SURGE: Understanding and Anticipating Unrest Events

Dr. Deepti Joshi

djoshi@citadel.edu

Project Personnel

Faculty

PI, Dr. Deepti Joshi, Computer Science, The Citadel, djoshi@citadel.edu

Co-PI, Dr. Ashok Samal, Computer Science, University of Nebraska-Lincoln, samal@cse.unl.edu

Co-PI, Dr. Leen-Kiat Soh, Computer Science, University of Nebraska-Lincoln, lksoh@cse.unl.edu

Co-PI, Dr. Regina Werum, Sociology, University of Nebraska-Lincoln, rwerum2@unl.edu

Co-PI, Dr. Mike Hayes, Climatology, University of Nebraska-Lincoln, mhayes2@unl.edu

Graduate Research Assistants

Sudeep Basnet, Computer Science, University of Nebraska-Lincoln, sbasnet@cse.unl.edu

Shawn Ratcliff, Sociology, University of Nebraska-Lincoln , sratcliff@huskers.unl.edu

Daniel Schaefer, Sociology, University of Nebraska-Lincoln , dschaefer2@huskers.unl.edu

Praval Sharma, Computer Science, University of Nebraska-Lincoln , psharma4@huskers.unl.edu

Undergrad Research Assistants

Timothy Clark, Computer Science, The Citadel, tclark6@ciadel.edu

Nathanial Ballard, Computer Science, The Citadel, nballard@citadel.edu

SURGE: Social Unrest Reconnaissance Gazetteer and Explorer

- ***Long Term Goal***: develop an integrated model-driven and data-driven framework to anticipate social unrest events in a broad range of countries
- Improve situational awareness by identifying *short-term triggers* and *long-term factors* that fuel unrest at multiple geographic scales.

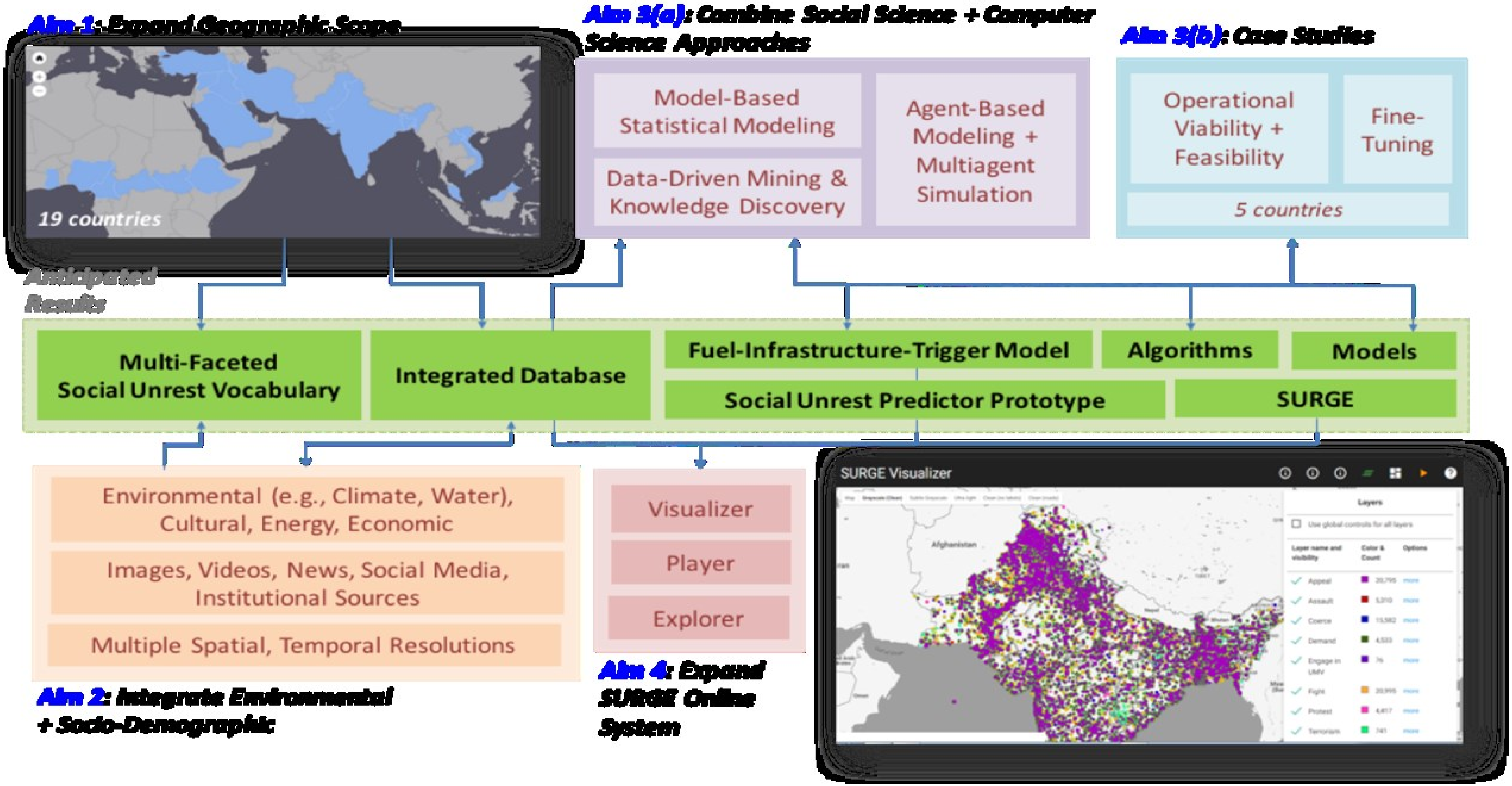


Figure: Overall framework

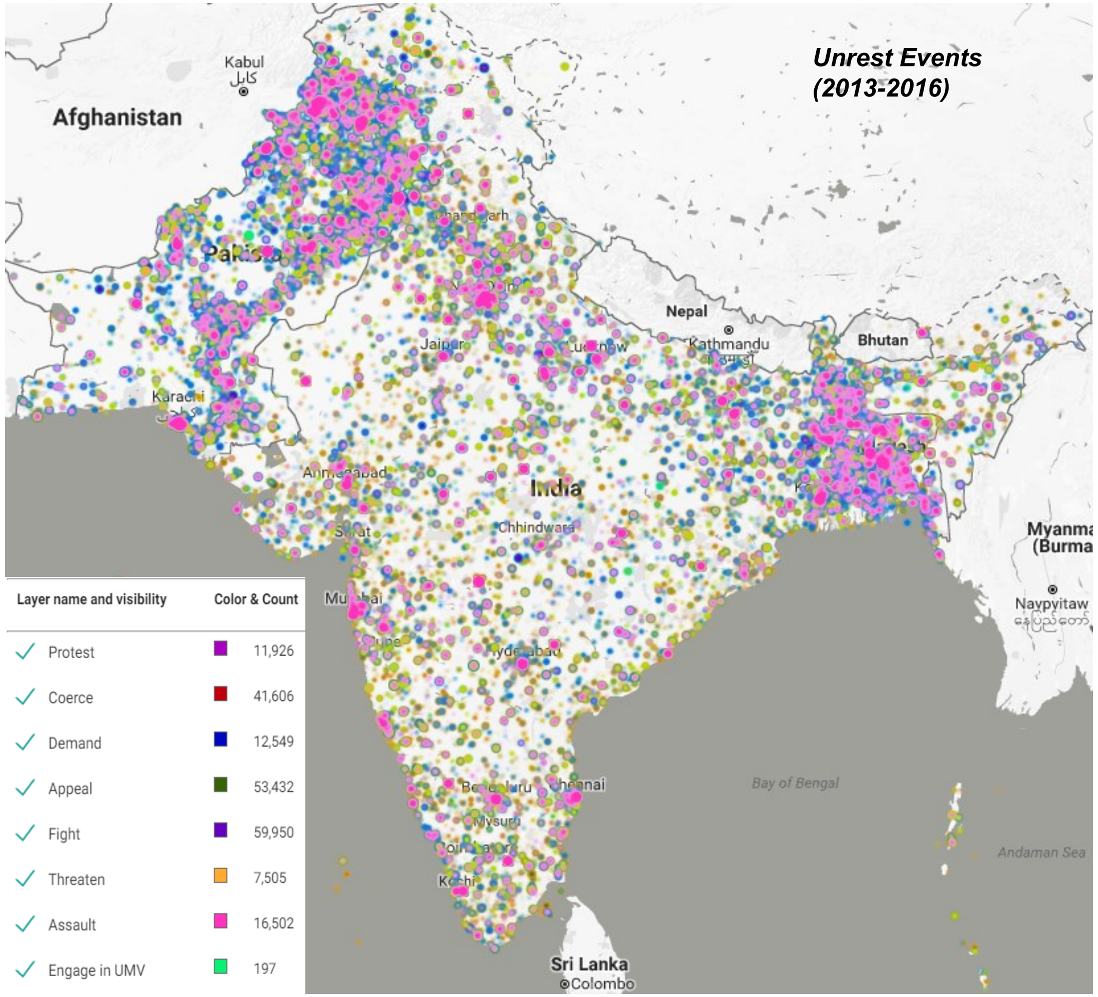
SURGE: Social Unrest Reconnaissance Gazetteer and Explorer

- ***Project Goals***:
  - examine the relationship of diverse thematic data that are increasingly becoming available in digital form including **Socio-demographic, Cultural, Environmental, Infrastructure, Geographic, Economic (SCEIGE) data**
  - Build an integrated database of unrest events and their relationships with the SCEIGE factors
  - Focus on the fuel-infrastructure-trigger model

SURGE: Visualizing Unrest Events

SURGE: Social Unrest Reconnaissance GazEtteer

- http://cse.unl.edu/~surge/
- GOAL: Understand the evolution of social unrest and develop a prediction model.
- Shows current and past locations of social unrest in India, Pakistan and Bangladesh.
- Will be extended to show locations of future social unrest.



SURGE: Data Sources (1)

- Currently the primary source for event data is the Global Database of Events, Language, and Tone (**GDELT**) - http://gdeltproject.org/.
  - GDELT offers a platform that monitors the news media from all over the world in print, broadcast, and web formats, in over 100 languages, every moment of every day.
  - It stretches back to January 1, 1979 through present day, with daily updates.
  - The raw data is in the form of a table with 58 columns containing information such as the following for each reported event:

    Actors, Targets, Types of unrest, Location of the unrest event

SURGE: Data Sources (2)

- From the raw GDELT data files, the events that occurred within India, Pakistan and Bangladesh were extracted.
- GDELT contains 20 categories of unrest events.
- We selected 8 categories (see next 2 slides) out of the 20 that were aimed at the state.
  - The categories come from the Conflict and Mediation Event Observations (CAMEO) Event and Actor Codebook (http://data.gdeltproject.org/documentation/CAMEO.Manual.1.1b3.pdf)

