

# Human Rights Event Detection from Heterogeneous Social Media Graphs

**Daniel B. Neill**  
**H.J. Heinz III College**  
**Carnegie Mellon University**  
**E-mail: [neill@cs.cmu.edu](mailto:neill@cs.cmu.edu)**

We gratefully acknowledge funding support from the National Science Foundation, grants IIS-0916345, IIS-0911032, and IIS-0953330, and the John D. and Catherine T. MacArthur Foundation.

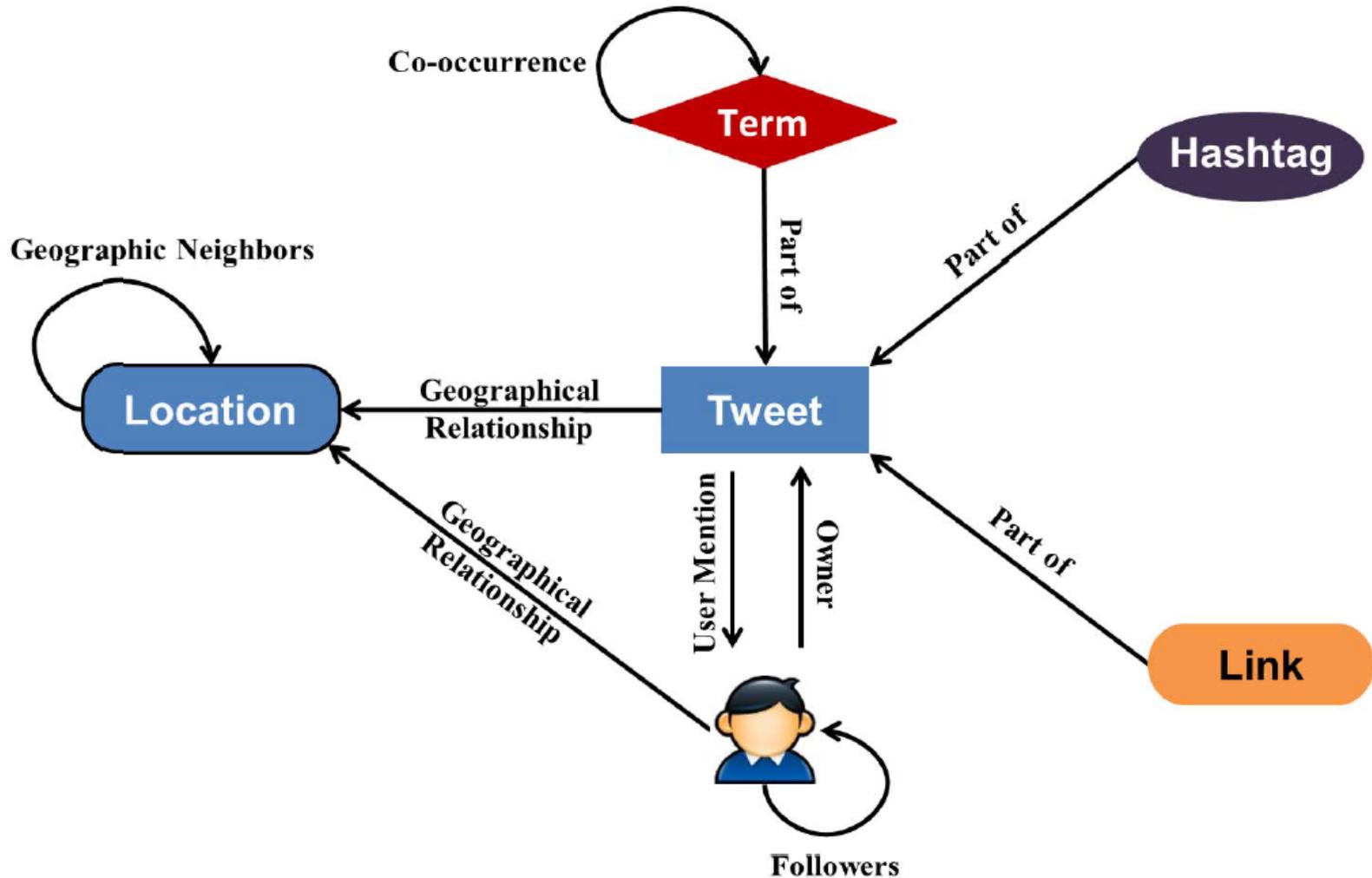
Carnegie Mellon University

**EPD Lab**

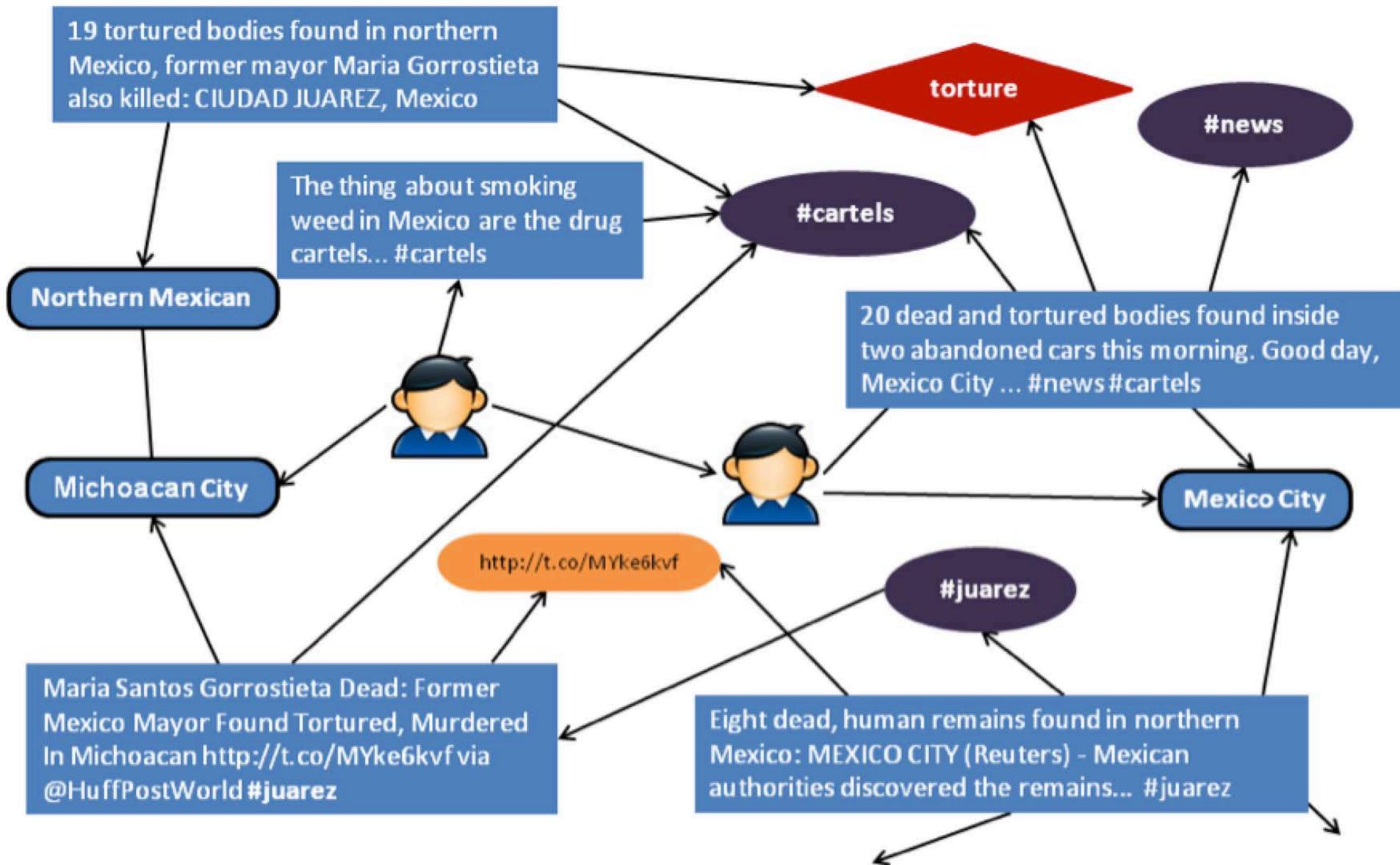
EVENT AND PATTERN DETECTION LABORATORY



# Twitter Heterogeneous Network



# Twitter Heterogeneous Network



# Step 1: Sensor network modeling

Each node reports a value measuring its current level of anomalousness for each time interval (hour or day).

