



Event Detection

Tarek Abdelzaher

University of Illinois at Urbana Champaign





Why You Liked It?

- Well thought-out/organized/clear content: 28
- Good visuals/diagrams: 22
- Interesting topic/idea: 17
- Good description of contribution/challenges (compared to prior art): 11
- Fluent/energetic/engaging speaker: 8
- Good interaction/Q&A with the audience: 4
- Good humor (made me laugh): 1

Research Goal



Source: <http://www.signagesolutionsmag.com/article/unleashing-new-media-the-rise-of-the-social-media-command-center-15501>



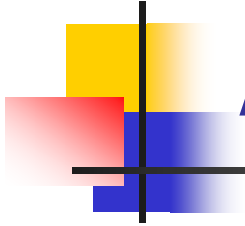
Research Goal

Empower the analyst to quickly detect, localize, and track ongoing events (normal and anomalous) by exploiting social media.



Service Design Philosophy

A Human-Machine Collaboration



Human

Good at natural
language processing



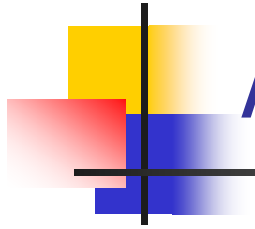
Time

Machine

Good at statistical
analysis and alerts

Service Design Philosophy

A Human-Machine Collaboration



Set tripwire:
Indicate interest area
(e.g., disasters, protests,
traffic, politics, etc.)



Human

Good at natural
language processing



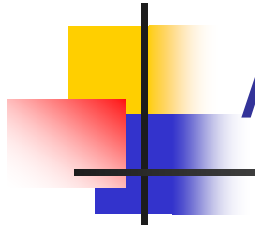
Time

Machine

Good at statistical
analysis and alerts

Service Design Philosophy

A Human-Machine Collaboration



Set tripwire:
Indicate interest area
(e.g., disasters, protests,
traffic, politics, etc.)

Human

Good at natural
language processing



Time

Alert when wire is tripped:
Report unusual events
matching interest area

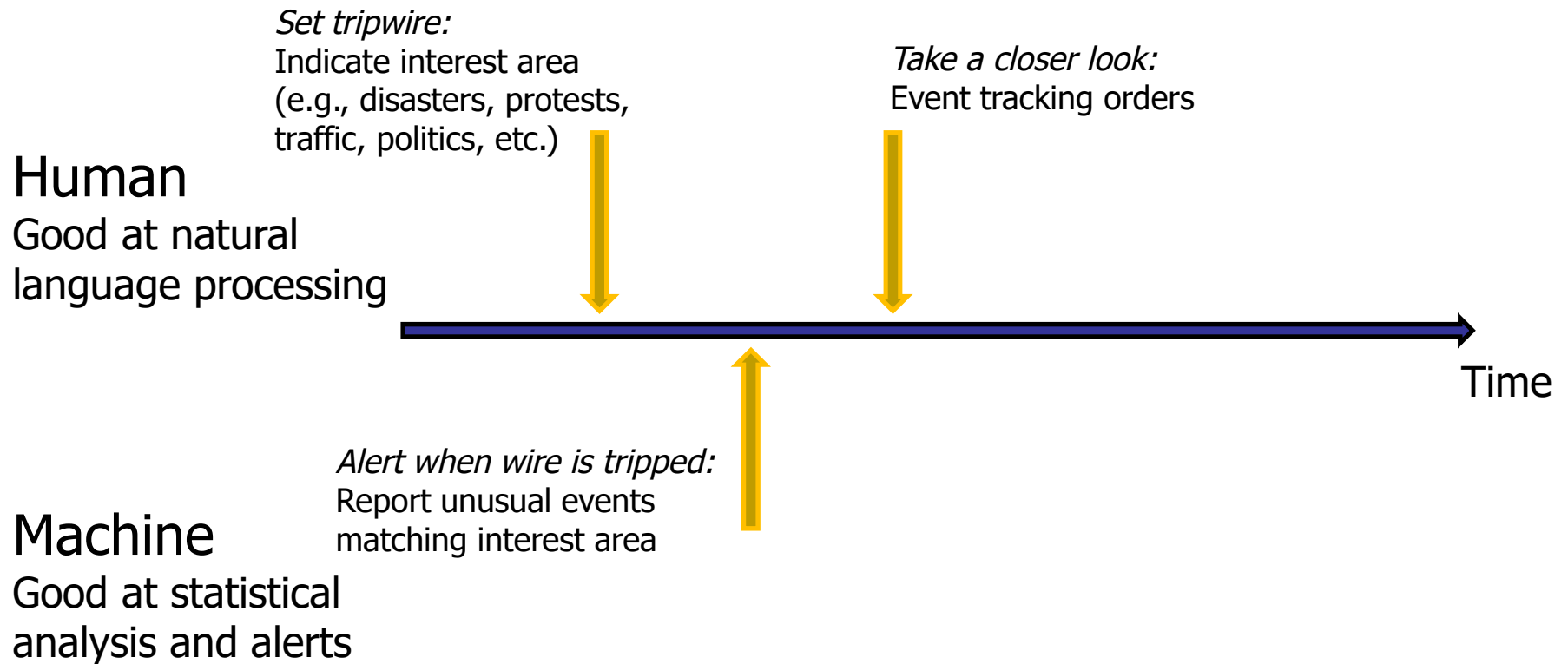
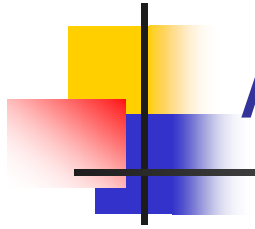
Machine