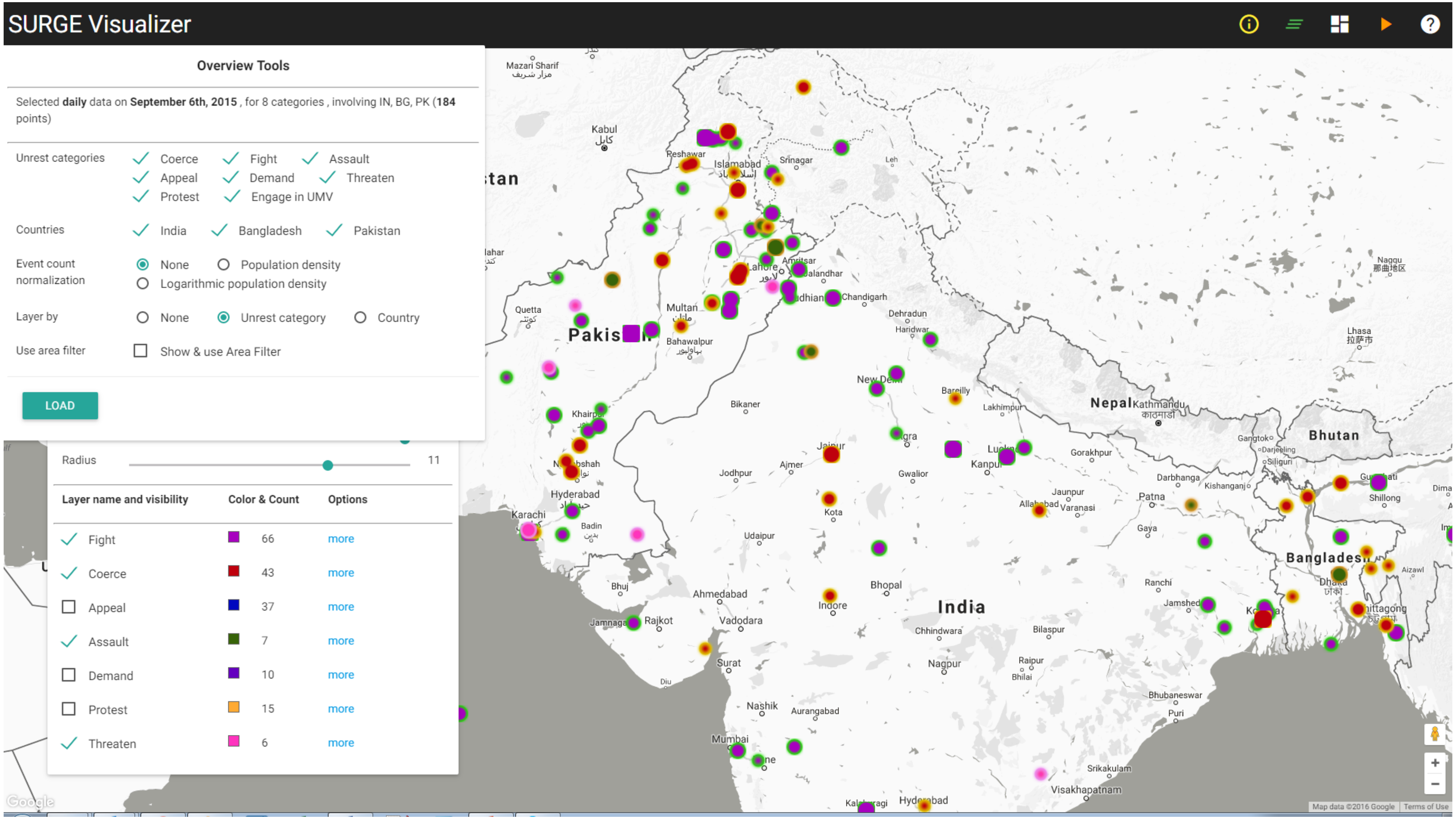
SURGE: Selected GDELT Unrest Categories (1)

Appeal	This category of unrest consists of different types of appeals that citizens can make regarding needs for certain items. This includes appealing for material cooperation, economic cooperation, military cooperation, and other types of cooperation from the state.
Demand	The public has requested a demand of the government or powers in the state. This can include the demand for economic cooperation, diplomatic cooperation, a policy change, or types of aid.
Threaten	This category is about the public threatening to boycott or even attack the state.
Coerce	These actions/events are about the destruction of items/places in order to get the outcomes that the people are interested in getting.

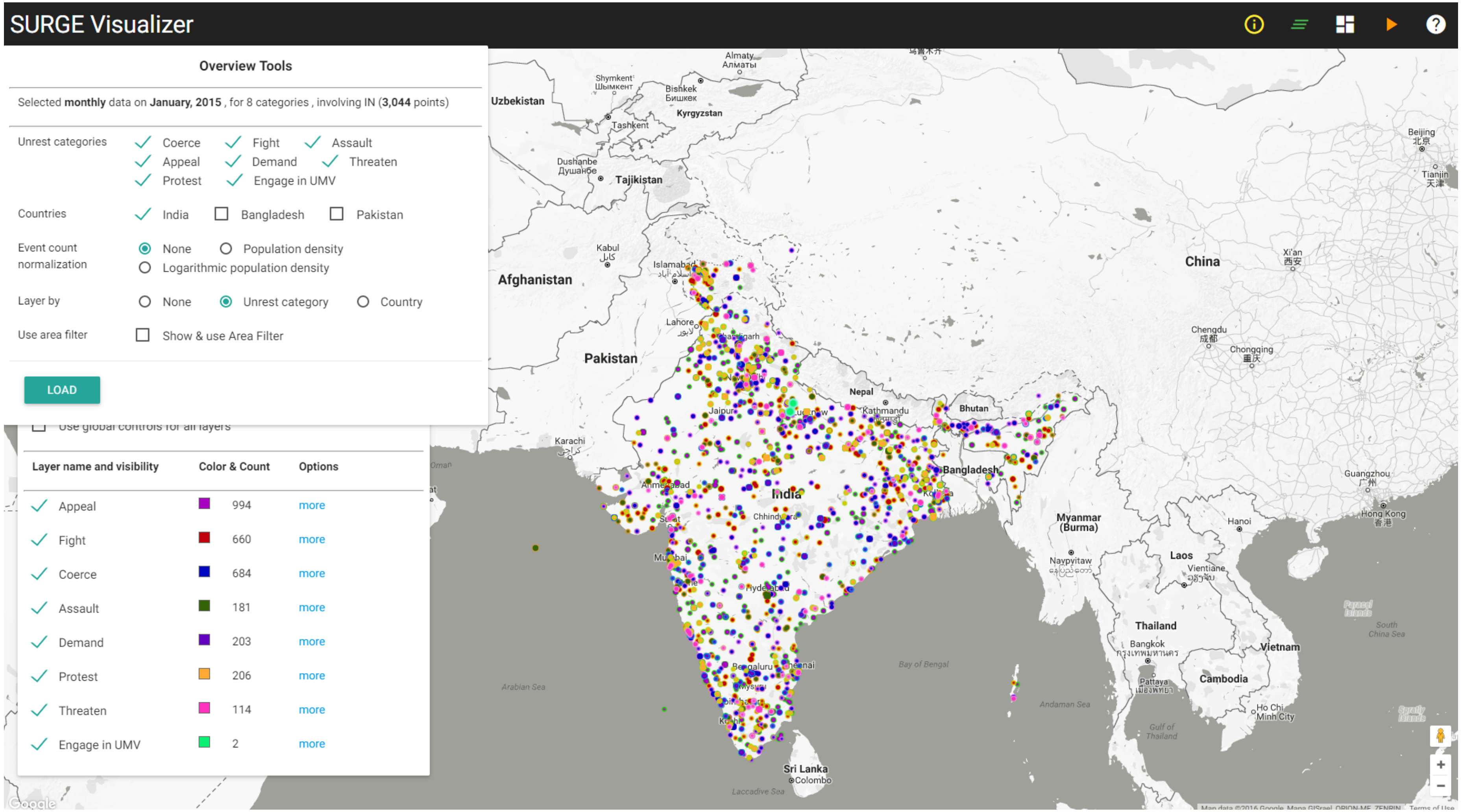
SURGE:: Selected GDELT Unrest Categories (2)

Protests	The people have engaged in some type of demonstration regarding an issue in which the public sees a problem. These demonstrations can be both violent and non-violent, but target the state/political powers.
Assault	The use of more hostile tactics, including abducting/hijacking, multiple forms of assault, bombings, and assassinations/attempts on ruling parties, by the people.
Fight	The general public has started to use non-violent tactics in order to fight back against the government. One example would be the use of small weapons or the occupation of a territory.
Engage in Unconventional Mass Violence	The country has started to experience mass killings, genocide, or other forms of mass violence.