Object Type	Features
User	# tweets, # retweets, # followers, #followees, #mentioned_by, #replied_by, diffusion graph depth, diffusion graph size
Tweet	Klout, sentiment, replied_by_graph_size, reply_graph_size, retweet_graph_size, retweet_graph_depth
City, State, Country	# tweets, # active users
Term	# tweets
Link	# tweets
Hashtag	# tweets

Features

empirical  
calibration

Individual p-value  
for each feature

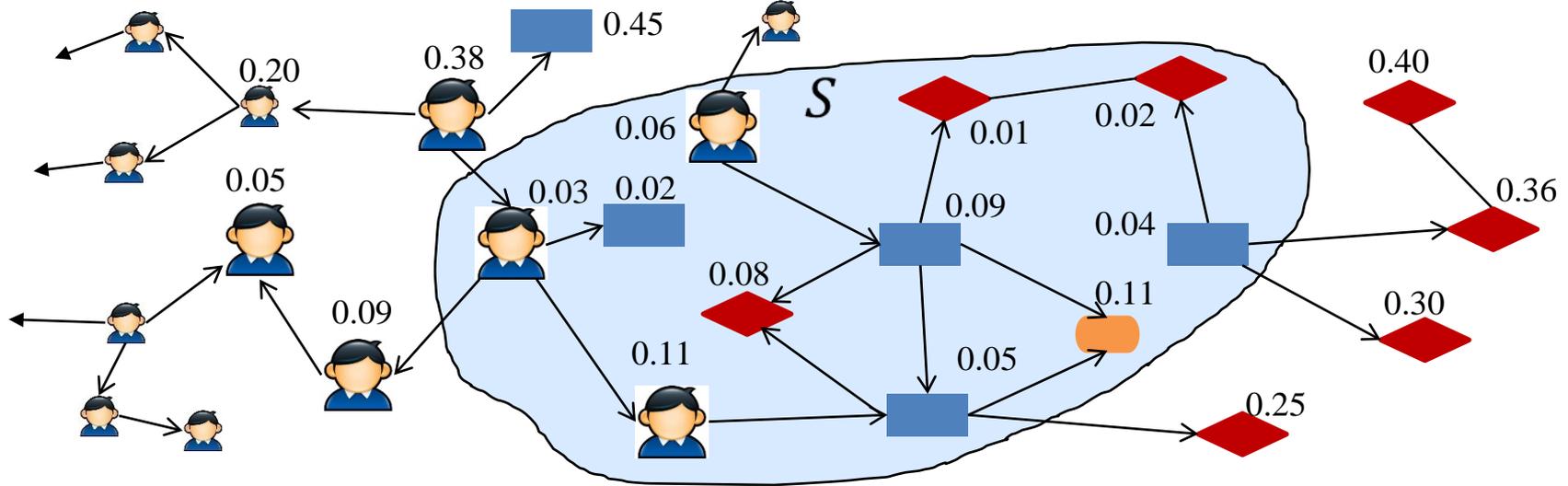
min

Minimum  
empirical p-  
value for  
each node

empirical  
calibration

Overall p-value  
for each node

# Step 2: Nonparametric graph scan



$$S^* = \underset{S \in V: S \text{ is connected}}{\operatorname{argmax}} F(S)$$

This step allows us to find groups of nodes (users, keywords, tweets, hashtags, etc.), that are most anomalous when considered collectively.

# Empirical Results

We performed an exploratory analysis of human rights-related events in Mexico from January 2013 to June 2014, using Twitter data (10% sample, filtered using relevant keywords).

