



USC University of  
Southern California



# Cannabis Surveillance with Twitter Data: Emerging Topics and Social Bots

Jon-Patrick Allem, Ph.D.

# Background

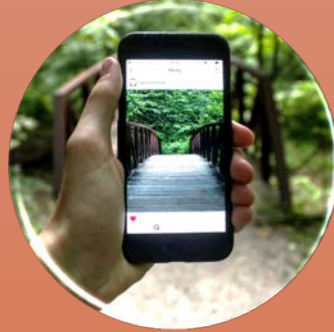
- State-level medical cannabis and adult recreational use laws have lowered prices and increased the availability of cannabis.
- Health implications associated with cannabis use:

Acute health outcomes	Repeated cannabis use
Impaired short term memory	Dependence
Impaired attention	Increased risk for other substance abuse
Impaired coordination	Increased risk for schizophrenia
Sleep problems	

# Access to massive real-time focus groups



Near real time data  
collection



Unprimed



Massive numbers



Own voice



The goal of this study was to identify and describe cannabis-related topics of conversation on Twitter to inform the public health community.

# Data Collection

- Tweets containing cannabis-related terms (e.g. “marijuana”, “THC”)
- Initial sample = 53,117,048 tweets
- Duration: May 1- December 31, 2018

Initial sample

## Data cleaning

- Retweets
- Non-English tweets
- Unrelated tweets (e.g. Emily Blunt, hash brown)
- Social bots

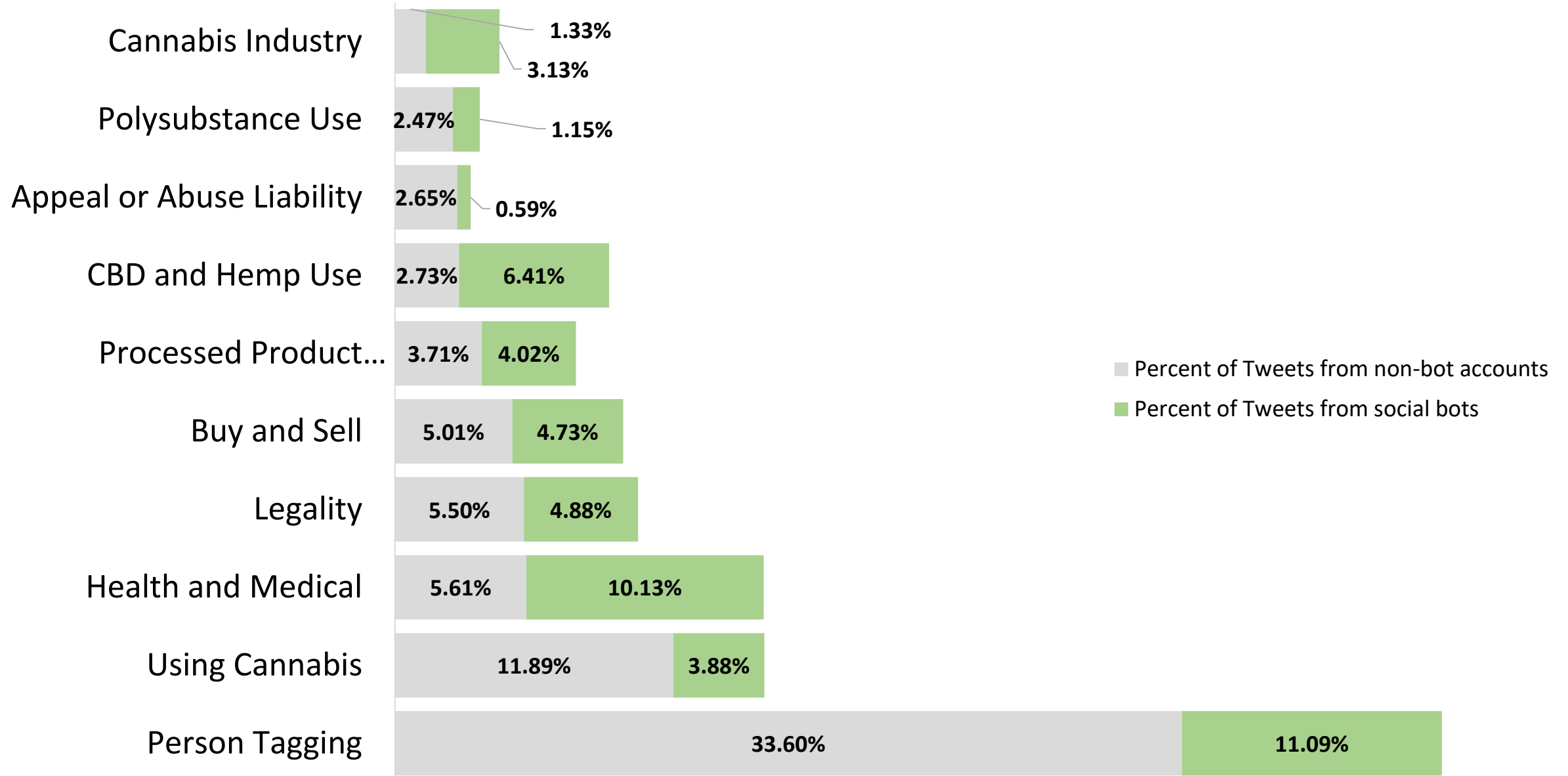
- n = 60,861 tweets from 47,760 non-bots
- n = 8,874 tweets from 2,871 social bots

