

# Big Data Analytics and Crimes in the Community

Kevin Wang (Criminology)

Yasin Yilmaz (Electronic Engineering)

Morris Chang (Electronic Engineering)

Mahsa Mozaffari (Electronic Engineering)

Emre Yilmaz (Electronic Engineering)



USF

# Development of Information Technology

- Cyberspace – the virtual space constructed by inter-connected computing and storage devices
- Optical communication
  - How much time does it take to transmit the encyclopedia of 50 million words across the Pacific Ocean in one fiber?
  - Less than 1/10,000 second
- Trends: cloud computing, big data, mobile revolution, IoT

# Societal Impact

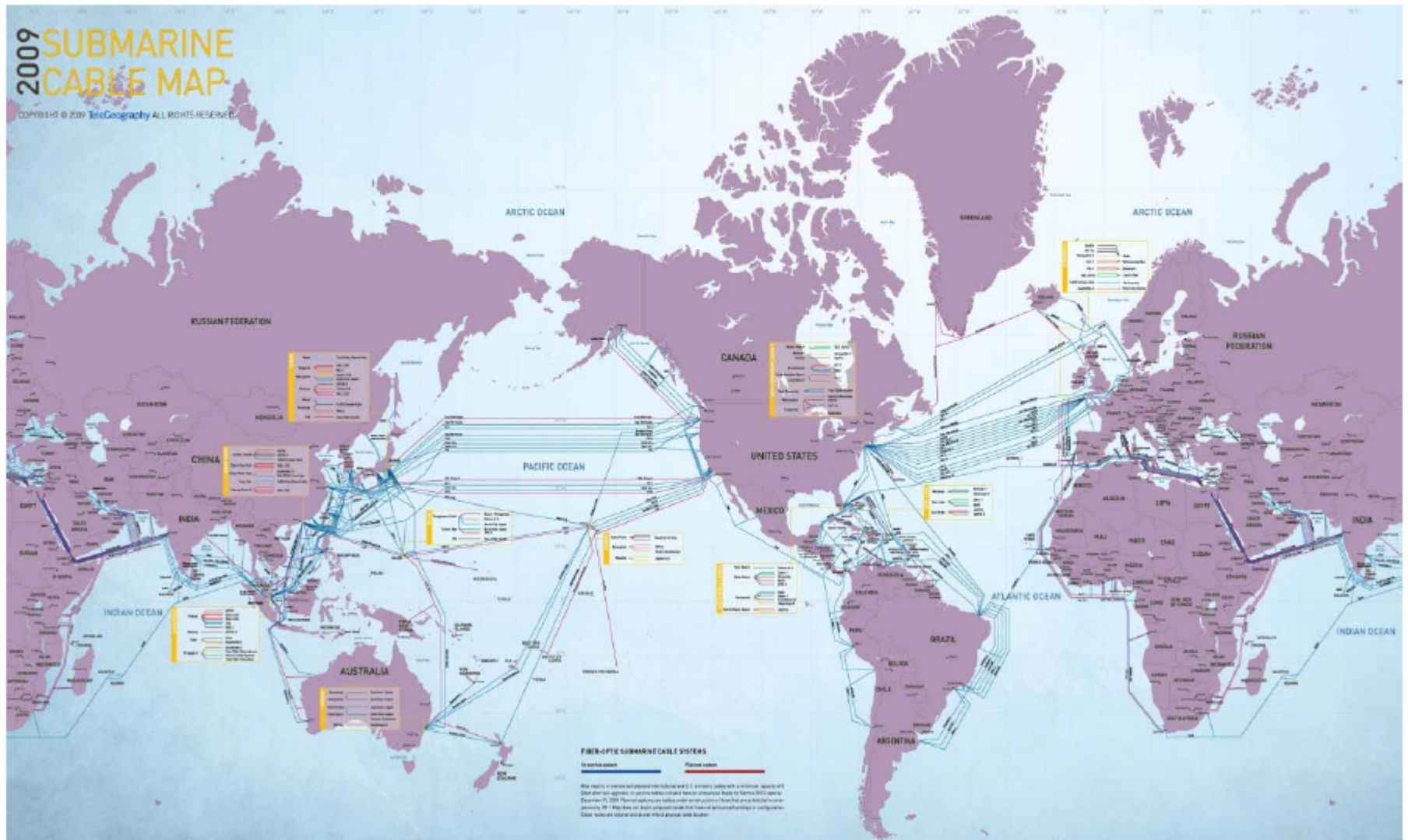