SURGE: Daily Data Snapshot

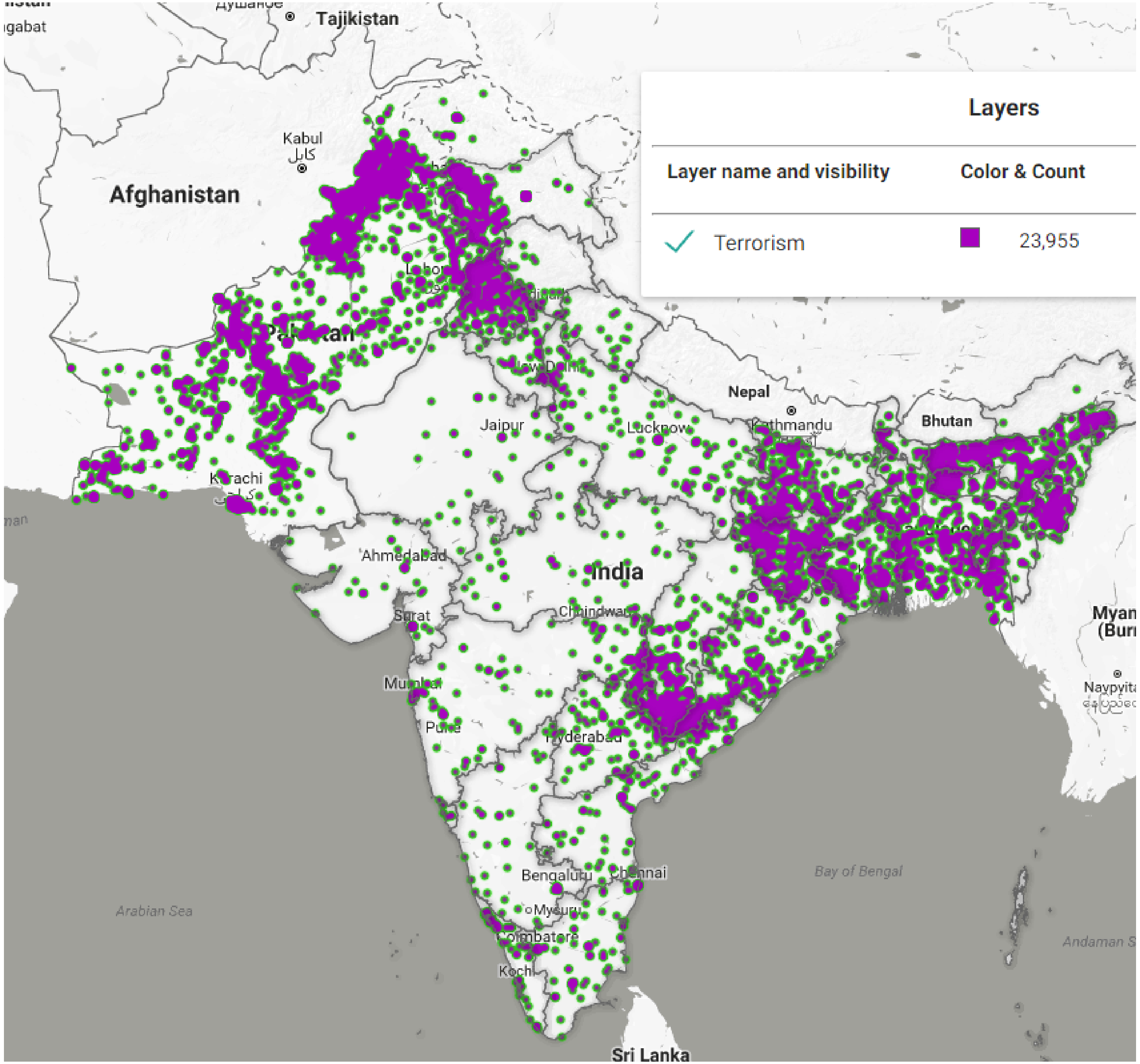


SURGE: Monthly Data Snapshot January 2015



SURGE: Additional Data Sources

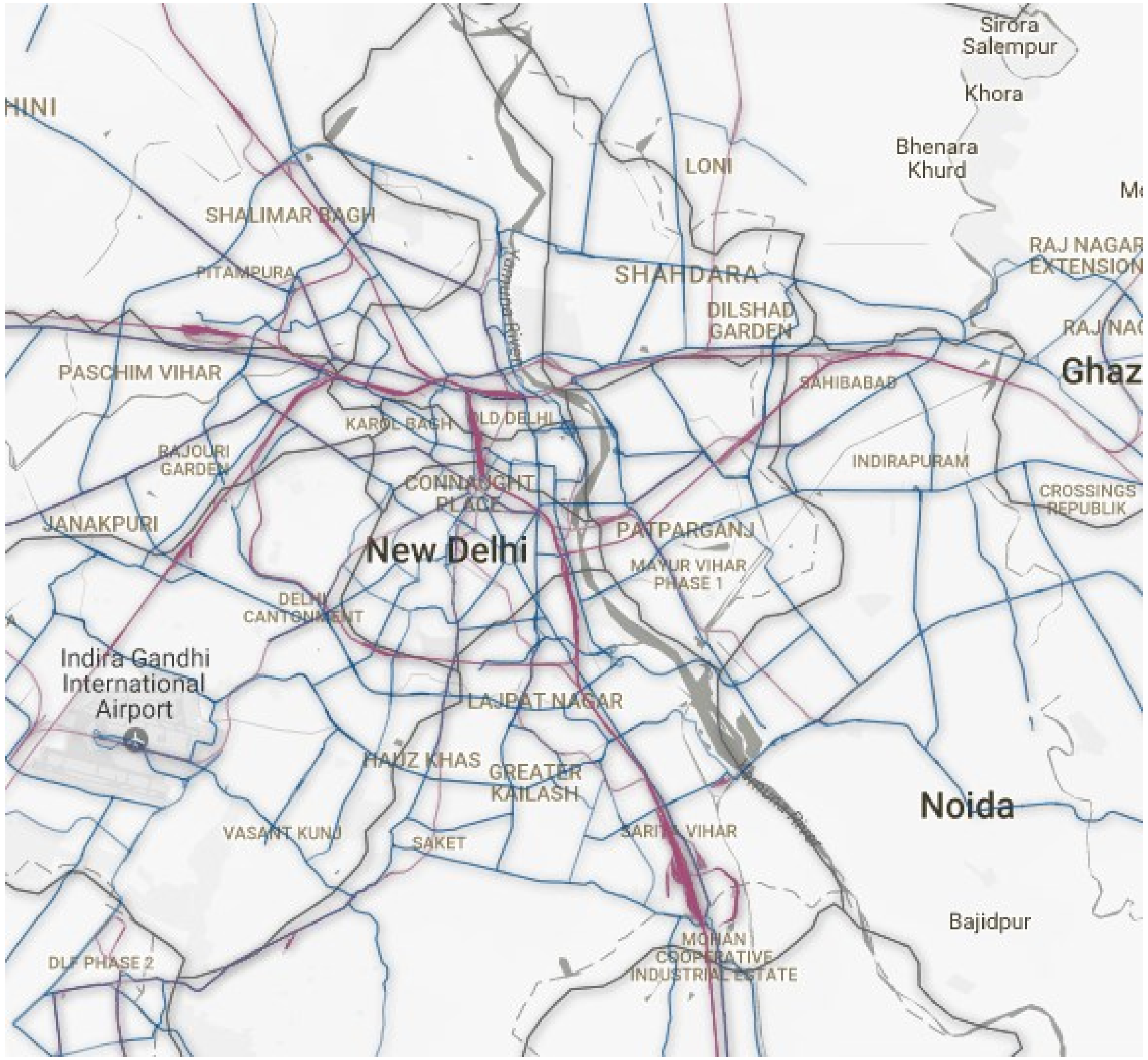
- The Global Terrorism Dataset from START, University of Maryland (http://www.start.umd.edu/gtd/) has also been added to our database to be visualized on SURGE



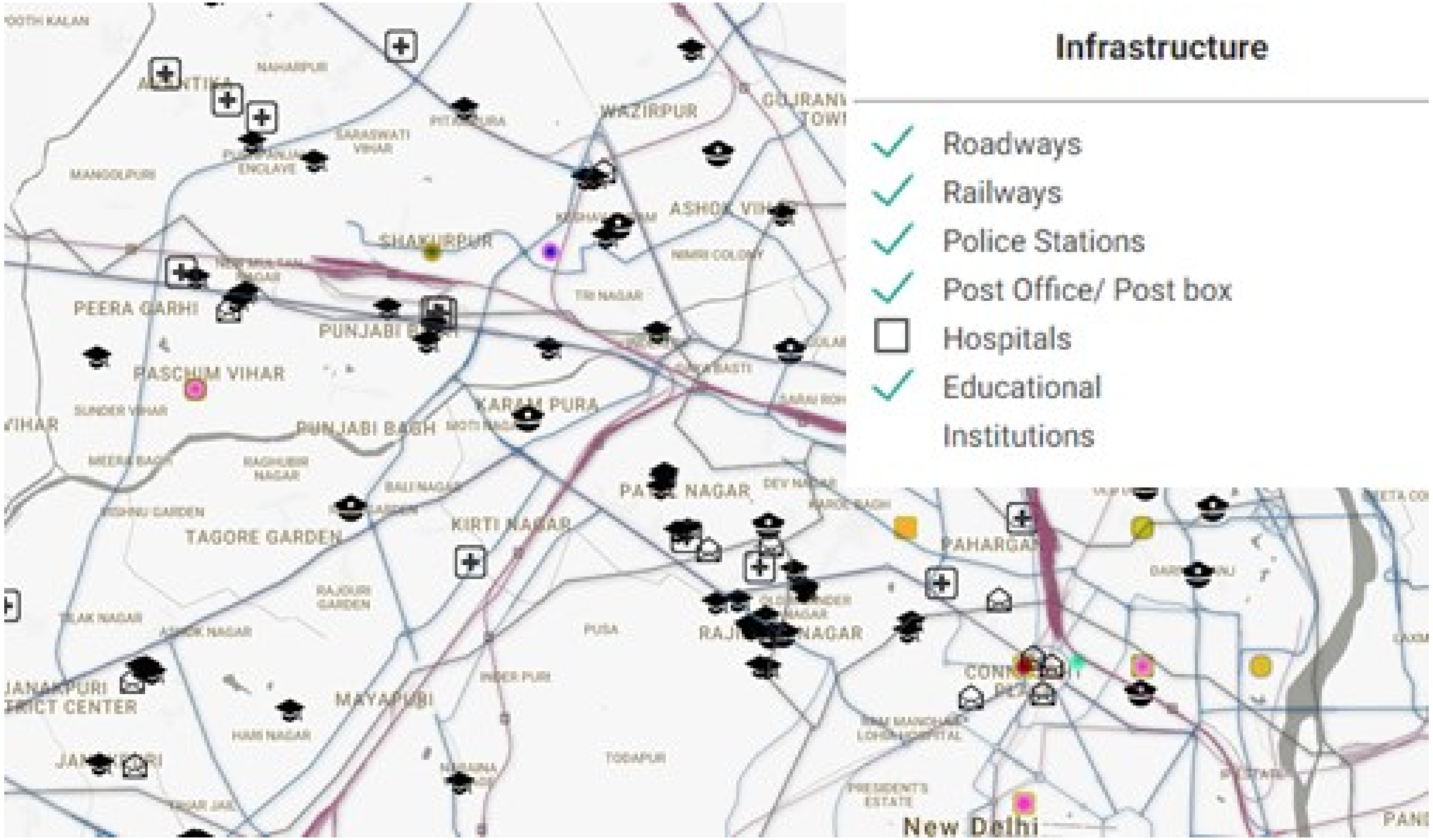
SURGE: Visualizing Infrastructure

Data obtained from OpenStreetMap (http://planet.openstreetmap.org/)

SURGE: Road and Rail Networks



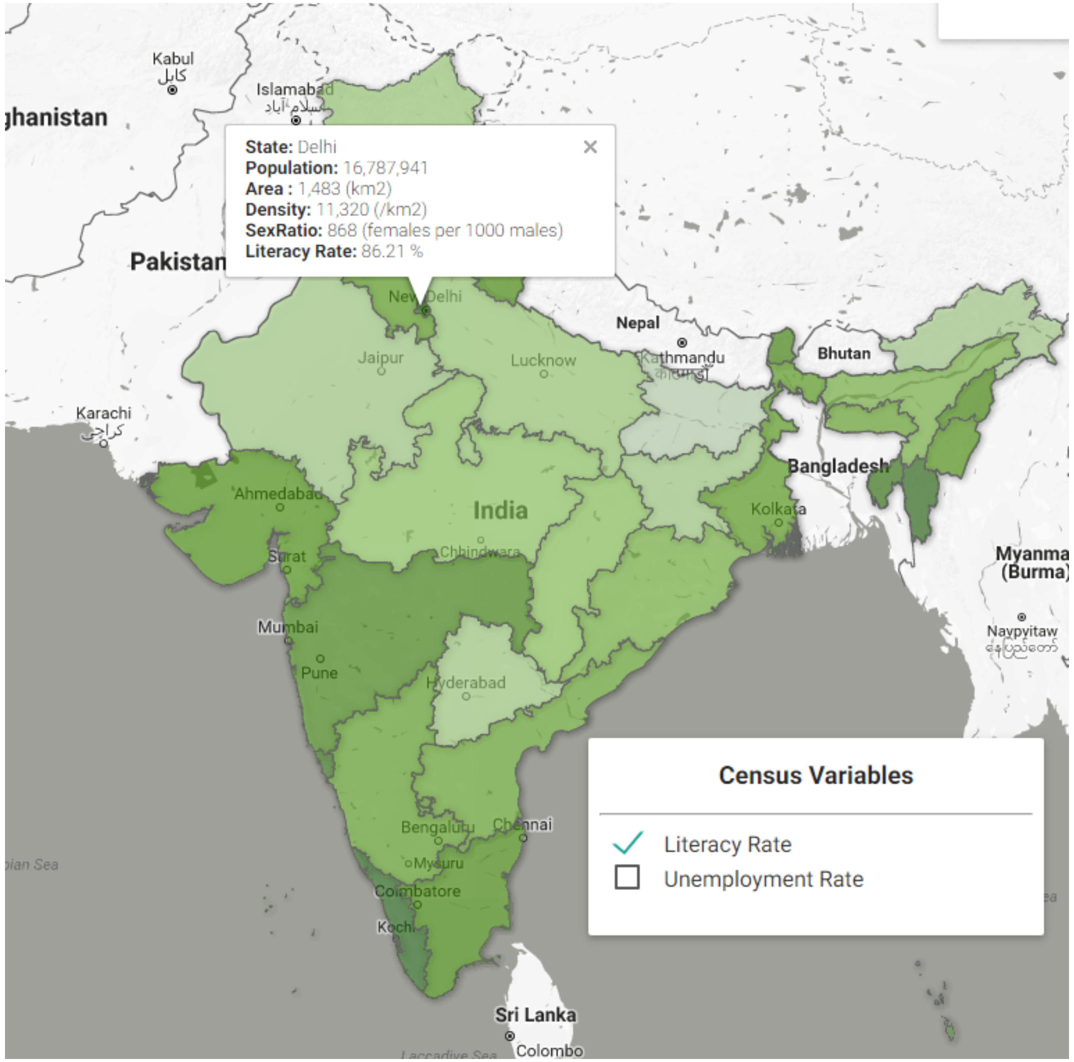
SURGE: Visualizing Police Stations, Universities, Hospitals and Clinics, and Post Offices along with unrest events