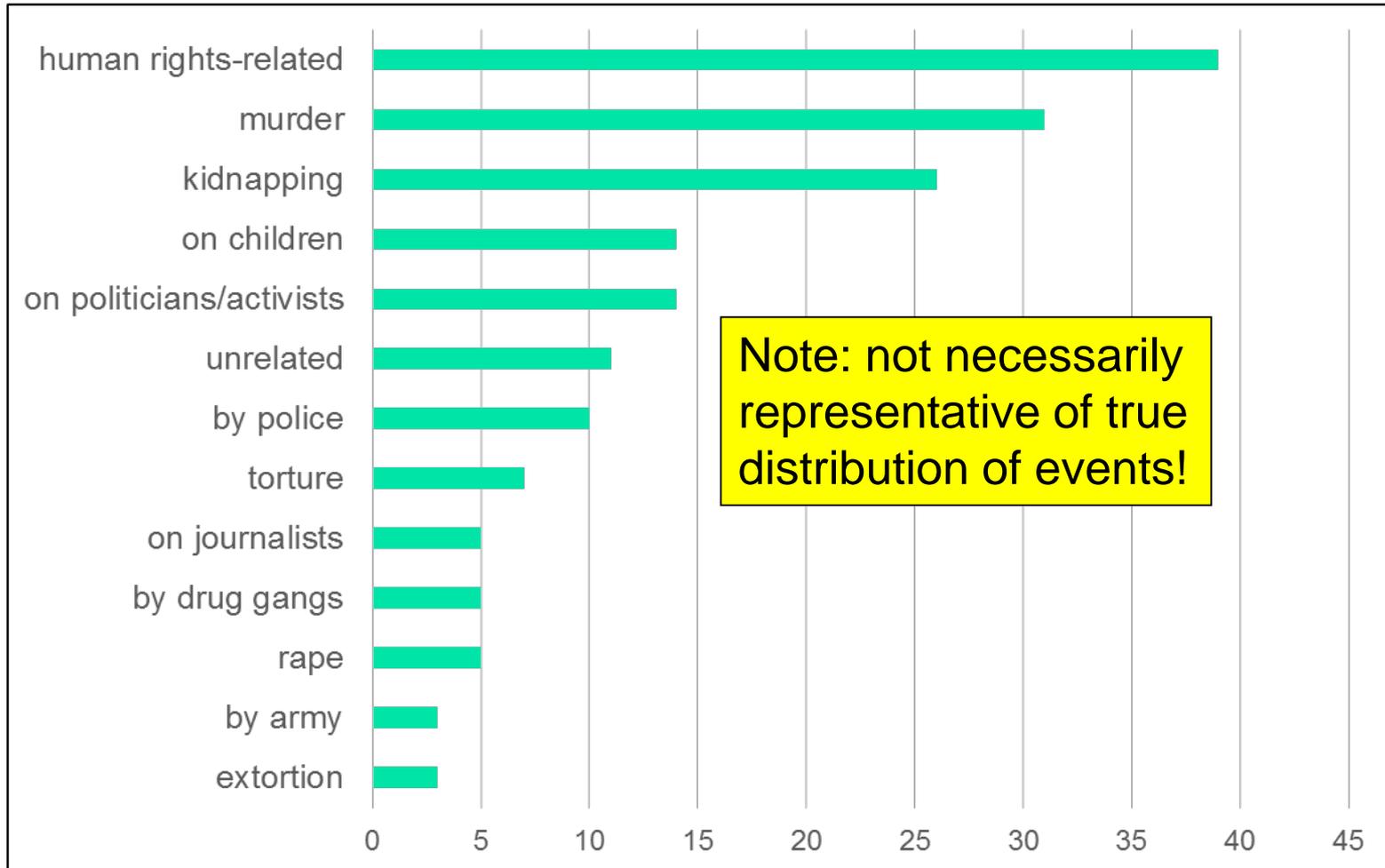
The top 50 identified clusters over the entire study period were analyzed manually to identify:

- (1) whether the cluster was human rights related
- (2) the types of human rights violations
- (3) the victims of the violations
- (4) the alleged perpetrators.

NPHGS was able to identify some human rights events of interest before international news sources...  
... and in some cases, before local news sources.

# Cluster characteristics

(top-50 detected clusters)



# For more details...

## Methodology:

- F. Chen and D.B. Neill. Non-parametric scan statistics for event detection and forecasting in heterogeneous social media graphs. *Proc. 20th KDD Conf. on Knowledge Discovery and Data Mining*, 2014.

## Application to human rights:

- F. Chen and D.B. Neill. Human rights event detection from heterogeneous social media graphs. *Big Data* 3(1): 34-40, 2015.

Both papers are accessible from the Event and Pattern Detection Laboratory website:

<http://epdlab.heinz.cmu.edu>

Interested in learning more? Feel free to contact me:

[neill@cs.cmu.edu](mailto:neill@cs.cmu.edu)