## Optical Fiber Everywhere

### Undersea

# Societal Impact

## Optical Fiber Everywhere

### Undersea



# “Big Data”

- Types: health records, government records, GPS data, sensor data, call logs of mobile phones, digital traces, online transactions/activities, activities on social media
- 4Vs of “big data”
  - Volume
  - Velocity
  - Variety
  - Veracity

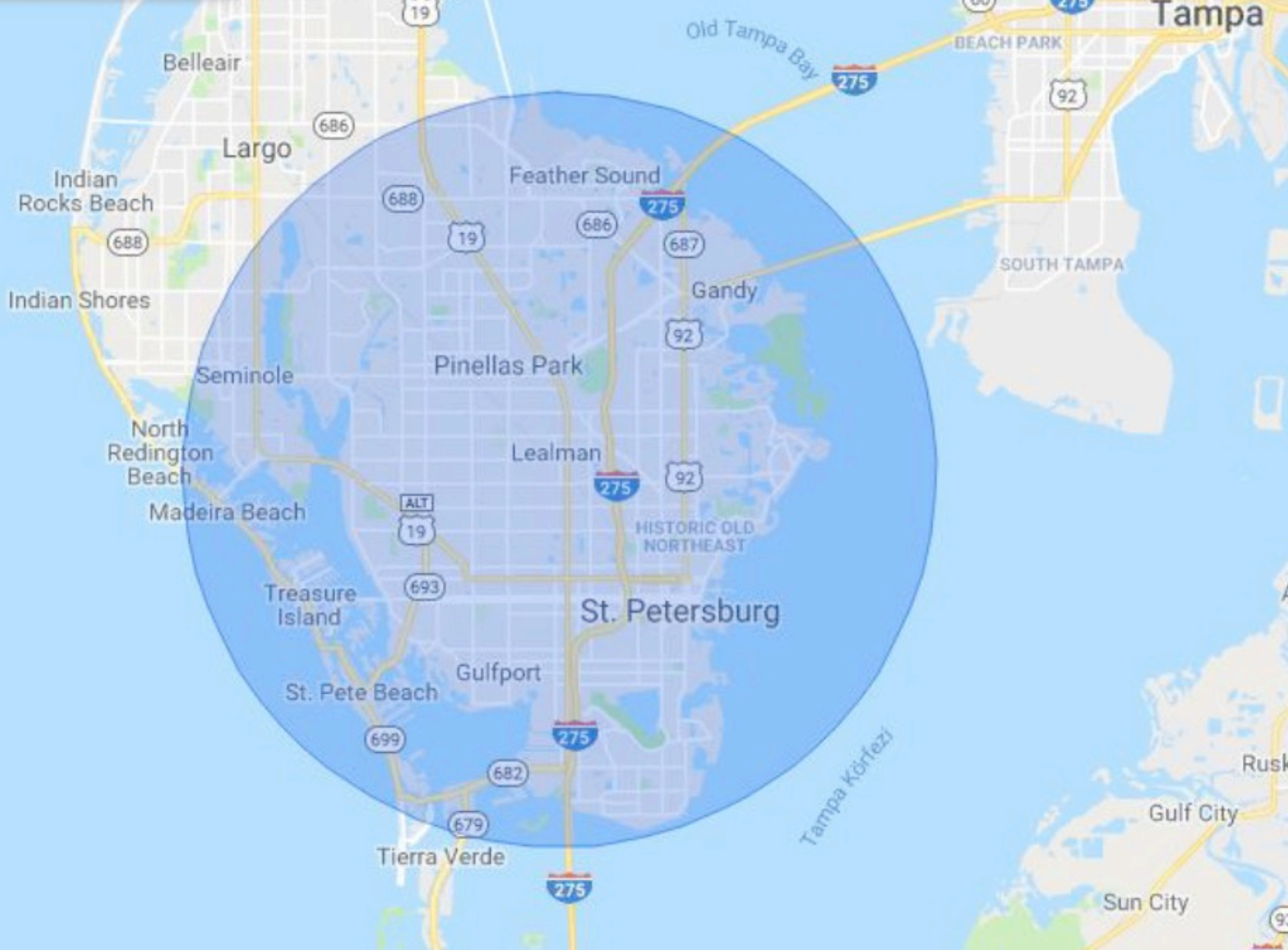
# “Big Data” and Challenges

- Representational validity
- Big data generated from social media like Twitter are contributed by active users
- Mobile devices – users
- Unknown uncertainty in geographic location