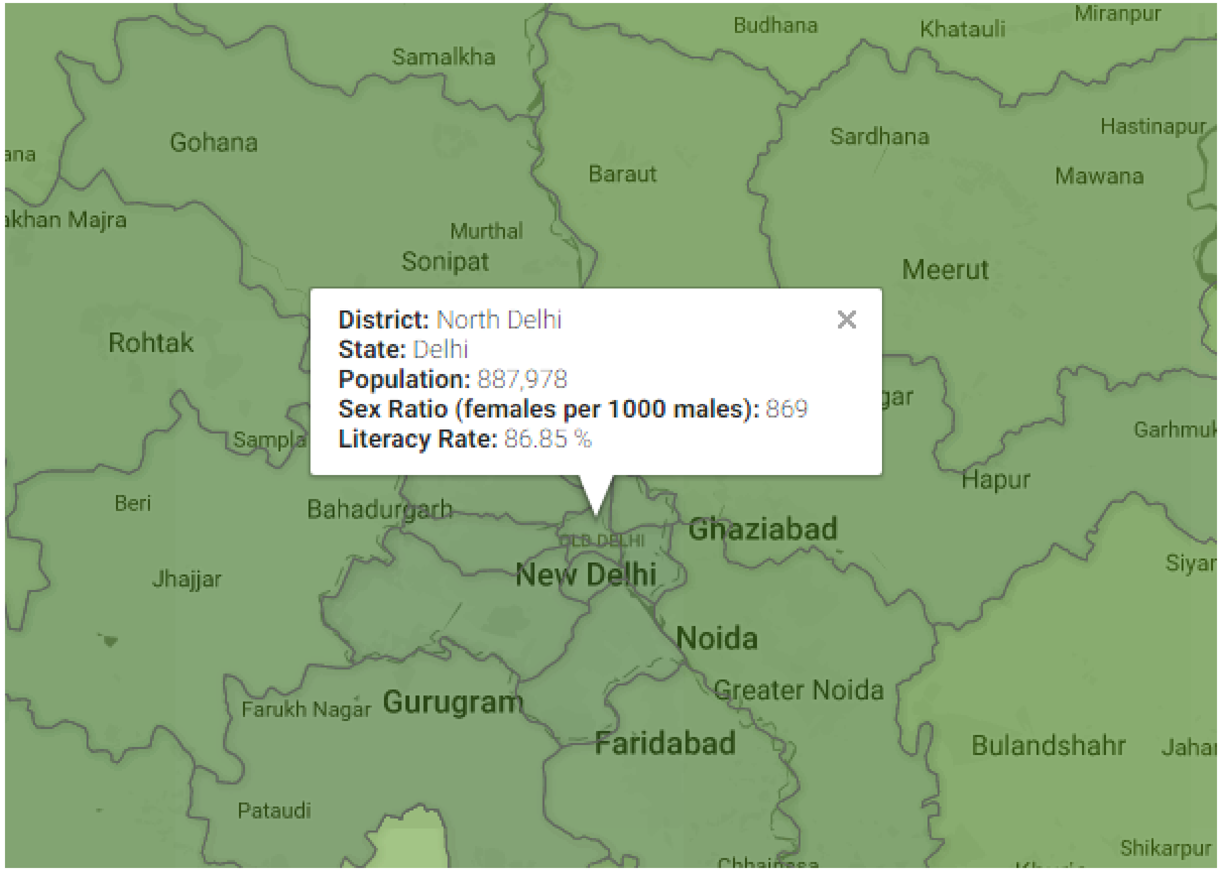


SURGE Infrastructure Data as clickable layers

SURGE: Visualizing Socio-Economic Factors

SURGE: Visualizing Literacy Rates for India at the State and District Levels





SURGE: Social Unrest Vocabulary

SURGE: Initial Social Unrest Vocabulary

Appeal	***reform*, union*, *safe*, *secur*, protect*, resist**** , appeal, cooperat*
Demand	***reform*, union*, free*, *safe*, *secur*, protect*, right*, resist****, demand, change
Threaten	***rebel*, threat****, ****safe*, *secur*, right*, resist*,*** boycott, attack
Coerce	****pressi*, enemy, hostage*, truce, threat*, boycott, attack***, destr*, forc*
Protests	***mass*,*** strik*, unrest, protest*, demonstrat*, ***rebel*,*** defen*, ***resist*,*** *violen*
Assault	***rebel*, defen*, violen*, *arm*, fight*, *terror*, extrem*, bomb*, IED, weapon*, gun*, suicid*, murder*, kill*, death*, explo*, enemy, hostage*, truce,*** assault, attack, abduct*, hijack*, hostile
Fight	***rebel*, defen*, violen*, war*, *arm*, fight*, *terror*, extrem*, bomb*, IED, weapon*, gun*,*** *fire, ***resist*, enemy, hostage*, truce,*** occup*, attack
Engage in Unconventional Mass Violence	***mass*, rebel*, defen*, violen*, war*, *arm*, fight*, *terror*, extrem*, bomb*, IED, weapon*, gun*,*** WMD, ***suicid*, murder*, kill*, death*, explo*, enemy, hostage*, truce,*** genocid*

Multilingual Social Unrest Vocabulary

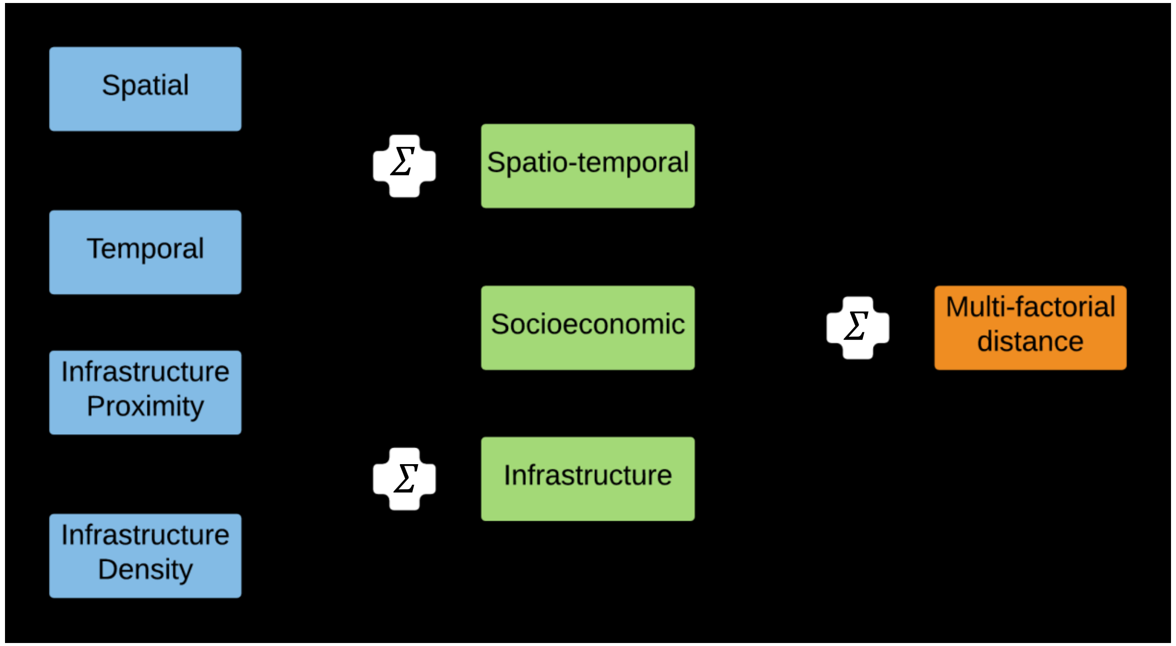
- For all the base words, translations for the words in Hindi and Bengali were written using the English alphabet.
- The process is also being implemented for Urdu words.

	A	B	C	D	E
1	**Top ~60 words:**	**English 1**	**Hindi 1**	**Hindi 2**	**Bengali 1**
2	strik*	strike	Akraman	Hartal	Hortal
3	unrest	unrest	ashanti		oshanti
4	mass* ?	masses	Jansamuh	Bahut Sara	jangan
5	protest*	protest	Virodh	Prativad	protibad
6	demonstrat*	demonstration	Pradarshan	Namuna	bikhob
7	work* ?	worker	Karamchari	Naukar	kormochari
8	(labor) *union*	union	Shramik Sangh		songho
9	compan* ?	company	Jansamuh	Mandali	songothon
10	caste	caste	Jati	Varg	jai
11	religi*	religious	Dharmik	Shraddha	dharmik
12	ethnic*	ethnic	Prajatiya	Manavjatiya	projati
13	reform*	reformed	Sudhar		sudhrono
14	rebel*	rebellion	Vidroh		bidroho
15	defen*	defense	Raksha	Suraksha	surokha
16	violen*	violence	hinsa	ugrata	hingsha
17	war*	war	yuddh	yudh	judhho
18	*arm*	armed	sena	paltan	shena
19	fight*	fight	ladai		maramari
20	(human/ labor/ civil/ religious) *right*	Right	adhikar		odhikar
21	free*	free	mukt	azaad	mukto
22		freedom	mukti	azaadi	mukti

SURGE: Data-Driven Methodologies

Past Work:: Multifactorial Distance Function

- Multi-factorial distance function combines different types of distances to give a *composite conceptual distance between each pair of events*
- We use this distance function to:
  - perform *clustering* to find patterns
  - establish event-event influence in *agent-based simulations*



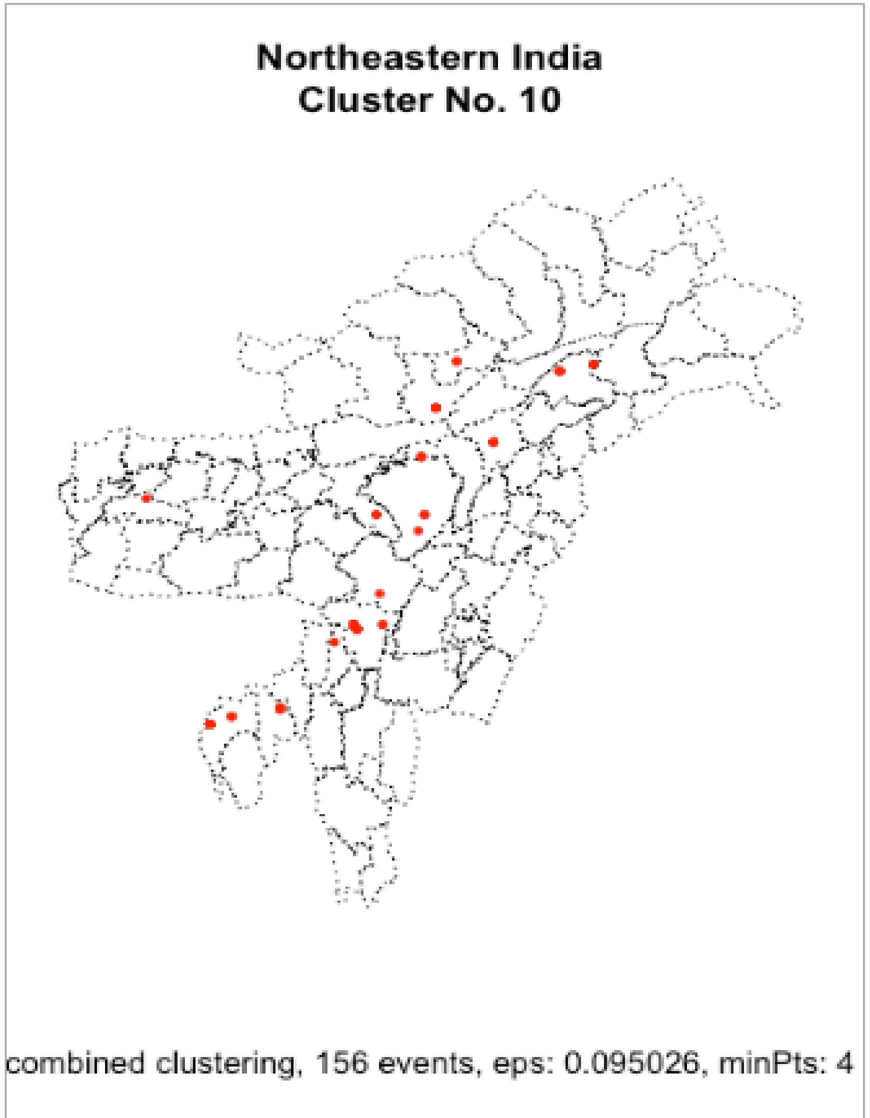
Multifactorial Distance Calculation
(weights add up to 1 at each summation)