Good at statistical
analysis and alerts



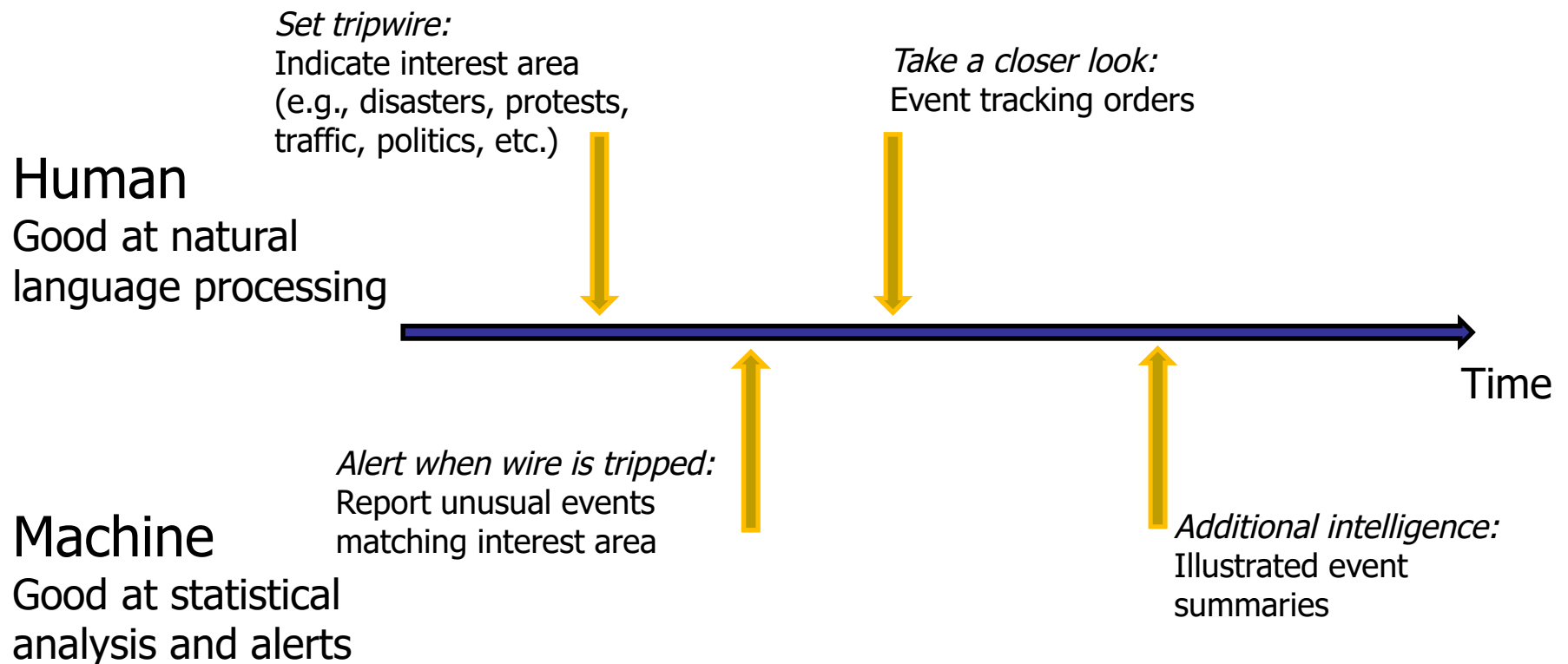
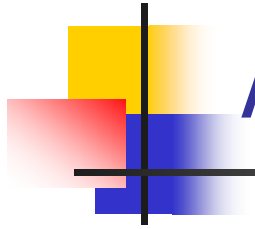
Service Design Philosophy