# Data & Methods

- Explorative nature of the study with data from St. Petersburg, FL between 2012 and 2016
- Crime data were collected from St. Petersburg Police Department (SPPD)
  - Crimes are grouped and aligned with UCR
- Tweet data were collected from APIs (application programming interfaces) by March, 2018
  - Contributors are viewed as human sensors





# Results – lexical approach

1	panhandler, beggar, vagrant, tramp, hobo, homeless, loiter, wander, unhoused, vagrancy, homelessness, vagabond, roam, sprawl, prowler, prowl, houseless, loiterer, wanderer
2	graffiti, mural, spraypaint, vandalism
3	drunk, bottle, alcohol, alcoholic, intoxicated, overdose
4	disturb, noisy, yell, shout, annoy, nuisance, argue, scream, noise, chaos, fight
5	smell, stink, stench, pee, urine, wee-wee, weewee, piss, poo, poop, defecate, urinate, spit, vomit, puke, excrement
6	steal, theft, thief, robbery, burglary, stolen, burglarize, trespassing, invasion, thieves, robber, burglarizing, burglar, burgle, thief, loot, trespasser
7	harass, assault, threaten, attack, threat, harassment, abuse, danger, risk, harm, offend, offender, vigilant, harmless, aggression, rape, lewd, abusive, harassing, stalker, offense, harassing
8	aggressive, aggressively
9	trash, garbage, litter, litterbug, dump, paraphernalia, dumpster, filth, dirty, junk, rat
10	addict, addiction, druggie, addicted, junkie
11	meth, methamphetamine, marijuana, cocaine, drug, heroin, dealer, smoke, amphetamine, weed, smoker, narcotic



# Results – lexical approach

12	gun, handgun, armed, shoot, weapon, stab, knife, shooting, unarmed, holster, rifle, stabbings, firearm, taser
13	violence, violation, ransack, rage, violent, attacker, clash
14	unsafe, uneasy, worry, intimidate, nervous, hassle, scare, frighten, frustrate, scared, insecurity, discomfort, unease, dread, appall, intimidation, terrorize, intimidating
15	police, patrol, cop
16	gang, gangster
17	stranger, suspect, motorist, creep, teenager, freak, creepy, suspicious, teenage, mob, riffraff, juvenile
18	murder, killing, kill, killer, victim, injure, injury, kidnapping, carjacking
19	insult, curse, swear, disrespect, unruly
20	collision, accident, speed, traffic, incident
21	crime, misdemeanor, felony

