



Combatting Corruption and Promoting Equity in the Health Sector

July 10 - 11, 2019
UC San Diego

Detection and Characterization of Corruption and Bribery on Twitter using Machine Learning

Jiawei Li, MS
UCSD
Data Scientist

Twitter




- Popular online social network service
- 6,000 tweets / second
- 500 million tweets / day

Study Objective

- **Main objective:** Detect and characterize tweets related to corruption from potentially millions of messages
- **Sub-analysis:** Identify health corruption-related tweets

MTLStockPromoters @MTLStockPromo · 1 Feb 2019

A police investigation needs to be initiated into allegations of #AMF and #UPAC collusion bit.ly/2zrli2k #FabriceBenoit, #NathalyMarcoux, #pennystocks, #deals, #ericvanguyen, #Osler, #TSX, #TSXv



Ouellette says police probe into collusion allegations is necessary
What has now become a saga over the arrest of Chomedey MNA Guy Ouellette is dragging on even longer with Ouellette and his fellow whistleblower, Annie Trudel
montrealgazette.com

1:00 PM - 1 Feb 2019

Christine M Ngaruiya @c_ngaruiya · 9 Mar 2019

Inspiring & candid speech by Dr. Patricia Garcia, former #MOH of #Peru with a call to action for increased research on #corruption in #health sector, #funding & adding this to #research agendas. Comparison of priorities, w a little humor (ref pic) #CUGH2019 @CUGHnews #globalhealth



8:09 AM - 9 Mar 2019

U of T Public Health Retweeted

Trudo Lemmens @TrudoLemmens · 29 Sep 2017

Jillian Kohler on #pharma sector vulnerabilities to #CORRUPTION #pharmagov2017



Dina Balabanova Retweeted

Global Integrity @GlobalIntegrity · Jul 8

New #GI_ACE #blog from @LizDavidBarrett & @mihaly_fazekas: Using 'red flag' indicators to identify #corruption and analyse reform efforts in the #procurement process: globalintegrity.org/2019/07/08/red ...



Data Collection & Processing



Twitter Stream API



Keywords :
'bribery', 'bribe', 'corruption', 'nepotism', 'collusion', 'kickback'



Collect Period : January 15 – April 15



Data Size: 22,179,869 tweets



Data Processing: Clean text and build word dictionary

corruption	people	practicing	poverty	make	built
000	qjn	900	scandals	bout	strategy
it	think	3	7	fritvqaf9	million
	alleged	1	p	forget	
	get	years	law	naira	
	10	50	n	year	

democrats	job	people	lie	dmk black	part	clarred
vote	house	r	dove	day	cong	mainers
democracy	white	president	types	serve	lucky	
it	voices	forth	cornerstone			
	americans	lift	right			

BTM (Biterm Topic Model)

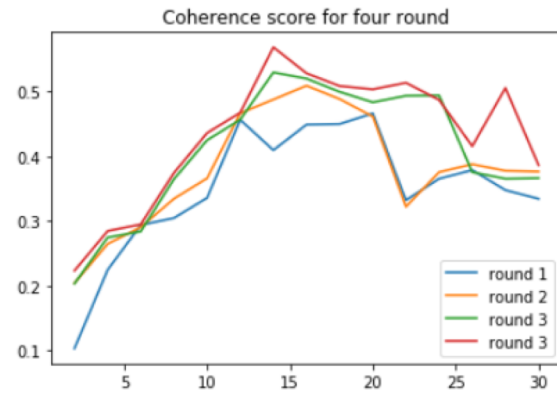
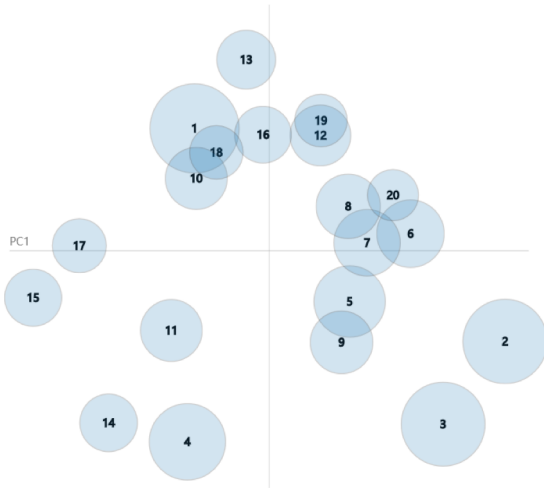
- Topic Model based on NLP (Natural Language Processing) and an unsupervised machine learning approach
- Categorize the data into k different clusters (k-topics)
- Objective of BTM is to take large volumes of unstructured data and conduct topic exploration to find signal



BTM Round 1&2

BTM Round 3


Coherence and BTM Parameters




Data Analysis

Results (Corruption)

- Number of tweets? 2054
- Corruption topics: bribery, police corruption, administrative (document processing), international development?

