Analyzing the Social Media Footprint of Street Gangs

Presented at IEEE International Conference on Intelligence and Security Informatics (IEEE ISI 2015)
Baltimore, MD, USA, May 27-29, 2015

1Sanjaya Wijeratne
sanjaya@knoesis.org

1Derek Doran
derek@knoesis.org

1Amit Sheth
amit@knoesis.org

2Jack L. Dustin
Jack.dustin@wright.edu

1Ohio Center of Excellence in Knowledge-enabled Computing (Kno.e.sis)
2Center for Urban and Public Affairs (CUPA)
Wright State University, Dayton, OH, USA
What does gang related research tell us?

“Gangs use social media mainly to post videos depicting their illegal behaviors, watch videos, threaten rival gangs and their members, display firearms and money from drug sales” [Patton 2015, Morselli 2013]

Studies have shown,

✓ 82% had used the Internet and 71% of them had used social media [Decker 2011]
✓ 45% participated in online offensive activities and 8% recruited new members online [Pyrooz 2013]
Gang related research cont.

“Someone says something to me on Facebook, I don't even write a word. The only thing I do is post my 30-popper, my big banger”¹

¹http://www.wired.com/2013/09/gangs-of-social-media/all/
There are other spectators too…

Eight out of every 10 law enforcement professionals (81%) actively use social media as a tool in investigations.

40% use it to monitor special events

34% use it to notify the public of crimes

34% use it to notify the public of emergencies or disasters

But there’s a problem...

Over half (52%) of agencies still don’t have a formal process for using social media for investigations. The single biggest barrier to use is a lack of access during work hours.

A platform to analyze social media

- Monitor negative community effects of gang activities
- Discover opinion leaders who influence the thoughts and actions of other gang members
- Evaluate the sentiment of posts targeting communities, locations, and groups
Architecture of our platform

Data Collection & Filtering
- Tweets
- Gang Violence Related Tweets and User Profiles Filtering
- Data Cleaning & de-identification

Data Processing
- Entity Identification
- Entity Disambiguation
- Sentiment Extraction
- User Location Estimation
- Social Network Analysis

Data Access Tools
- Search and Browsing
  - User Profile Explorer
  - Tweet Content Explorer and Search
  - User Network Explorer

Data Analysis & Interpretation
- Example Questions:
  - Eg1. Who are the violent gang leaders operate in Chicago area?
  - Eg2. What is the sentiment of gang A towards gang B?
  - Eg3. Who are the other gang members that user A communicates with and what are their locations?
  - Eg4. What are the gangs user A is affiliated with?

Slang Term Dictionary
ML
NLP
Data Store
## Data collection

<table>
<thead>
<tr>
<th></th>
<th>Gang Related Dataset</th>
<th>Location Related Dataset</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Data Collection</strong></td>
<td>91 known gang member profiles from a Chicago gang (Gang X) + Gang related keywords (eg. #BDK, #GDK etc.) [Decary-Hetu 2011]</td>
<td>10 neighborhoods in Chicago - South Landale, North Landale, West Elsdon, Gage Park, West Lawn, Chicago Lawn, New City, Humboldt Part, Logan Square and Belmont Cragin</td>
</tr>
<tr>
<td><strong>APIs</strong></td>
<td>Twitter REST + Twitter Streaming</td>
<td>Twitter Streaming</td>
</tr>
<tr>
<td><strong>Size</strong></td>
<td>$56,368 + 49,079 = 105,447</td>
<td>383,656</td>
</tr>
<tr>
<td><strong>Geo Information</strong></td>
<td>57.07%</td>
<td>100%</td>
</tr>
</tbody>
</table>
Spatio-Temporal-Thematic Analysis

lemme hear you say Gang X and we finna murder you

On location_name We Drillin Fuck Da Opps

@user1 @user2 check out this 7414 track,

url_to_file
People-Content-Network Analysis

Nodes – 1,322
Edges – 19,539
Avg. Degree – 14.77
# of Clusters – 72
Modularity – 0.859
Avg. Nodes in a Cluster – 15
Sentiment-Emotion Analysis
Twitter profile analysis
Challenges and Future work

- Automatically identify gang member profiles on Twitter
- Automatically detect beefing among gangs
- Integrate additional slang dictionaries
- Hipwiki.com
- Improve sentiment analysis algorithm
- Evaluating the findings
- Beefing identified by the system VS beefing taken place in the real world
- Filter noisy data (Wijeratne 2014)

Questions?

Connect with me

sanjaya@knoesis.org
@sanjrockz
http://bit.do/sanjaya
References