A Human-Machine Collaboration



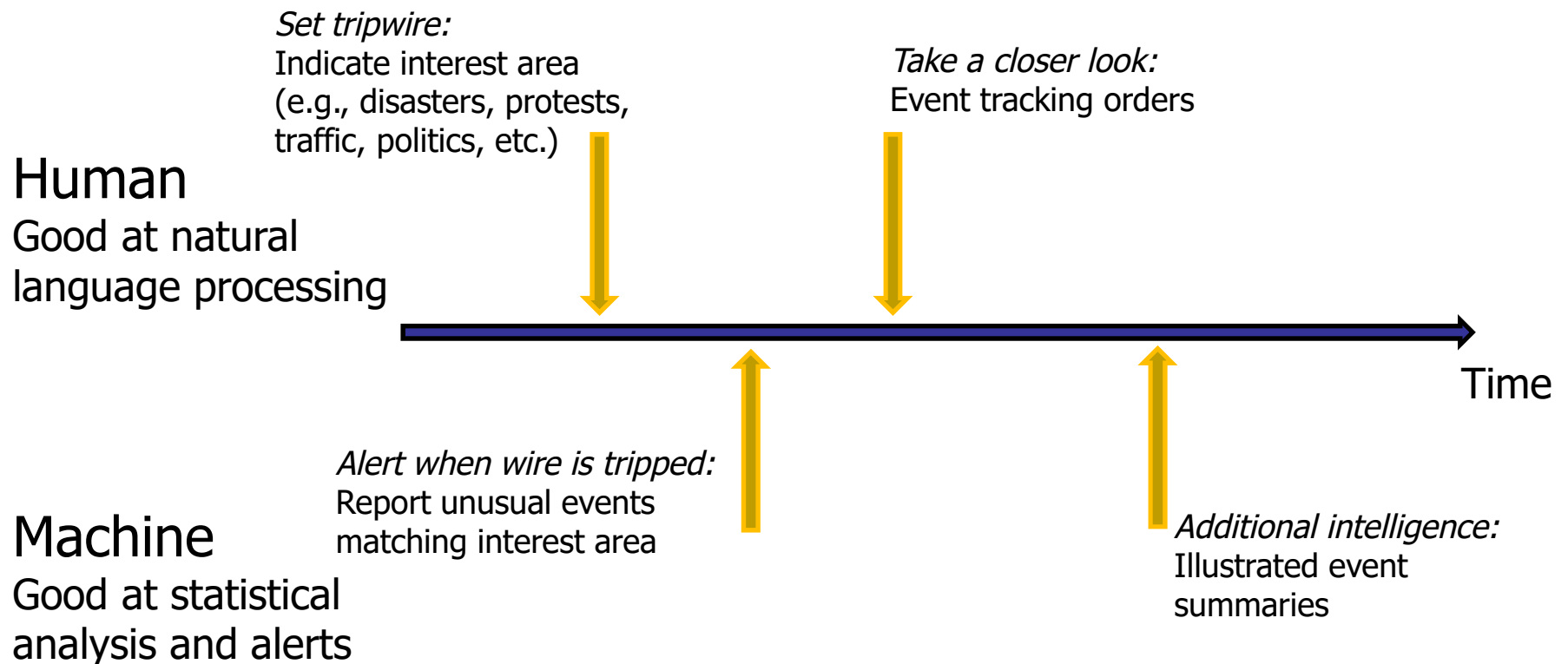
Service Design Philosophy

A Human-Machine Collaboration



Service Design Philosophy

A Human-Machine Collaboration



Design decision: Machine detects and tracks events and patterns. The human interprets the specifics.