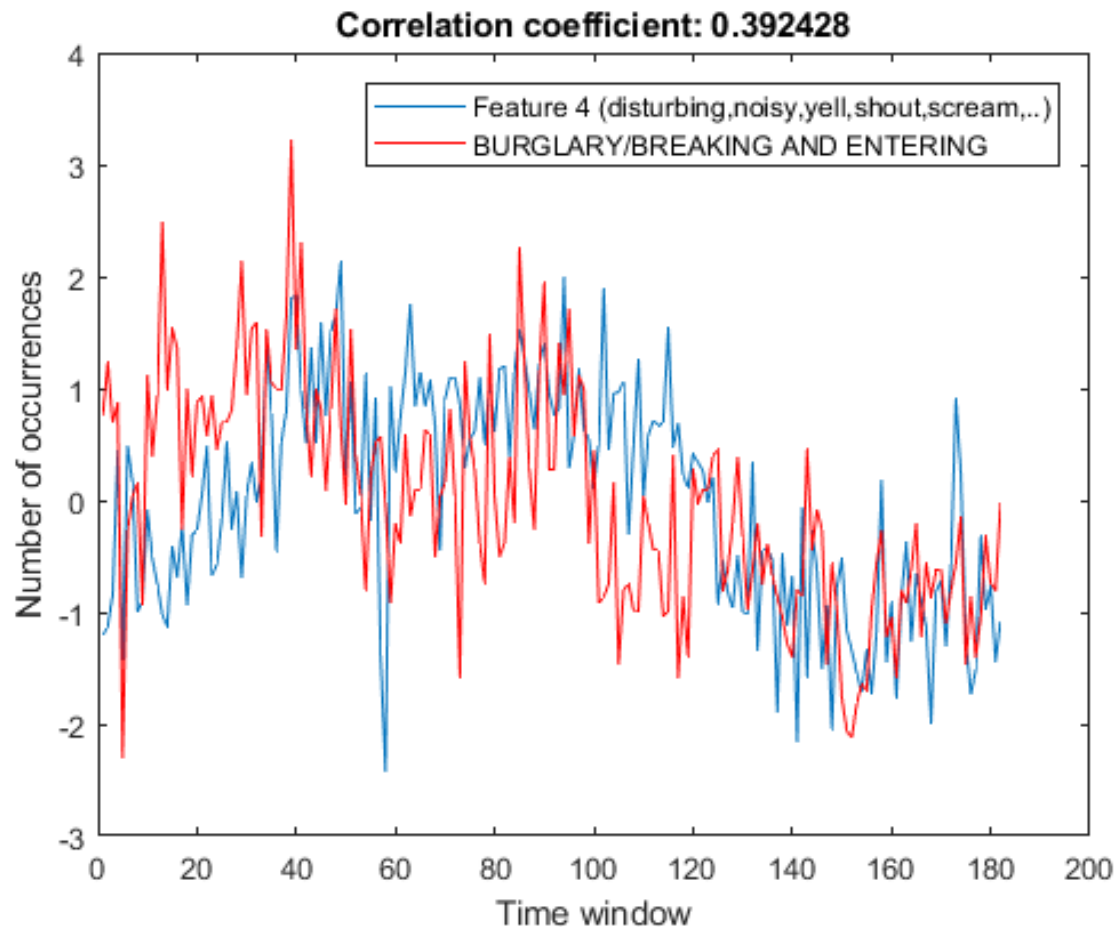
# Data & Methods

- A time lag between tweets and the occurrence of crimes
- The correlation between the average occurrence of words between days  $i$  and  $i+W_f$  with the average number of crimes between the days  $i$  and  $i+W_I$ 
  - $W_I = 10$  and  $W_f = 5$

