importation network. *Global Crime*, 8(2):109–130, 2007.
- [33] Carlo Morselli, Mathilde Turcotte, and Valentina Tenti. The mobility of criminal groups. *Report No. 004*, 2010.
- [34] Williamson Murray. Some thoughts on war and geography. *Journal of Strategic Studies*, 22(2-3):201–217, 1999.

- [35] M. Natarajan. Understanding the structure of a large heroin distribution network: A quantitative analysis of qualitative data. *Journal of Quantitative Criminology*, 22(2):171–192, 2006.
- [36] M. E. J. Newman. Power laws, pareto distributions and zipf’s law. *Contemporary Physics*, 46:323–351, December 2005.
- [37] Brendan O’Connor, Ramnath Balasubramanyan, Bryan R. Routledge, and Noah A. Smith. From tweets to polls: Linking text sentiment to public opinion time series. In *Proceedings of the International AAAI Conference on Weblogs and Social Media*, 2010.
- [38] Diego Enrique Osorno. El cartel de sinaloa: Un historia del uso politico del narco. In *Random House Mondadori*, 2009.
- [39] L. Paoli and P. Reuter. Drug trafficking and ethnic minorities in western europe. *European Journal of Criminology*, 5(1):13–37, 2008.
- [40] Ricardo Ravelo. Osiel. vida y tragedia de un capo. In *Ediorial Grijalbo*, 2009.
- [41] John A. Rice. *Mathematical Statistics and Data Analysis*. Duxbury Press, April 2001.
- [42] R. Rodríguez Castañeda. El méxico narco. In *Editorial Planeta*, 2010.
- [43] Markus Schedl, Tim Pohle, Peter Knees, and Gerhard Widmer. Exploring the music similarity space on the web. *ACM Trans. Inf. Syst.*, 29(3):14:1–14:24, July 2011.
- [44] Maximillian Schich and Michele Coscia. Exploring co-occurrence on a meso and global level using network analysis and rule mining. *Minign and Learning with Graphs, KDD Workshop*, 2011.
- [45] Mark Shaw. Organized crime in post-apartheid south africa. *Organized Crime: Uncertainties and Dilemmas*, 1994.
- [46] Louise Shelley. Identifying, counting and categorizing transnational criminal organizations. *Transnational Organized Crime*, 1:1–1, 1999.
- [47] M. V. Simkin and V. P. Roychowdhury. Theory of Aces: Fame by chance or merit? *eprint arXiv:cond-mat/0310049*, October 2003.
- [48] Arthur Spirling. Us treaty-making with american indians. *American Journal of Political Science*, 56(1):84–97, 2012.
- [49] E. Tagliacozzo. Contraband and violence: lessons from the southeast asian case. *Crime, law and social change*, 52(3):243–252, 2009.
- [50] P. C Van Duyne. Organized crime markets in a turbulent europe. *European Journal on Criminal Policy and Research*, 1(3):10–30, 1993.

- [51] F. Varese. *Mafias on the move: How organized crime conquers new territories*. 2011.
- [52] K. von Lampe. The cigarette black market in germany and in the united kingdom. *Journal of Financial Crime*, 13(2):235–254, 2006.
- [53] Klaus Von Lampe. Transnational organized crime challenges for future research. *Crime, Law and Social Change*, 2012.
- [54] P. Williams and R. Godson. Anticipating organized and transnational crime. *Crime, Law and Social Change*, 37(4):311–355, 2002.
- [55] Phil Williams. Organizing transnational crime: Networks, markets and hierarchies. *Transnational Organized Crime*, pages 57 – 87, 1998.
- [56] D. Zaitch. Trafficking cocaine: Colombian drug entrepreneurs in the netherlands. *The Hague: Kluwer Law International*, 2002.