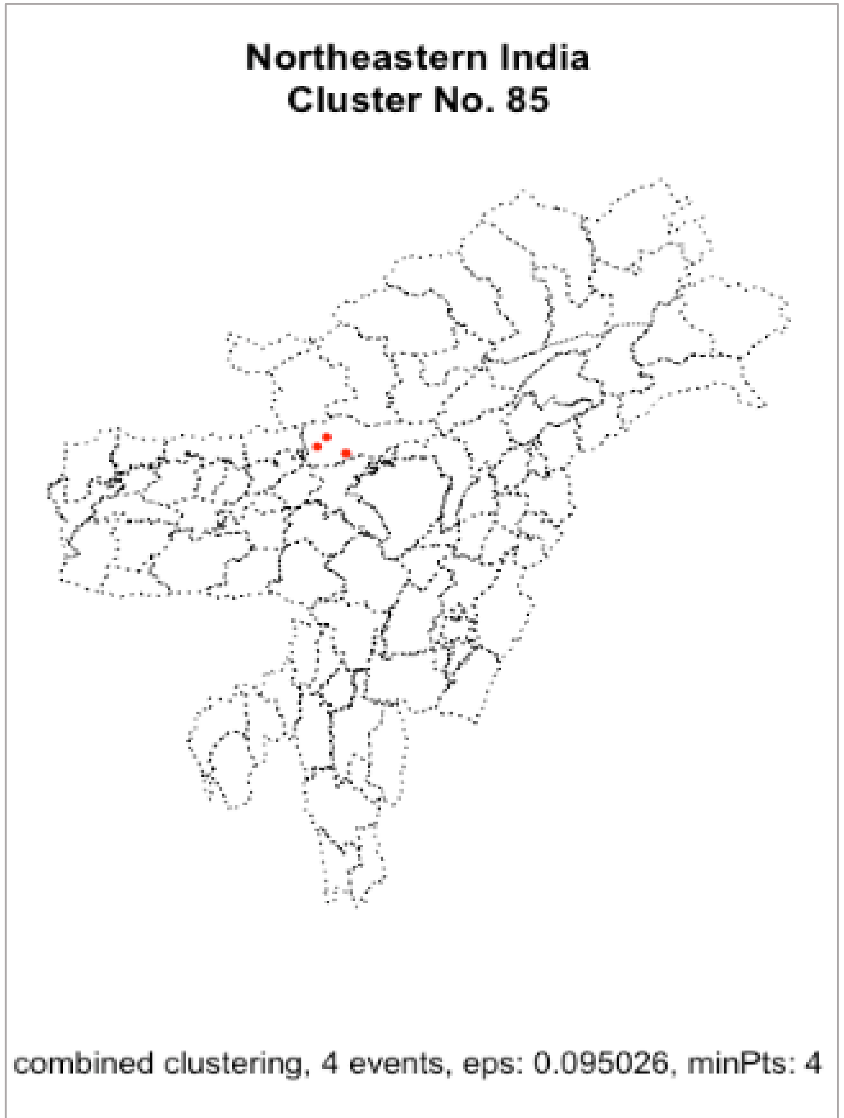
Past work:: Spatio-temporal clustering

- Data clustering allows grouping of objects based on their similarities determined by using a distance function
    - Allows data analysts to interpret groups of data and identify distinguishing patterns
- We refer to groups of related events in terms of spatial, temporal or conceptual similarities, as an *episode*
- We use the density-based clustering approach (DBSCAN) as the as it is efficient in finding clusters of similar densities in spatial database with noise
- **Goal:** discovery of episodes with their inflection points, beginning and ending points

Clustering Results

Clustering results based on combined social unrest distance function





Past work:: Agent-Based Modeling

- Due to the dynamic nature of the environment, we use an agent-based solution to model ***social unrest events as intelligent agents***
  - Complex behaviors emerge through independent actions of myopic agents without the need of a central control unit
- Each event/agent
  - has an *intensity* value
  - studies its environment and performs actions which translate into the increase or decrease of its intensity
- Loosely based on the *N-body gravitational model* where each object is trying to pull objects towards itself
- **Goal:** allow for projection of unrest events and/or what-if studies

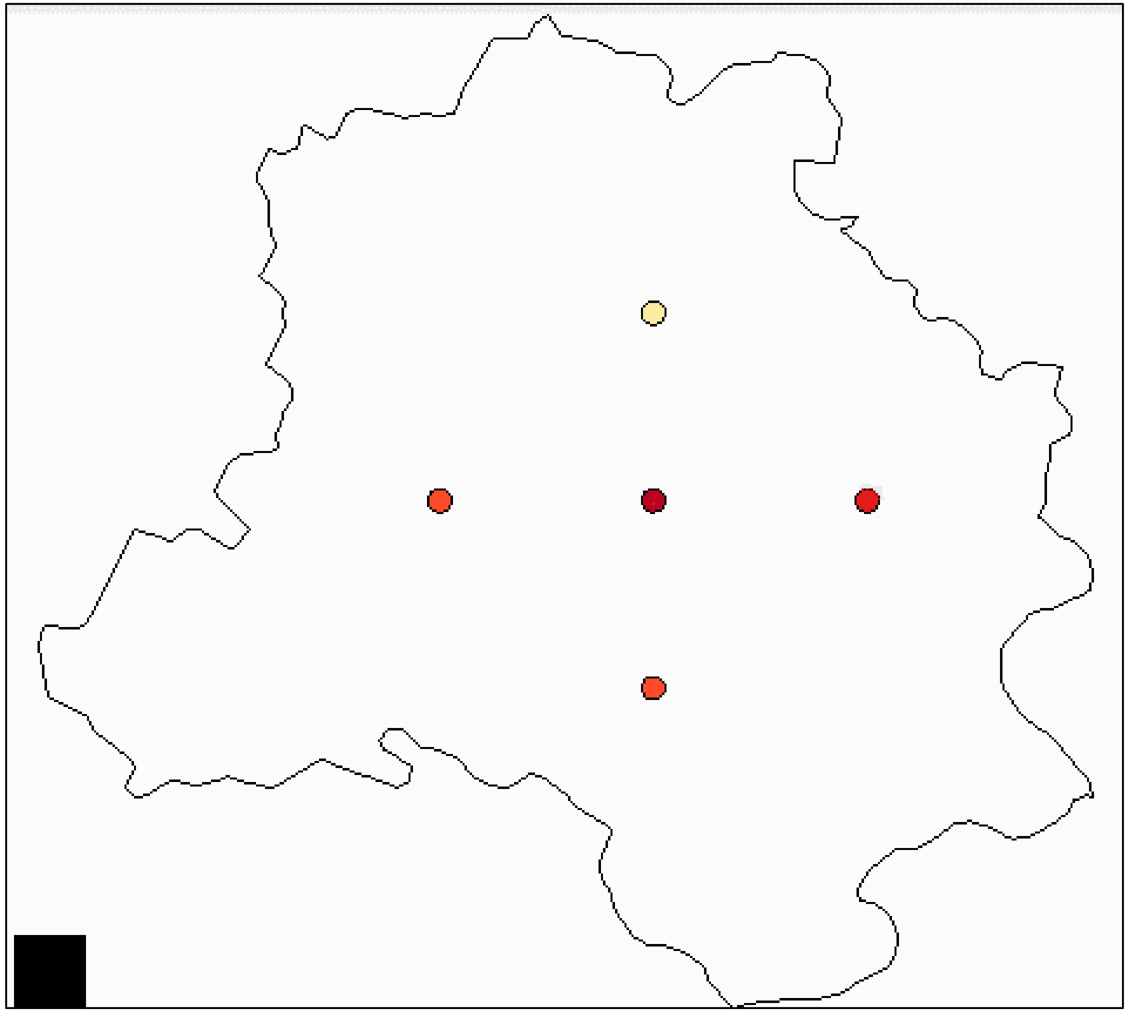
Past work:: Agent Parameters and Neighborhood

Parameters	Description
Location (l)	The geographic location where the occurrence of an unrest event has been observed. (longitude and latitude)
Event-date (t)	The date of occurrence of the event representing a specific day.
Intensity (I)	The intensity of any event e_1 at time t_1 is the energy associated with it, representing its severity.
Socioeconomic variables and Area (a)	Socioeconomic variables such as the literacy rates or employment rates are calculated for a region or area (a), we assign the socioeconomic variables of the area to all the events occurring within it.
Infrastructure variables and spatial radius (r_s)	These variables are measures of how close infrastructure objects are to an event and how many of these infrastructure objects are within a certain radius of any event.
Neighborhood (N) and Radius (R)	Any agent e_1 that is within a distance of R from another agent e_2, is considered a neighbor of the agent e_2. R can only be between [0 - 1].

Simulation of Events with no neighbors: default behavior.
Small Neighborhood Radius

Scenario 1: Events with no Neighbors

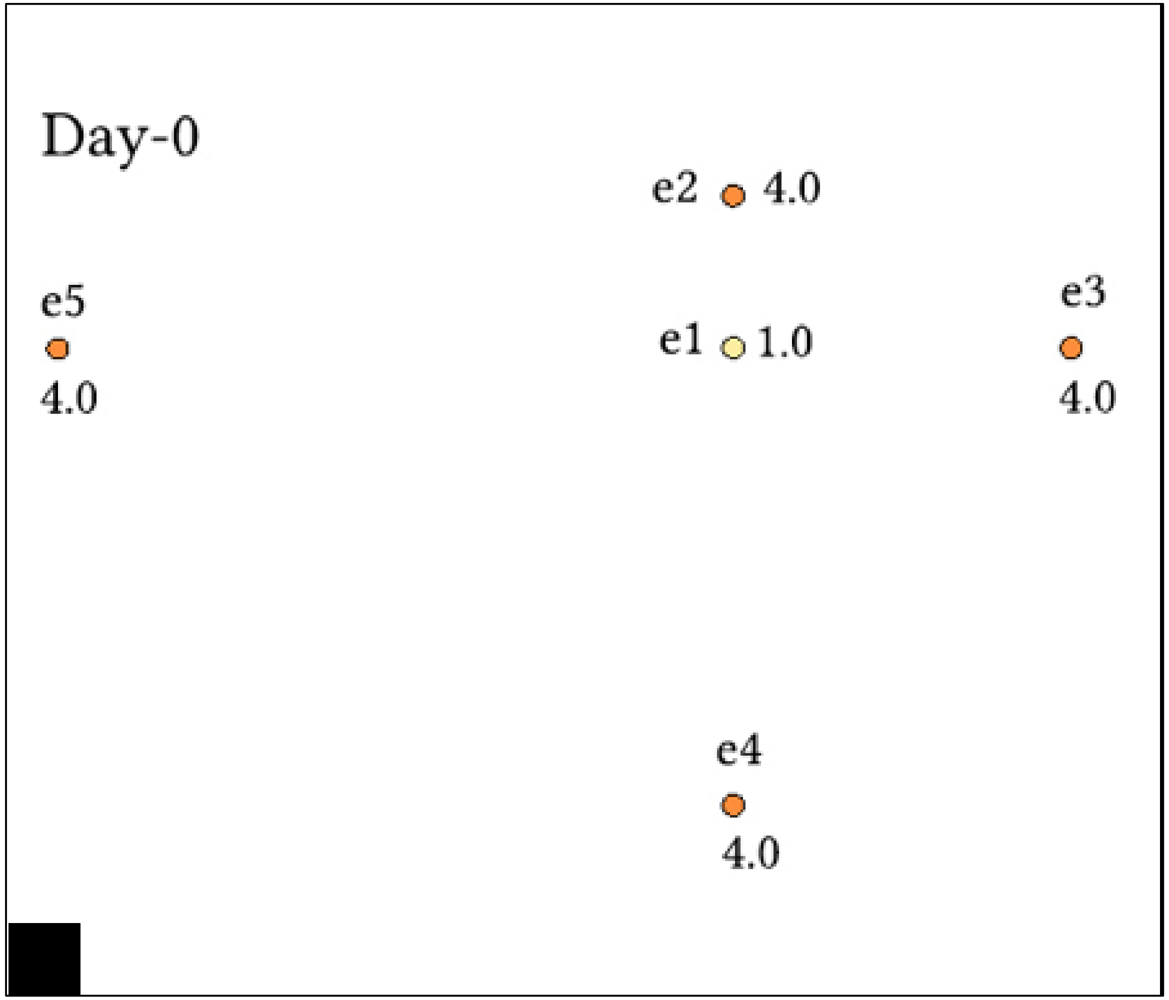
- Neighborhood Radius = 0.0
- Recovery Rate (λ) = 0.5
- Since the neighborhood radius is very small, none of the agents are able to form a neighborhood.
- The intensities can be seen to be gradually decreasing following the Decay Principle.