# Results

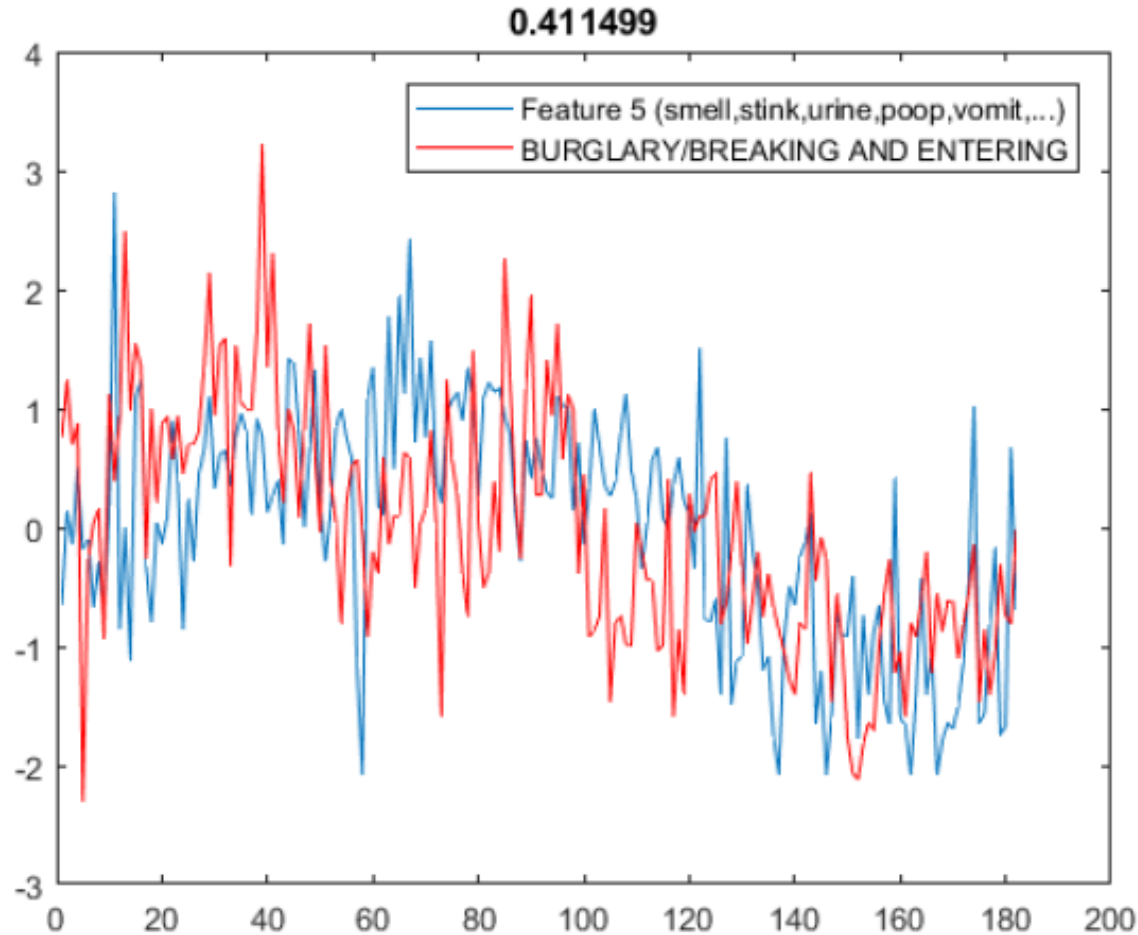
- 198,460 incident reports in 5 years
  - Following the national downward trend (burglary drops from 3,432 to 2,382 incidents)
- In total, there were 2,373,780 tweets in 5 years
  - The daily tweets was nearly 2,000, but it dropped to 500 from April 28, 2015
  - A policy and product updates aimed at combating abuse announced April 21, 2015

Results – indicators of social disorder (disturb, noisy, yell, shout, annoy, nuisance, argue, scream, noise, chaos, fight)