Final samples

# Data Preparation

- 1) **Basic normalization** e.g., lower case all text, remove punctuation
- 2) **Stop word removal** e.g., words such as “a,” “the,” etc.
- 3) **Normalization of Twitter user mentions** e.g., “@johnsmith” is converted to “@person”
- 4) **Lemmatization** e.g. “dog,” “dogs,” “dog’s,” and “dogs’” are all converted to “dog”
- 5) **Non-printable character removal** e.g., emojis and symbols from other languages

# Results



# Example Tweets

Person  
Tagging

"I can't wait to smoke a blunt with @person"

Polysubstance  
Use

"Pass me the beer and bong"

Health and  
Medical

"Cannabis cures triple negative breast cancer"

Legality

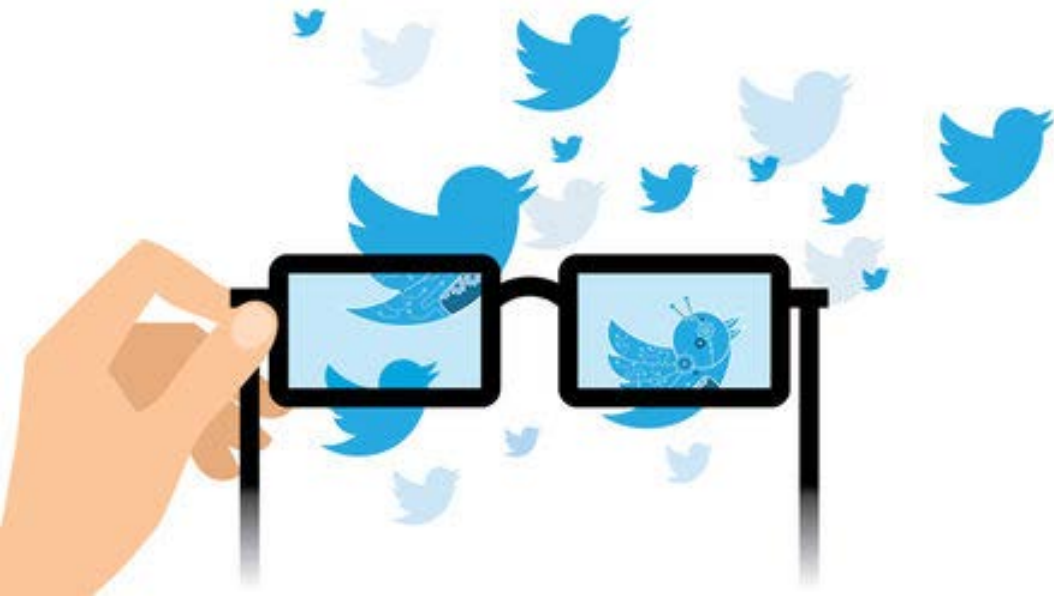
"California now allows cannabis when flying within state lines"

Buy/Sell

"Message me for the best weed"



# Discussion and Implications



- Social bots perpetuating unsubstantiated health claims pose a public health concern.
- Online exposure to pro-cannabis messaging may impact perceived risks, initiation, and further use.
- Polysubstance use may be a target for future surveillance.

# Acknowledgements

The Regents of the University of California, Research Grants Program Office, Tobacco-Related Diseases Research Program, Grant Number No.28KT-0003.

Allison Dormanesh

Likhith Dharmapuri

Patricia Escobedo

**THANKS!**

**JON-PATRICK  
ALLEM**

**E: [allem@usc.edu](mailto:allem@usc.edu)**

**Twitter: [@somalabusc](https://twitter.com/somalabusc)**

**Visit: [somalab.usc.edu](http://somalab.usc.edu)**