 **Manoranjan Shaw** @ShawManoranjan · Feb 8
@n
arendramodi


1 1


 **Manoranjan Shaw** @ShawManoranjan [Follow](#)

Sir I loose my judicial casee because my lawyer took bribe from my dependant and didn't appealed In the court of law and suppressed the matter and they evedted my property . So how can I get helm from you.


10:12 AM - 8 Feb 2019

1 1

 **Kushal Kumar Sinha** @KushalSinha001 · Jan 13
youtu.be/0XfUcleXfwU
Here is the old video. After the PCR van was removed, I thought that I've won the small battle against corruption. But (3)


 **Delhi Police caught taking bribe**
This is a video showing Delhi Police Con able taking bribe from poor tractor drivers. Incident happened on January 7, 2016 (time- 11:50) Near the Mayur Vihar...
youtube.com

1 2 1

 **Kushal Kumar Sinha** @KushalSinha001 [Follow](#)

Last night i again saw [@dtptraffic](#) constables taking bribe, near the same location (on the Delhi noida border). All hopes of my fight against corruption, shattered...
Unfortunately I didn't have my phone at that time, otherwise I would have recorded their faces. (4)

10:25 PM - 13 Jan 2019

 **Riaz Awan** @RiazAwan2 [Follow](#)

[@mubasherlucman](#) Aslam o Alikum I hope u r fine and have good health I have some doc of corruption in world bank loan funded project plz help me to handover said doc to IK for recovery

8:14 AM - 7 Mar 2019

1 1

Policy Bribery Example



#welcomeHomeFrank

@Asfrancisoff1

facebook.com/AmoahSamFrancis die hard @ChelseaFc fan, football analyser, presenter, journalist, Host of Ofie ni fie (Highlife Mmree on Busy fm)

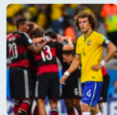
Kasoa, Ghana

chelseafc.com

Joined January 2015


Born December 26, 1987

351 Photos and videos



#welcomeHomeFrank @Asfrancisoff1 [Follow](#)

The man is annoyed the police is taking bribe him

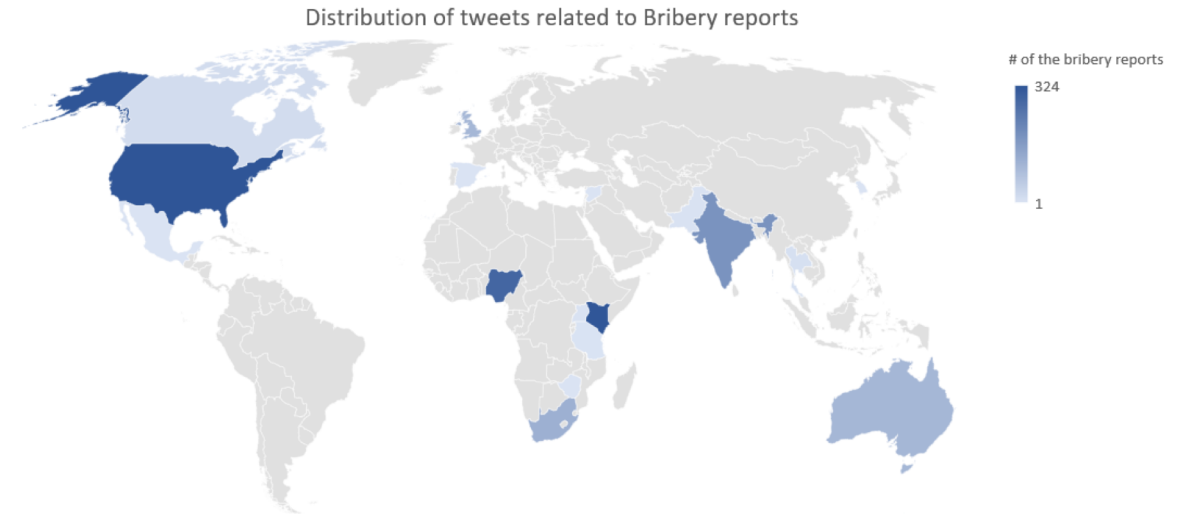
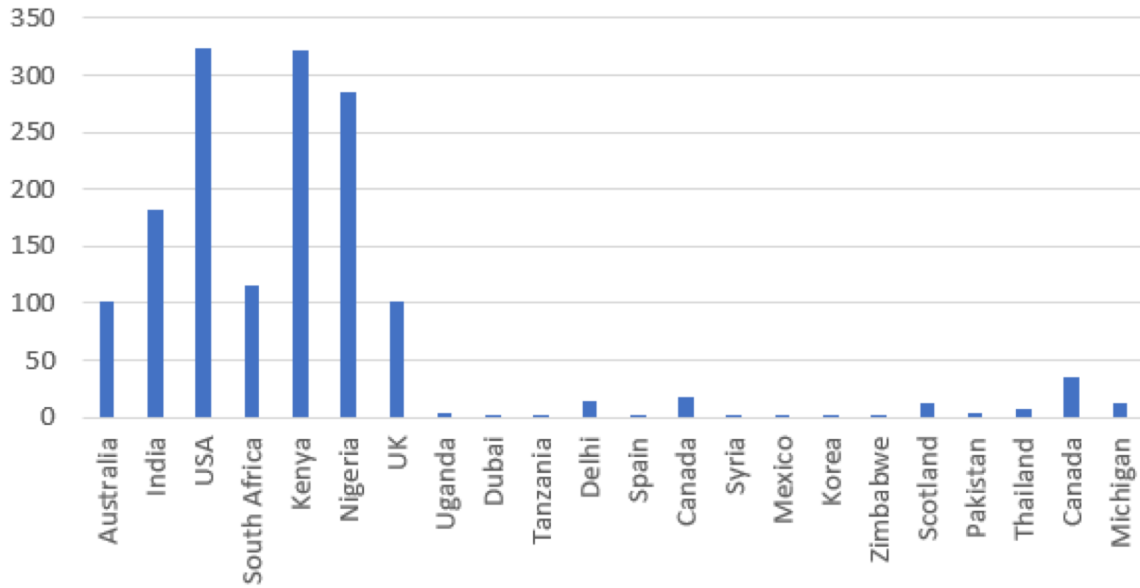