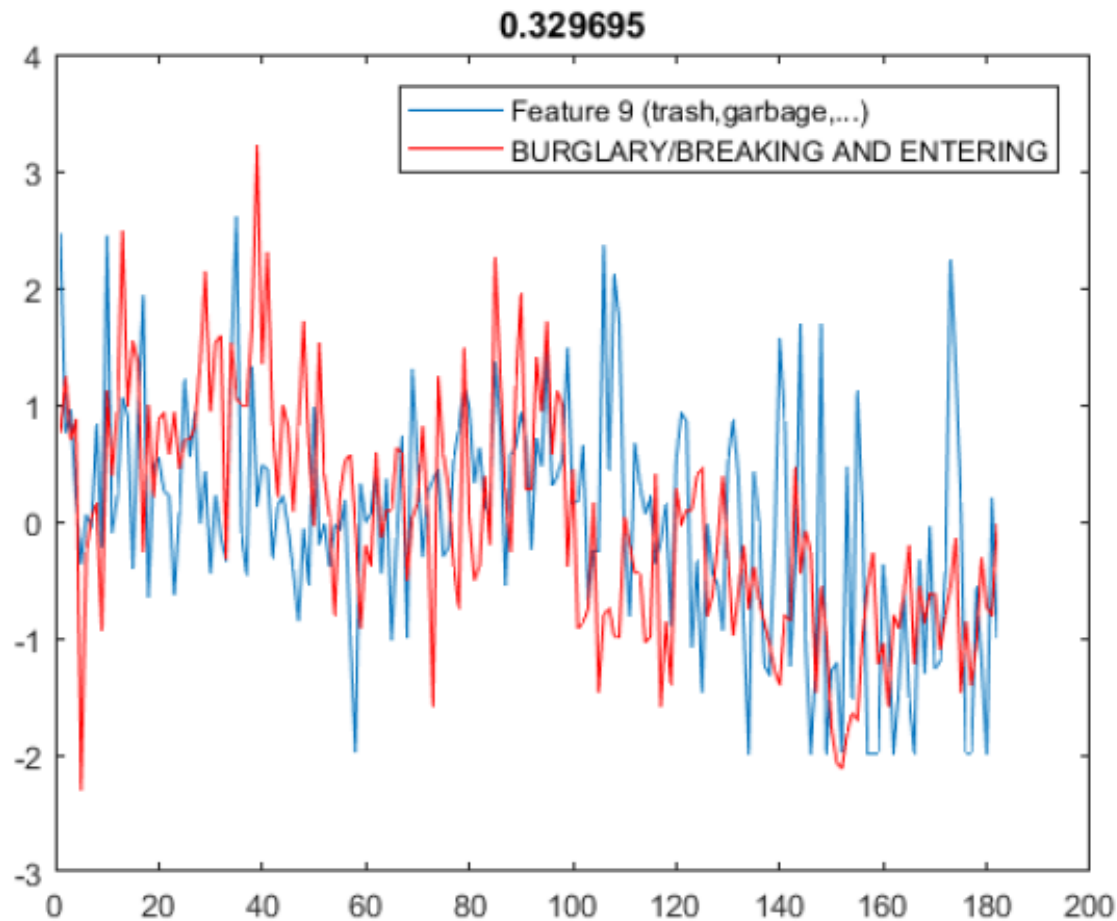


## Results – indicators of physical disorder 1

(smell, stink, stench, pee, urine, wee-wee, weewee, piss, poo, poop, defecate, urinate, spit, vomit, puke, excrement)



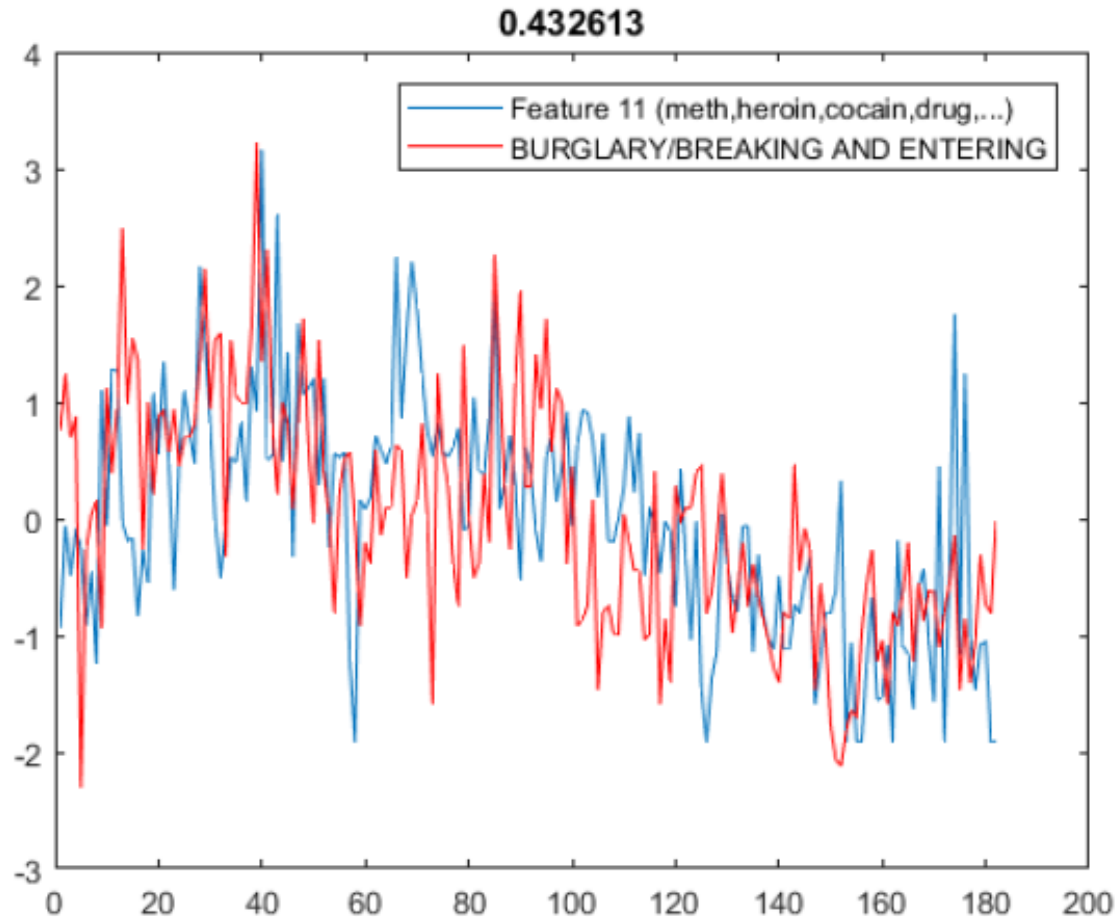
## Results – indicators of physical disorder 2 (trash, garbage, litter, litterbug, dump, paraphernalia, dumpster, filth, dirty, junk, rat)