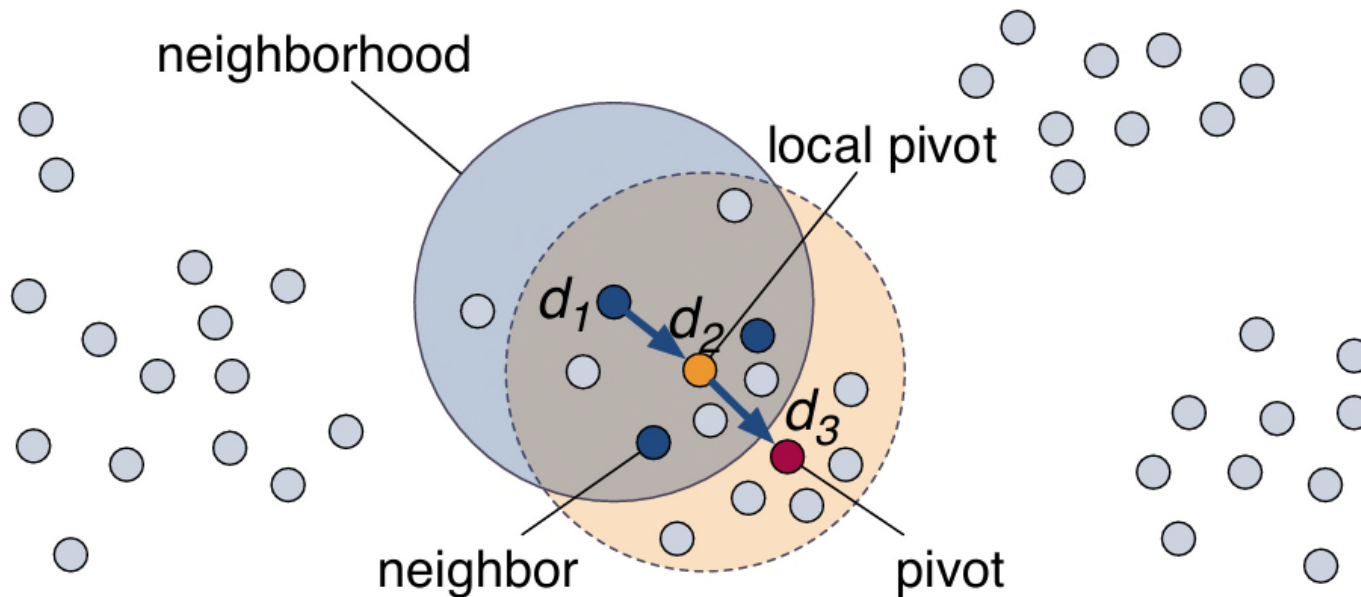


Event Detection

- Cluster tweets
 - In time (All)
 - In space (Geoburst)
 - By content (ET, StoryLine)

