Knowing Where and How Criminal Organizations Operate Using Web Content

Michele Coscia\textsuperscript{1} and Viridiana Rios\textsuperscript{2}

\textsuperscript{1}KddLab - ISTI CNR - michele.coscia@isti.cnr.it
\textsuperscript{2}Department of Government - Harvard University - vrios@fas.harvard.edu
The Intelligence Problem

Information complexity critically affects the ability of security agencies to collect intelligence information by making it more costly.
Criminal activities happen every day...
But they are often blurred

Criminal organizations usually have high interest in hiding their activities from the public
However...
Moreover...
EL CADÁVER DEL ENJUZADO NAZARIO MOLERO GÓMEZ
5 MILLONES DE DÓLARES, EUROS O MONEDAS DE OTRO VALOR

JESUS MENDEZ VARGAS
5 MILLONES DE DOLARES, EUROS O MONEDAS DE OTRO VALOR

ENRIQUE TECATEL
ALIAS "EL PUEBLE", 500 MIL DOLARES

QUIEN INFORME QUE ME LLAME A SU CÁRCEL, SERÉ COMO LOCALIZARLOS

CUANDO ME NADA PARA TODO LOS CRIMENES

NO DEJA MARAVILLOSA LOS ECOTAMOS

NUESTRA LUCHA
Problems

• The Web's knowledge is:
  – Too large to be analyzed as a whole;
  – Subject to reliability concerns.
Our Approach

Already deals with data selection and cleaning to index only reliable and relevant news

We design an intelligence-oriented, topic-specific automatic query and data cleaning process

We derive an operative picture of the world and we test it against ground truth
MOGO: Making Order using Google as an Oracle
Entity Definition Rules

• Firstly define the “actor lists”
  – In our case: drug traffickers' names and municipalities

• Classify words into:
  – Generic
  – Common
  – Unique
  – State Unique
First Query Generation Step

• A query is either:
  – Unique word by itself (because unambiguous)
  – State Unique + State Name (it makes it unambiguous)
  – Prune out Generic and Commons

• Other rules to avoid ambiguity (i.e. a municipality with the same name of a state)
The Popularity Check

• Pruning out clear outliers that escaped the query generation step
  – Exaggerate popularity
  – Extremely high deviation
Data Statistics
Fat-tail distributions
Logarithmic growth
Data Cleaning & Validation
Cleaning

• Just because a trafficker co-appears a lot with a municipality, it does not mean that the connection is strong
• We validate the results against a null model
\[ PMF(t_i = \bar{t}_i) = \frac{ \begin{pmatrix} T_i \\ \bar{t}_i \end{pmatrix} \begin{pmatrix} M - T_i \\ \bar{m}_i - \bar{t}_i \end{pmatrix} \begin{pmatrix} M \\ \bar{m}_i \end{pmatrix} }{ \begin{pmatrix} M \\ \bar{m}_i \end{pmatrix} } \]

\[ CDF(t_i = \bar{t}_i) = \sum_{a=0}^{\bar{t}_i} PMF(t_i = a) \]
Validation using politician activities
Results
Drug-Active Municipalities
Competition Among Cartels

![Graph showing the number of municipalities over years]

- Active Cartels
- Competitive Cartels
# Drug Cartels Classification

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

The migration pattern for Zetas cartel.
Thank You

Questions?