2:29 PM - 12 Feb 2019

2 Retweets 2 Likes

© 2019 Twitter [About](#) [Help Center](#) [Terms](#) [Privacy policy](#) [Cookies](#) [Ads info](#)

Geographic Distribution (bribery)



Results (Healthcare)

- BTM Topic Clusters: 20K (2K related to healthcare)
- 316 tweets from users directly reporting corruption
- Themes included: healthcare-related bribery, conflicts-of-interest (industry relationships)



betterafrika
@prosperousafrik

Follow

@MBuhari @nassnigeria @yemiosinbanjo
Psychiatrist association once attempted to sponsor a Mental Health Bill that will benefit Nigerians but this was stopped due to a demand for bribe as I was informed by a classmate who is a psychiatrist. Please help them in that regard.

8:08 AM - 26 Jan 2019



SOS Stomp Out Stupid SOS Don't Be a...
@Stop_UTK_Now

Follow

I almost died from a drug my sleep doctor (who I was also interning with) pushed on me for restless legs. She was making big bucks from the Pharma co. The drug pushing & corruption in US #medicine is truly mind-boggling. #research #BigPharma #Pharmaceutical #corrupt

...to were directed to the... Why some worry about it? Good back hold and avoided everything. They use that potential for billions of dollars and just want what it's... Push with each. Further promised a high-end promotion strategy, effectively providing kickbacks - which were legal under American law - to each part of the distribution chain. Wholesalers get volume in exchange for keeping the cost of off-gate authorization. In... Pharmacies get volume in their retail sales. Patients get expense for thirty-day starter supply.