Simulation of Events with high Recovery Rate and sufficient Influence Rate.

Scenario 2 - Same intensity neighbors, at different distances

- Agent e_1 is the observed agent.
- Neighborhood Radius (R) = 1.0
- Recovery Rate (λ) = 0.9
- Influence Rate (γ) = 0.1



Document Analysis using the 5Ws

- **5Ws: Where, When, Who, What, Why**
- News articles and social media report unrest events
- Deeper content analysis by discovering the 5Ws will *help improve the accuracy of the distance function*:
  - Events that share the same actors (who) are closer in distance
  - Events that involve the same activities (what) are closer in distance
  - Events that are motivated by the same cause (why) are closer in distance
  - Events that occur at the same time (when) are closer in distance
  - Events that occur at the same place (where) are closer in distance

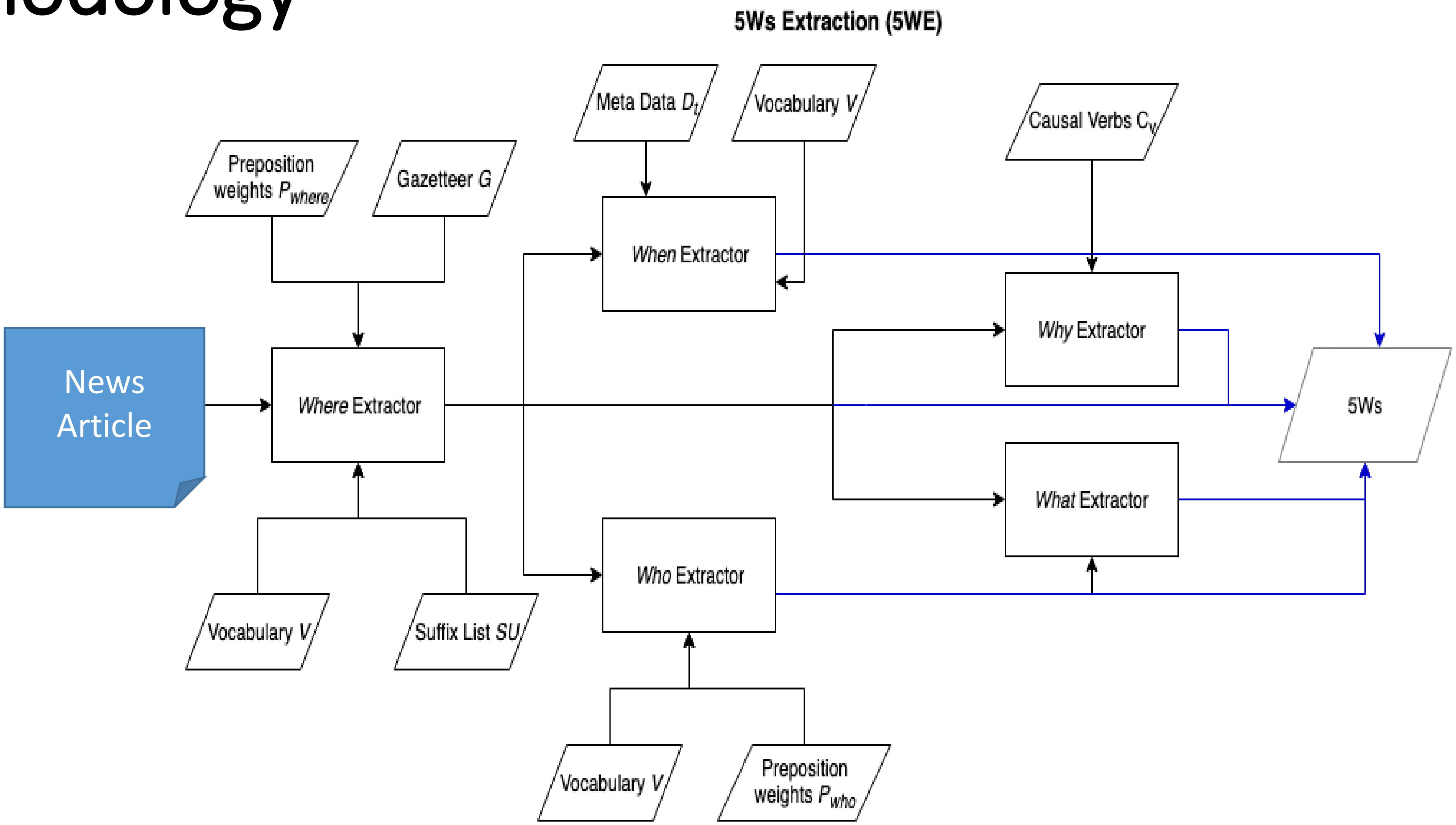
Document Analysis:: 5W Extraction (5WE)

- Each individual *W* extraction process is divided in to two tasks:
  - Candidate Identification

    Identifying all the potential candidates for a *W* using the syntactic structure of an article
  - Candidate Ranking

    Ranking the identified candidates using syntactic and semantic cues. Also utilizes domain specific keywords

Methodology



Sample 5Ws – Hand Coded – Ground Truth

Bengaluru: Hundreds march on Borewell Road against government s apathy.

Bengaluru: More than 200 residents of BBMP s Mahadevapura Zone walked peacefully this morning on Borewell Road in Whitefield, demanding restoration of the road and civic amenities. Citizens marched the 1.5 km stretch from the post office to the Ambedkar statue, while around 200 children and residents stood with placards in solidarity for the cause. Organised by Nallurahalli Rising along with Whitefield Rising, the protest in Hagadur ward drew people Kadugudi, Garadacharpalya and Hoodi wards.Led by a band of drummers and accompanied by a contingent of police, traffic police and traffic wardens, the protesters held placards, wore black and donned masks, to symbolise the pathetic condition of one of Whitefield s oldest roads. The reasons for the protest are many: poor road conditions, garbage strewn roads, fatal accidents due to water tankers that even the police cannot seem to control and all this in Nallurahalli and Whitefield itself. Residents say that six fatal road accidents have taken place over the past year. And there are many minor accidents too.The road is too narrow for the volume of traffic, there are also numerous shops that encroach footpaths and make parking difficult.Nearby roads from Ramagondanahalli and Siddhapura, the Nallurahalli New Temple Road, Outer Circle and Inner Circle are also similarly affected.Half of the streetlights, Half of the streetlights say, do not function. One of Half of the streetlights main problems is regarding the Under ground drainage work that began in August 2016, and was scheduled to be completed in October, of the same year. The BWSSB has dug up the road and left The BWSSB open, this despite, repeated requests made to the BWSSB and BBMP.Residents say that the BWSSB contractor has just poured quarry dust and pebbles wherever pits and channels have been dug, and that this is dangerous for pedestrians, cyclists and two wheelersAs a resident put it, Borewell Road (and much of Whitefield) is sinking and stinking! A portion of Outer Circle caved in when an SUV passed by on Monday. This was a chance for people s voices to be heard.

Results – 5WE

- Ground truth obtained by combining (union) the annotations from the 2 coders

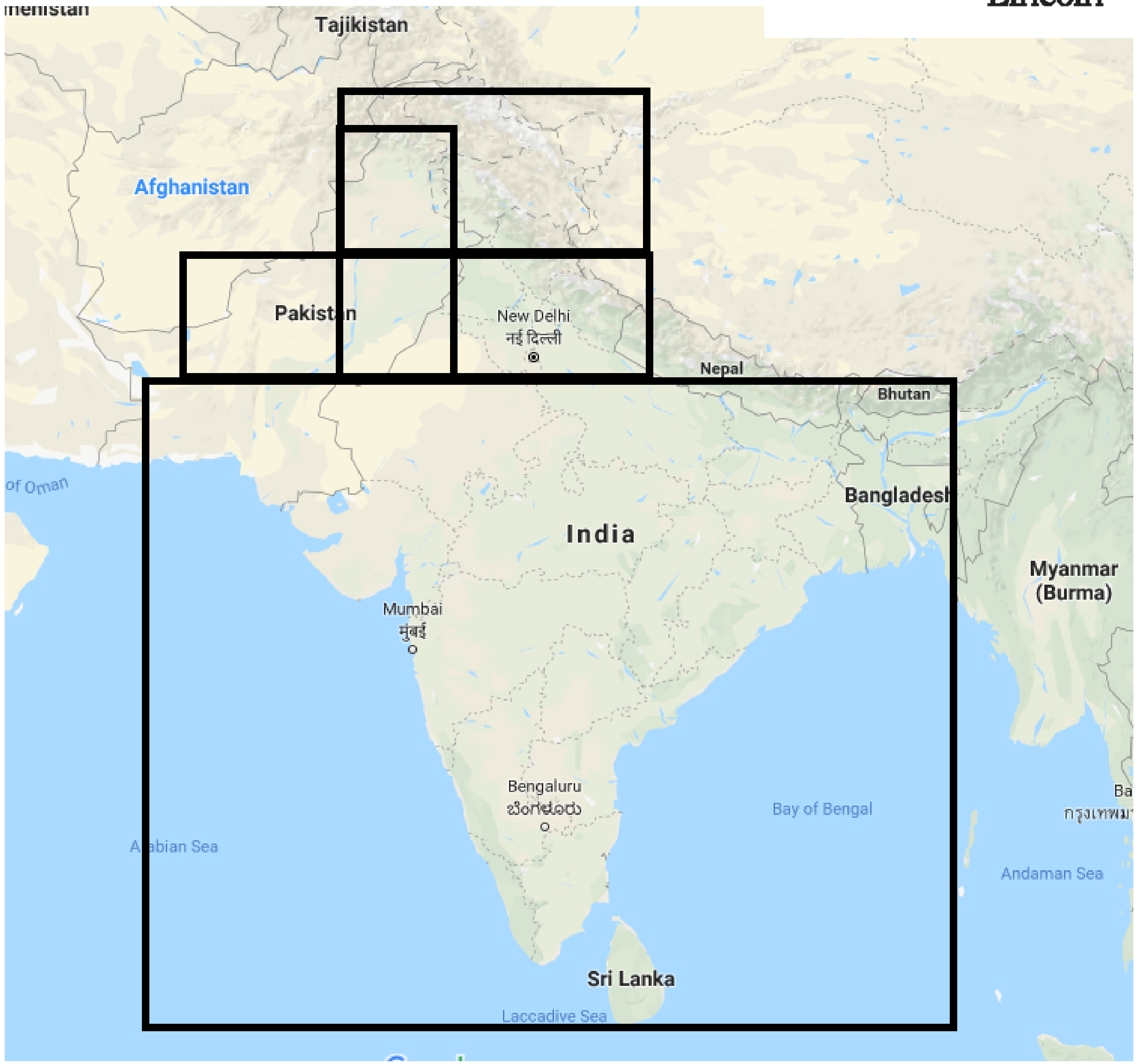
- Three evaluation strategies:
  - **Exact Match** : success if the top candidate is the ground truth
  - **Top-3 Match** : success if ground truth is in the top-3 candidates
  - **Candidate Check** : success if the ground truth is in the list of candidates

Results – 5WE

W	Exact Match (%)	Top-3 Match (%)	Candidate Check (%)
Where (139)	68.9	82.4	85.1
When (98)	71.6	91.8	91.8
Who (144)	*48.6*	74.3	*97.2*
What (84)	67.5	83.8	89.1
Why (61)	*32.4*	32.4	33.8
5Ws	**57.8**	**72.9**	**79.4**

Tweet Collection and Classification

- Tweet Collection
  - Using the Tweet2SQL package built on the Twitter Stream API
  - All the tweets collected have associated geocoordinates
  - Collected and stored publically available tweets from the identified regions of interest encompassing India, Pakistan and Bangladesh