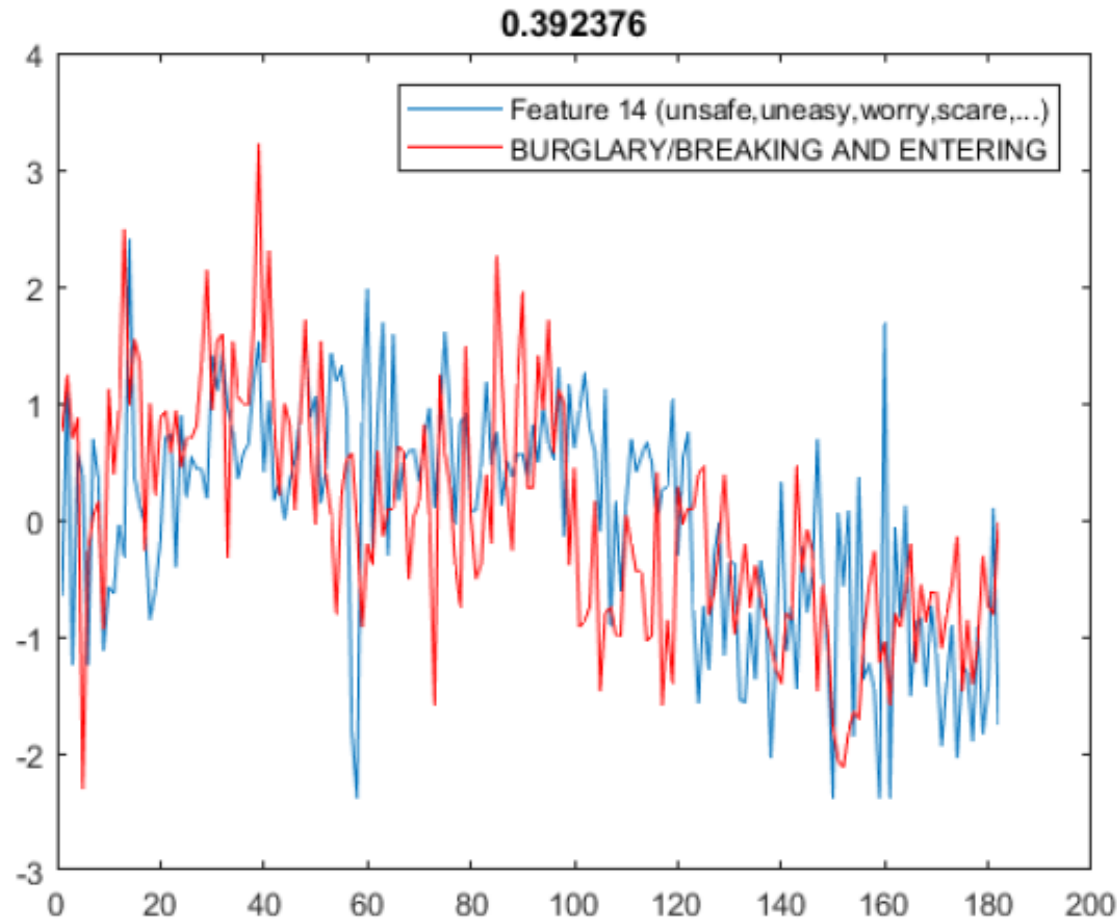


## Results – indicators of drugs

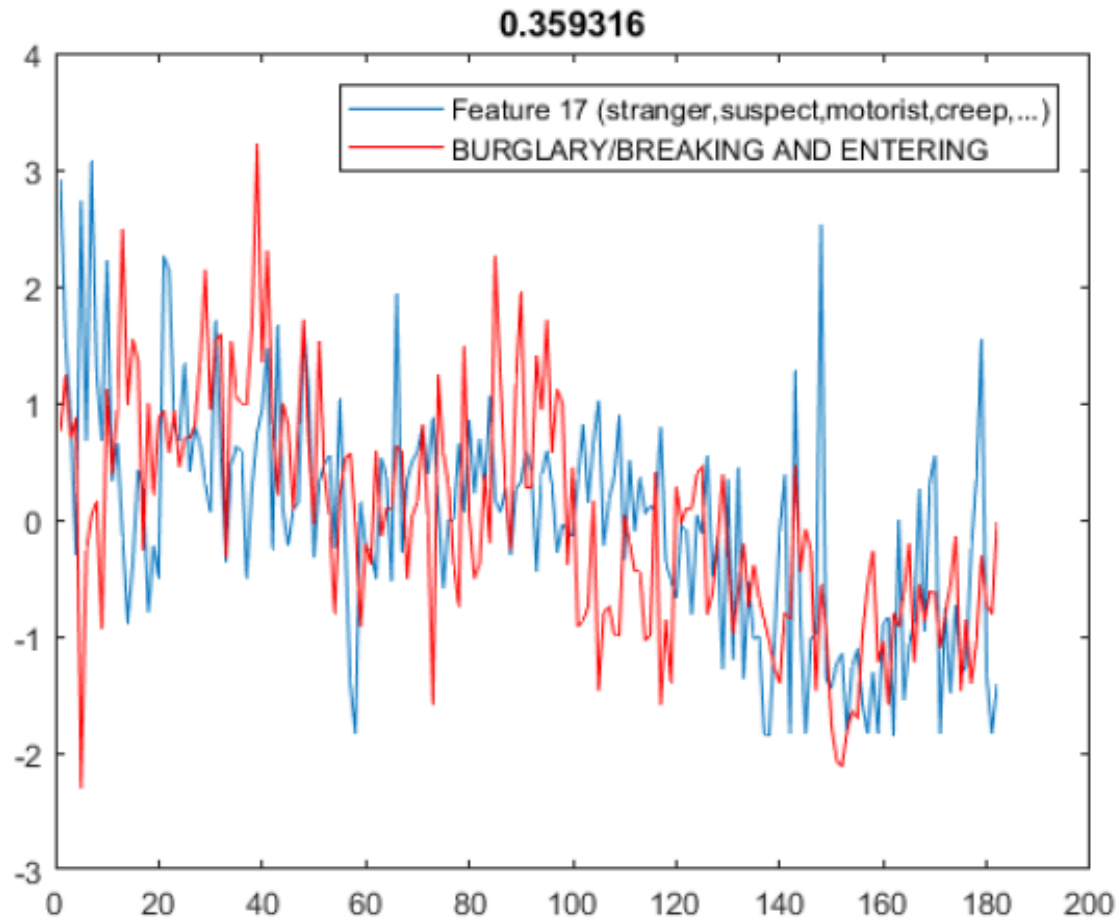
(meth, methamphetamine, marijuana, cocaine, drug, heroin, dealer, smoke, amphetamine, weed, smoker, narcotic)



Results – indicators of perception (unsafe, uneasy, worry, intimidate, nervous, hassle, scare, frighten, frustrate, scared, insecurity, discomfort, unease, dread, appall, intimidation, terrorize, intimidating)



Results – indicators of potential suspect  
(stranger, suspect, motorist, creep, teenager, freak, creepy,  
suspicious, teenage, mob, riffraff, juvenile)



# Implications

- Theory
  - Collective efficacy
  - Broken window
- Policing