seeshell @Tentoads4truth

Replying to @Stop_UTK_Now

8:47 PM - 1 Feb 2019

10 Retweets 20 Likes



2

10

20



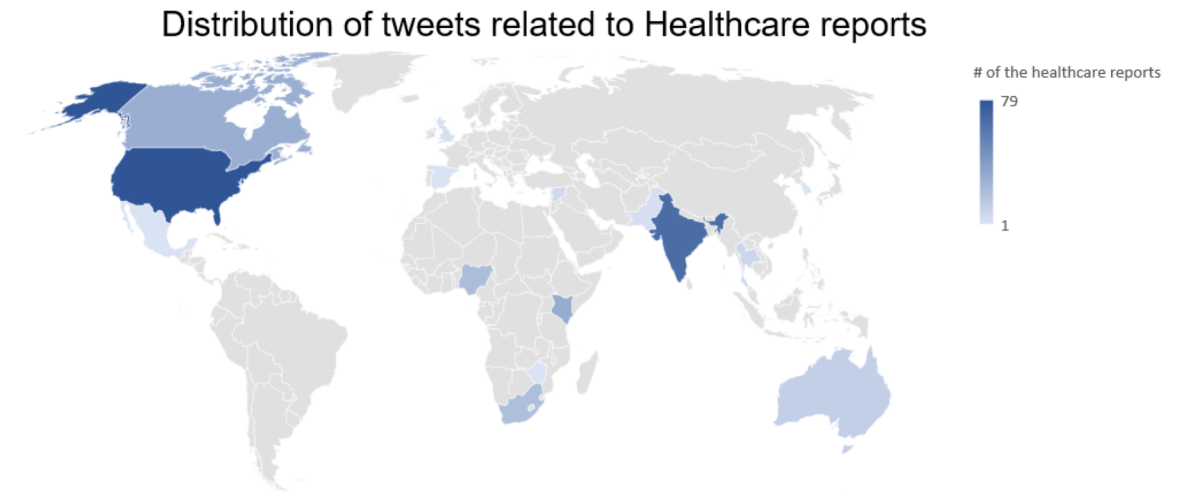
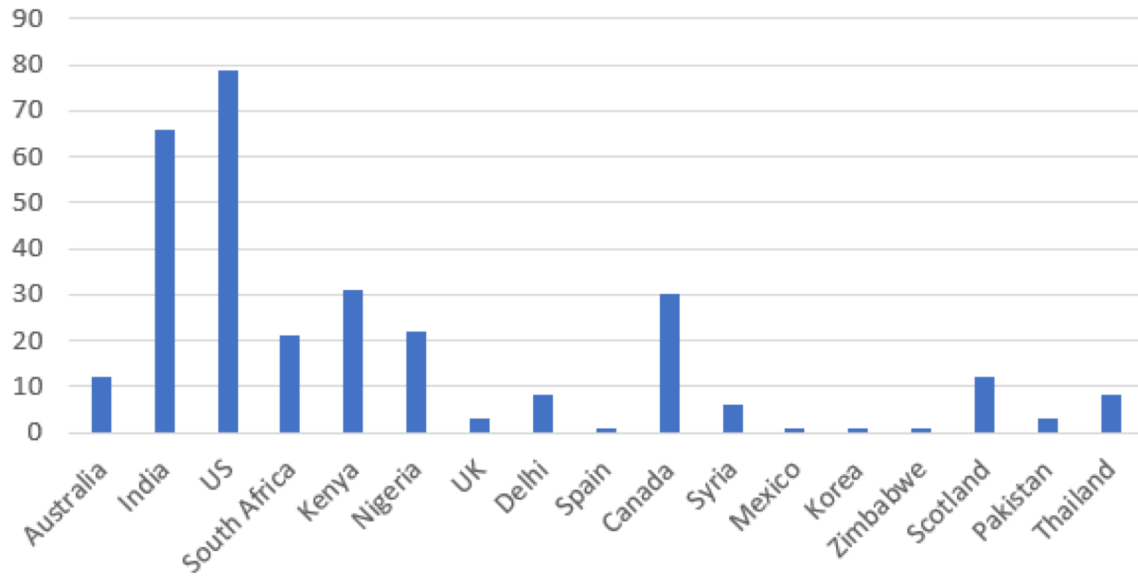
Josh M (Mister J.D.)
@bb_Joshy79

Follow

Replying to @Badflowers1Stan

Same. My doctor wanted to push medications that aren't covered by my insurance, and had bad side effects. He didn't want to give me Xanax, unless I agreed to his bribe.

Geographic Distribution (healthcare)



Study Limitations

- Limited keyword selection (i.e. no code or slang terms and language limited to English)
- Reliance on Ground Truth (Human Coding)
- Fake Tweets?

Areas of Collaboration

1. Help with keyword identification based on anti-corruption subject matter expertise and field experience
2. Better contextualization of manually annotated tweets to the real-world
3. Triangulation of self-report to actual corruption investigations or initiatives

Global Health Policy Institute

Welcome to the website for the Global Health Policy Institute (GHPi), a safe space for multidisciplinary research, education and advocacy aimed at improving domestic and global health outcomes

[Join our mailing list](#)

CONTACT INFO

Jiawei Li, MS
Email: jil206@ucsd.edu
Associate Professor,
UC San Diego - Extension
Dept of Healthcare Research & Policy

FUNDING SOURCES



UNIVERSITY OF TORONTO
LESLIE DAN FACULTY OF PHARMACY



WHO Collaborating Centre
for Governance, Accountability and Transparency
in the Pharmaceutical Sector