THE CITADEL | UNIVERSITY OF Nebraska Lincoln

Sample tweets - 1

	CreatedAt	Follower	Text	Latitude	Longitude	Source	UserID	TweetID
2	1555502239	123	Commissioned Work of the Living Legend #Gulzar . Faber Castell Polychromos Pencils on Custom Sized Leather Paper. . #Fabercastell #Polychromos #Pencils #ColorPencil #Colors #Colorpencils... https://t.co/ItD8MGCloD	12.89275	77.59833	Instagram	112158012	1118483576926818305
3	1555502254	2	Stop worrying about the things that you cannot control. @ Lahore, Pakistan https://t.co/Mf4e3HD6lO	31.5497	74.3436	Instagram	1011152523112132608	1118483637018619905
4	1555502256	1497	#Welcome_Party_2k19 #Nishtarian @ MGM Loddges Multan Cantt https://t.co/lFltDwGonc	30.1818	71.42538	Instagram	743163426	1118483647844311040
5	1555502273	84030	#gorillafrommay is now trending in #Chennai https://t.co/9j1VWFuiTB https://t.co/u2i58YJmq9	13.0604	80.2496	Trendsmap Alerting	132095682	1118483720258904064
6	1555502290	115	up above the world so high! pc & ec- me. @ Amity University Kolkata https://t.co/rT6CZeKoUn	22.5956	88.488	Instagram	2181561604	1118483788256800768
7	1555502290	184	WE NEVER LOOSE INFINITY HOPE #best #bestquotes #avengers #endgame #ok #infinitygauntlet #game #alonequotes #we #quick #friends #mkdnh #trending #mardkodardnahihota #war #famous @ India https://t.co/V9yqmVanUk	21	77	Instagram	946472424450678784	1118483789498335232
8	1555502301	4729	House/Villa in Pandeypur #2Bedroom #IndependentHouse #ForRent #Pandepur #Varanasi #Residential #Property https://t.co/OmLoFnGYhH	25.35299	82.9972	PropertyWala.com	88164343	1118483834603962373
9	1555502301	4729	Luxurious #3BHK flats #ForSale in gated community #Apartment #Flat #Gachibowli #Hyderabad #Residential #Property https://t.co/k9shtPqpj6	17.43623	78.34119	PropertyWala.com	88164343	1118483834612404224
10	1555502306	4729	#Commercial #Property #ForRent #OfficeSpace #HoChiMinSarani #Kolkata https://t.co/3bfjSfHEFN	22.72075	88.33532	PropertyWala.com	88164343	1118483858423451649
11	1555502348	2	. . . . . #photo #photos #pic #pics #picture #pictures #snapshot #art #beautiful #instagood #picoftheday #photooftheday #color #all_shots #exposure #composition #focus #capture #moment @... https://t.co/jBY69VfoXO	24.86	67.01	Instagram	1053983964715716609	1118484030926802944
12	1555502358	88	#monsoonseason #lonelygirl @ Bangalore, India https://t.co/LKGNXAyEyb	12.97112	77.59765	Instagram	116482470	1118484076371951616
13	1555502381	12	River flows in you... . . . . #river #fisherman #fishermanslife #boat #fishingboats #riverlife #riverphotography #riverphoto #hills #tree #nature... https://t.co/Te4jggYmMO	20.71667	92.36667	Instagram	588547517	1118484172895473665

Sample tweets - 2

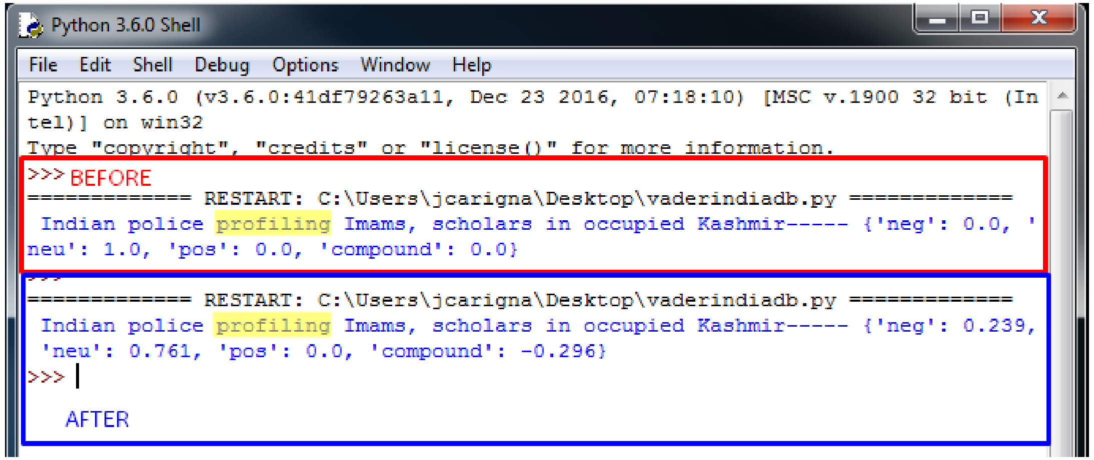
1555502457	2211	#MahavirJayanti #MahavirJayanti2019 #Jainism #Jains #LordMahavir #महावीरजयंती #जैन #महावीर #ध्यान @ Delhi, India https://t.co/p3XJ2I8fHG	28.63175	77.21967	Instagram	241474033	1118484490127642627
1555502466	4	#inspiration_for_youth #A_man_who_works_18Hours #A_man_who_think_about_only_for_his country_and_civil… https://t.co/jGtZNrClfR	21.20166	72.83196	Instagram	865762335163641856	1118484526827745281
1555502471	20	*For sale, any quantity, even 1. Can be used for gifts or pranks.* Ping for price or call 9820961376 or visit GiftWay, 7, Shiv Centre, Sector 17, Opp St Lawrence School, Next to Everest… https://t.co/9iLS7FeXmR	19.08407	72.99869	Instagram	910057666772443136	1118484549778804736
1555502478	669	Nervous to apply for a job like "Incident Manager" at Ericsson? Apply even if you're not a 100% match. You might be underestimating your value. Click the link in our bio for more info. #IT #Noida, UP	28.53552	77.39103	CareerArc 2.0	41013532	1118484576953757696
1555502482	0	क्रिकेटर युवराज ने केन्सर का ईलाज विदेश मे करवाया ।मनीषा कोईराला ने विदेश मे ईलाज करवाया, सोनिया गाधी विदेश मे ईलाज करवा रही है।स्वर्गवासी हुऐअनंत कुमारने विदेश मे ईलाज करवाया था।… https://t.co/BdXjGphsgF	25.31172	83.01212	Instagram	815530556	1118484595920343040
1555502484	79	CLEAN DO500 Dissolved Oxygen Meter Benchtop https://t.co/aojjFZEhM1 https://t.co/M7WzlNcW52	23.72202	90.41807	WordPress.com	268695248	1118484604246093825
1555502493	5	Maths Home Tutor in Delhi. Call Now: 9582317419: Maths Home Tutor in Delhi. Call Now: 9582317419 https://t.co/Cu9ezM5ToK https://t.co/xXZOZpZlnK	28.54304	77.18538	dlvr.it	993127650368540672	1118484642640740352
1555502494	4705	Mai Akela nhi hu !! Mai Mere sath hu .. Ab jaan gayi kisse muqabla hai ?? I have made some friends here also !! https://t.co/4p9kcFUh1i	21.1946	79.14258	Twitter for Android	853581736495554560	1118484644071034881
1555502510	45	Just posted a photo @ Dhaka, Bangladesh https://t.co/CfBY4GlcUv	23.7302	90.4152	Instagram	918387824910508032	1118484710986895362

Tweet Classification cont'd

- Goal: **Identify tweets related to social unrest**
- Keyword filtering insufficient
- Hand selected and labeled 110 tweets as related/unrelated to unrest
- Naïve Bayes Classifier results show 67% accuracy
- fastText Classifier results show 77% accuracy
- **Next steps:**
  - Building a larger manually labeled dataset of Tweets classified as fuel/trigger/unrest event/unrelated
  - Further testing with classification algorithms

Sentiment Analysis

- Sentiment Analysis was performed on the Twitter data collected from the South Asia region, using the VADER sentiment analysis package.
- The sentiment lexicon is being updated to include the unrest words. For example, the figure below shows the effect of the lexicon being updated.



Election Integrity Twitter Dataset

Election Integrity Twitter Dataset - 1

- Go to: https://about.twitter.com/en_us/values/elections-integrity.html#data
- Enter your email address towards the bottom of the page
- You should be now given access to data…

The follow datasets were released in October 2018

Download the corresponding Dataset Readme, and read more about these datasets on our blog.

Internet Research Agency (October 2018) - 3,613 accounts

- Account information
- Tweet information (1.2GB)
- Media (274GB, 300 archives)

Iran (October 2018) - 770 accounts

- Account information
- Tweet information (168MB)
- Media (65.7GB, 52 archives)

Election Integrity Twitter Dataset - 2

- The Dataset Readme provides a description for the data. Of particular importance are the fields included for each tweet:
  - tweetid - tweet identification number
  - userid - user identification number (anonymized for users which had fewer than 5,000 followers at the time of suspension)
  - user_display_name - the name of the user (same as userid for anonymized users)
  - user_screen_name - the Twitter handle of the user (same as userid for anonymized users)
  - user_reported_location - the user's self-reported location (*)
  - user_profile_description - the user's profile description (*)
  - user_profile_url - the user's profile URL (*)
  - follower_count - the number of accounts following the user (*)

Election Integrity Twitter Dataset - 2

- following_count - the number of accounts followed by the user (*)
- account_creation_date - date of user account creation
- account_language - the language of the account, as chosen by the user
- tweet_language - the language of the tweet
- tweet_text - the text of the tweet (mentions of anonymized accounts have been replaced with anonymized userid)
- tweet_time - the time when the tweet was published (UTC)
- tweet_client_name - the name of the client app used to publish the tweet
- in_reply_to_tweetid - the tweetid of the original tweet that this tweet is in reply to (for replies only)
- in_reply_to_userid - the userid of the original tweet that this tweet is in reply to (for replies only)
- quoted_tweet_tweetid - the tweetid of the original tweet that this tweet is quoting (for quotes only)
- is_retweet - True/False, is this tweet a retweet

Election Integrity Twitter Dataset - 2

- retweet_userid - for retweets, the userid who authored the original tweet
- retweet_tweetid - for retweets, the tweetid of the original tweet
- latitude - geo-located latitude, if available
- longitude - geo-located longitude, if available
- quote_count - the number of tweets quoting this tweet
- reply_count - the number of tweets replying to this tweet
- like_count - the number of likes that this tweet received (^)
- retweet_count - the number of retweets that this tweet received (^)
- hashtags - a list of hashtags used in this tweet
- urls - a list of urls used in this tweet
- user_mentions - a list of userids who are mentioned in this tweet (includes anonymized userids)
- poll_choices - if a tweet included a poll, this field displays the poll choices separated by |

Let's Explore
Dataset: Saudi Arabia (April 2019) - 6 Accounts

Saudi Arabia (April 2019) - 6 Accounts

- Account Information (1 KB)
- Tweet Information (38 KB)
- Media (357 MB, 1 archives)

Saudi Arabia (April 2019) - 6 Accounts

- Account Information (1 KB)
- Tweet Information (38 KB)
- Media (357 MB, 1 archives)

Dataset: **Saudi Arabia (April 2019) - 6 Accounts**

A	B	C	D
userid	**user_display_name**	**user_screen_name**	**user_reported_locatio**
4735379237	The Globus	TheGlobus	Global
zhBAcDBb6wboYvWkXJcSQ6wyhPucYvbkOGSDZkMAa	zhBAcDBb6wboYvWkXJcS	zhBAcDBb6wboYvWkXJcSQ6wyhPucY	Riyadh, Saudi Arabia
811342602607984000	Arabia	arabiadaily	
xAUWCikSC+FnQrsHVQcyV3+A3HGp8FrKvnssv3V+zA=	xAUWCikSC+FnQrsHVQcy	xAUWCikSC+FnQrsHVQcyV3+A3HGp8	The world
126601987	KSA TODAY	KSATODAY	
SvLwVcqf6ciLen09DmFjiCSFq4pkBS0JIlsy4oyIn1Q=	SvLwVcqf6ciLen09DmFjiC	SvLwVcqf6ciLen09DmFjiCSFq4pkBS0J	Global