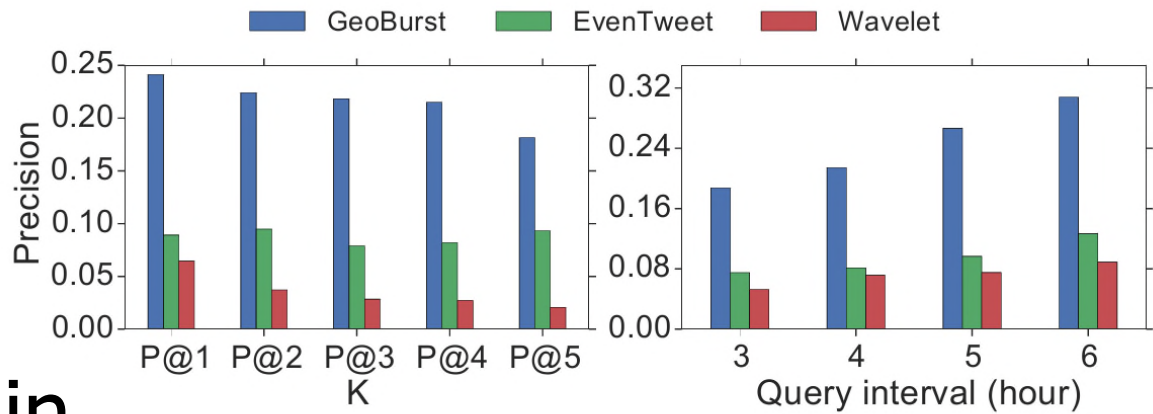
Geoburst

- Identify tweet clusters in space then decide on whether they represent events

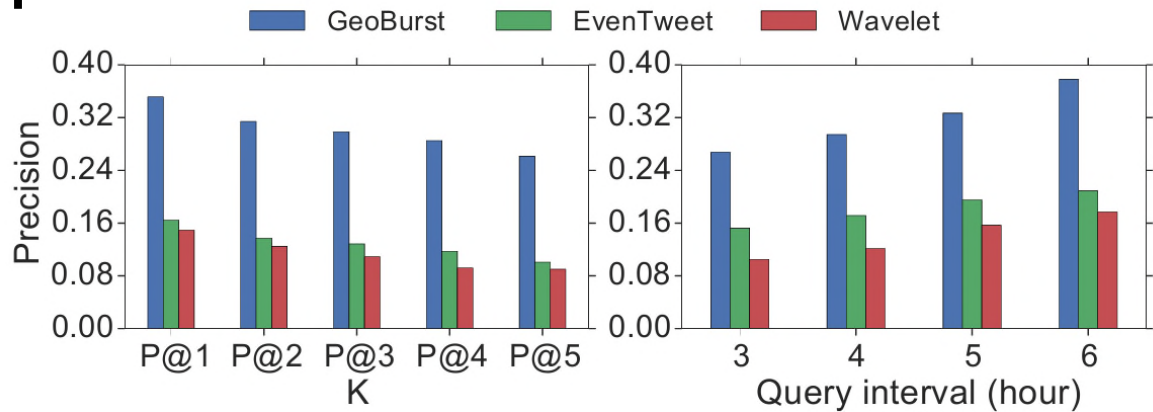


Geoburst Evaluation

- High precision in event detection



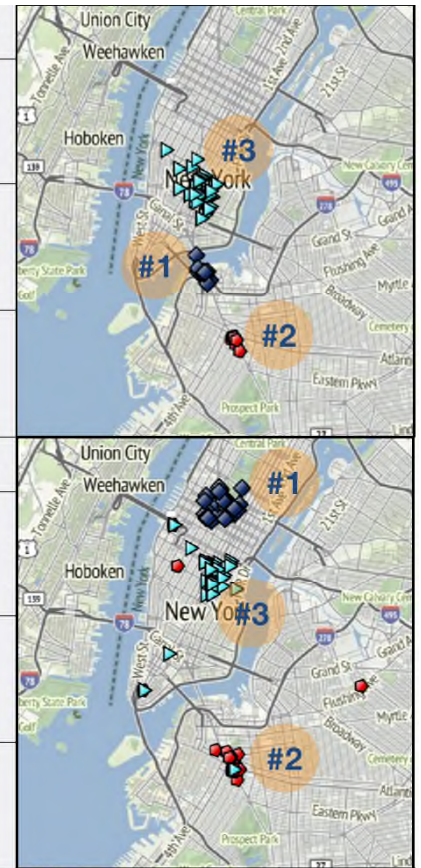
(a) Precision comparison (NY).



(b) Precision comparison (LA).