E	F	G	H	I	J
user_profile_description	**user_profile_url**	**follower_count**	**following_count**	**account_creation_date**	**account_language**
Stay Relevant. Briefs of the wo	https://t.co/BBroMHlA3K	26766	65	1/7/2016	en
Ø¥Ø-ØµØ§Ø¡Ø§Øª | Øªة-Ù„	https://t.co/p4ji8Yp2Xk	1367	3	11/10/2015	en
news, resources, inspiration a	https://t.co/8X8mFkm9Dr	95206	0	12/20/2016	en
Your daily brief of global news	https://t.co/dtkDhd0AVM	501	48	3/19/2019	en
The latest news, exclusive sto	https://t.co/dpgUnooxul	15093	9	3/26/2010	en
The pulse of the world: politic	https://t.co/EGKbPr3wsf	393	75	3/19/2019	en

Saudi Arabia (April 2019) - 6 Accounts

- Account Information (1 KB)
- Tweet Information (38 KB)
- Media (357 MB, 1 archives)

Dataset: **Saudi Arabia (April 2019) - 6 Accounts**

1	tweetid	userid	user_disp	user_scre	user_repc	user_prof	user_prof	follower_	following	account_creatio	account_l	tweet_lan	tweet_text	tv
2	730042419291525000	4735379237	The Globu	TheGlobu	Global	Stay Relev	https://t.c	26766	65	1/7/2016	en	en	RT @moethemyth: This isn't a picture from V	
3	693046274958974000	4735379237	The Globu	TheGlobu	Global	Stay Relev	https://t.c	26766	65	1/7/2016	en	en	RT @NewsweekME: #BREAKING: Saudi autho	
4	707180239630245000	4735379237	The Globu	TheGlobu	Global	Stay Relev	https://t.c	26766	65	1/7/2016	en	en	RT @Infographic_ksa: #Asiri: #North_Thunde	
5	728316229954510000	4735379237	The Globu	TheGlobu	Global	Stay Relev	https://t.c	26766	65	1/7/2016	en	en	RT @Infographic_ksa: How does #Iran suppo	
6	746480321424658000	4735379237	The Globu	TheGlobu	Global	Stay Relev	https://t.c	26766	65	1/7/2016	en	en	RT @Infographic_ksa: Treachery is a feature	
7	689784805848961000	4735379237	The Globu	TheGlobu	Global	Stay Relev	https://t.c	26766	65	1/7/2016	en	en	RT @Infographic_ksa: Al-jubeir: we are not	
8	695659646246391000	4735379237	The Globu	TheGlobu	Global	Stay Relev	https://t.c	26766	65	1/7/2016	en	en	RT @Infographic_ksa: Asiri: #Saudiâ€™s grou	
9	720683417873039000	4735379237	The Globu	TheGlobu	Global	Stay Relev	https://t.c	26766	65	1/7/2016	en	en	RT @Infographic_ksa: With his statesmanshi	
10	695286124378456000	4735379237	The Globu	TheGlobu	Global	Stay Relev	https://t.c	26766	65	1/7/2016	en	en	RT @Infographic_ksa: #Saudis combat #terro	
11	720783537608257000	4735379237	The Globu	TheGlobu	Global	Stay Relev	https://t.c	26766	65	1/7/2016	en	en	RT @Infographic_ksa: Ignoring #Rouhani,	
12	727633149434449000	4735379237	The Globu	TheGlobu	Global	Stay Relev	https://t.c	26766	65	1/7/2016	en	en	RT @Infographic_ksa: Supervisor of the	
13	728316271272562000	4735379237	The Globu	TheGlobu	Global	Stay Relev	https://t.c	26766	65	1/7/2016	en	en	RT @Infographic_ksa: Daesh/ISIS tactics to	
14	742000197719457000	4735379237	The Globu	TheGlobu	Global	Stay Relev	https://t.c	26766	65	1/7/2016	en	en	RT @Infographic_ksa: @CIA Director:	
15	751383541074853000	4735379237	The Globu	TheGlobu	Global	Stay Relev	https://t.c	26766	65	1/7/2016	en	en	RT @Infographic_ksa: Iran harboring	
16	740302892297555000	4735379237	The Globu	TheGlobu	Global	Stay Relev	https://t.c	26766	65	1/7/2016	en	en	RT @Infographic_ksa: Arab Coalition	
17	727124978404040000	4735379237	The Globu	TheGlobu	Global	Stay Relev	https://t.c	26766	65	1/7/2016	en	en	RT @Infographic_ksa: Supervisor of the	
18	712766527087489000	4735379237	The Globu	TheGlobu	Global	Stay Relev	https://t.c	26766	65	1/7/2016	en	en	RT @Infographic_ksa: #Terrorist	
19	714467331565551000	4735379237	The Globu	TheGlobu	Global	Stay Relev	https://t.c	26766	65	1/7/2016	en	en	RT @Infographic_ksa: World media stands	
20	751383474804756000	4735379237	The Globu	TheGlobu	Global	Stay Relev	https://t.c	26766	65	1/7/2016	en	en	RT @Infographic_ksa: American Iranian	
21	867525086349275000	126601987	KSA TODA	KSATODAY		The latest	https://t.c	15093	9	3/26/2010	en	en	RT	
22	867200054293143000	126601987	KSA TODA	KSATODAY		The latest	https://t.c	15093	9	3/26/2010	en	en	RT	
23	867228350783381000	126601987	KSA TODA	KSATODAY		The latest	https://t.c	15093	9	3/26/2010	en	en	RT	
24	867228270307233000	126601987	KSA TODA	KSATODAY		The latest	https://t.c	15093	9	3/26/2010	en	en	RT	
25	869307213541761000	126601987	KSA TODA	KSATODAY		The latest	https://t.c	15093	9	3/26/2010	en	en	RT	

Saudi Arabia (April 2019) - 6 Accounts

- Account Information (1 KB)
- Tweet Information (38 KB)
- Media (357 MB, 1 archives)

Dataset: **Saudi Arabia (April 2019) - 6 Accounts**

UserIDs

Name	Date modified	Type	Size
126601987	10/7/2019 10:33 AM	File folder	
4735379237	10/7/2019 10:34 AM	File folder	
811342602607984641	10/7/2019 10:34 AM	File folder	
zhBAcDBb6wboYvWkXJcSQ6wyhPucYvbkOGSDZkMA...	10/7/2019 10:36 AM	File folder	
126601987	9/6/2019 10:23 PM	zip Archive	8,819 KB
4735379237	9/6/2019 10:23 PM	zip Archive	167,576 KB
811342602607984641	9/6/2019 10:23 PM	zip Archive	189,669 KB
zhBAcDBb6wboYvWkXJcSQ6wyhPucYvbkOGSDZkMA...	9/6/2019 10:23 PM	zip Archive	177 KB

Saudi Arabia (April 2019) - 6 Accounts

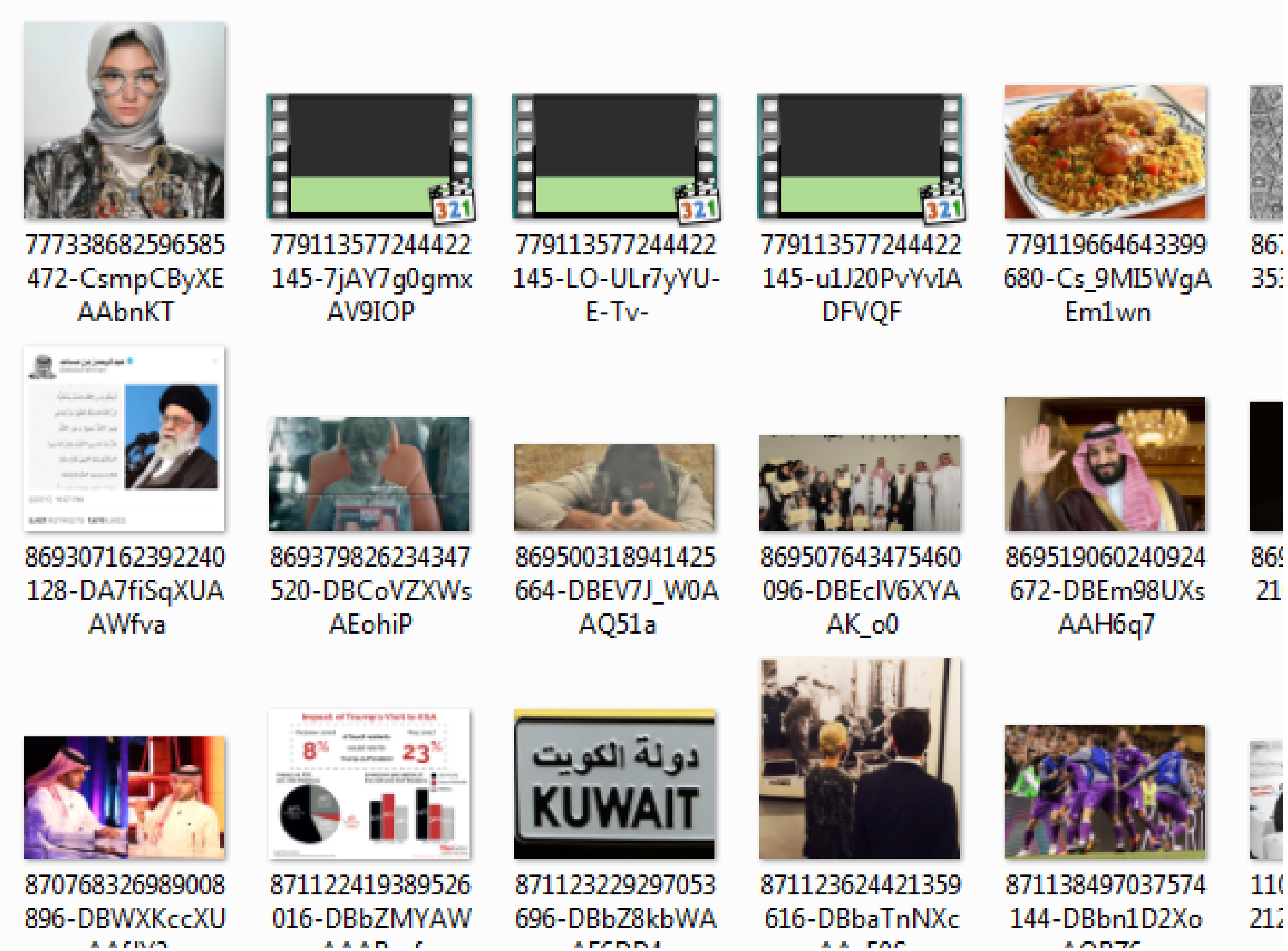
- Account Information (1 KB)
- Tweet Information (38 KB)
- Media (357 MB, 1 archives)

Dataset: **Saudi Arabia (April 2019) - 6 Accounts**

Name

126601987

TweetIDs



Analysis with tweets: Different pathways

- Create word clouds and word frequency lists
- Frequency analyses and charts (see figure for examples)
- Discover events – place and time references
- Cluster Tweets
- Trend analysis to discover changes over time
- Sentiment analysis
- Topic detection

Chart	Bars represent	Categories
Weekday	Number of tweets	Days
Time	Number of tweets	Hours
Type	Number of tweets	Tweet, retweet, reply
Most frequent words	Number of tweets	15 most frequently used words (options for making this anaylsis case-sensitive and/or taking word cloud stop-lists into account). The category “Other” can also be displayed.
Most frequent hashtags	Number of tweets	15 most used hash tags. The category “Other” can also be displayed.
Author by number of tweets (real name) / Author bz number of tweets (Twitter name)	Number of tweets	15 authors with the highest amount of tweets. The category “Other” can also be displayed.
Authors followers	Number of followers	15 authors with the most followers. The category “Other” can also be displayed.
Source	Number of tweets	15 most common sources. The category “Other” can also be displayed.
Retweets	Number of retweets	Specified categories from 0 to a 100+
Likes	Number of likes.	Specified categories from 0 to a 1,000+

Thanks!

djoshi@citadel.edu