Geoburst Evaluation

GeoBurst		Is event?
# 1	<ol style="list-style-type: none"> 1. Festival of Light! #nyfol (@ The Archway under the Manhattan Bridge in Brooklyn, NY) 2. #Lasers and beats under the Manhattan Bridge! #NewYorkFestivalofLight #NYFOL @ DUMBO 3. New York Festival of Lights #nyfol #dumbo @ DUMBO, Brooklyn 	Yes
# 2	<ol style="list-style-type: none"> 1. Knicks vs. Nets at Barclays Center. @ Barclays Center http://t.co/PILk1xK3Tn 2. Brooklyn go hard @ Barclays Center http://t.co/iVUsJJ5TNG 3. Let's go Knicks! #NETS1107 (@ Barclays Center - @brooklynnets for @nyknicks vs @BrooklynNets) 	Yes
# 3	<ol style="list-style-type: none"> 1. #Thai Restaurant #spicythaifood (@ 104 2nd Avenue in New York, NY) 2. The ASIAN DISHES here are always my favorite. @ Ugly Kitchen 3. Dinner time with my family. Suuuuper Nice Indian RESTAURANT! @ Malai Marke Indian Cuisine. 	No
EventTweet		Is event?
# 1	<ol style="list-style-type: none"> 1. I practiced... Almost time for Amy Schumer. Jennifer (@ Carnegie Hall) https://t.co/HfqfTLMK2y 2. 2014 Gold Glove Awards Ceremony with Hall of Famers, All-Stars Jay Leno @ The Plaza Hotel 3. My best attempt at a selfie with Hugh Jackman after The River at CITS @ The River on Broadway 	No
# 2	<ol style="list-style-type: none"> 1. Knicks vs. Nets at Barclays Center. @ Barclays Center http://t.co/PILk1xK3Tn 2. Budweiser brings everyone together #family #nonewfriends @ Alchemy Tavern, Brooklyn 3. #Knicks vs #nets with my best gal. @ Barclays Center Brooklyn http://t.co/eXXMUKxpIs 	Yes
# 3	<ol style="list-style-type: none"> 1. #katespade @ Kate Spade / Jack Spade HQ http://t.co/g6jiFwyc4M 2. Inspiring keynote by Twitter CEO, Dick Costolo @GirlsWhoCode Gala. http://t.co/yEGh803CuT 3. I wonder if Jake from Statefarm covers Jumanji? 	No



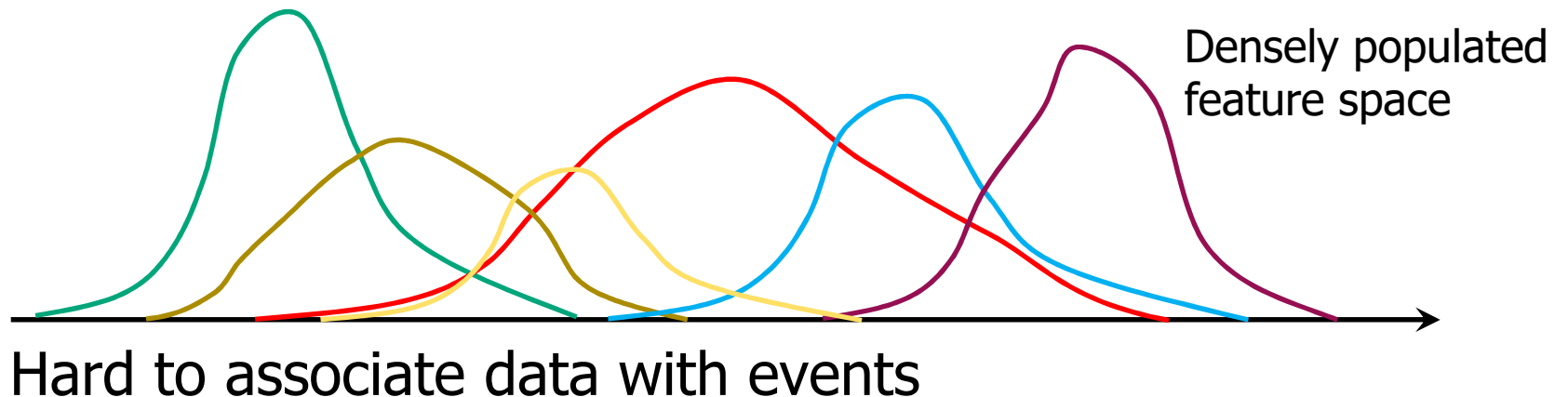
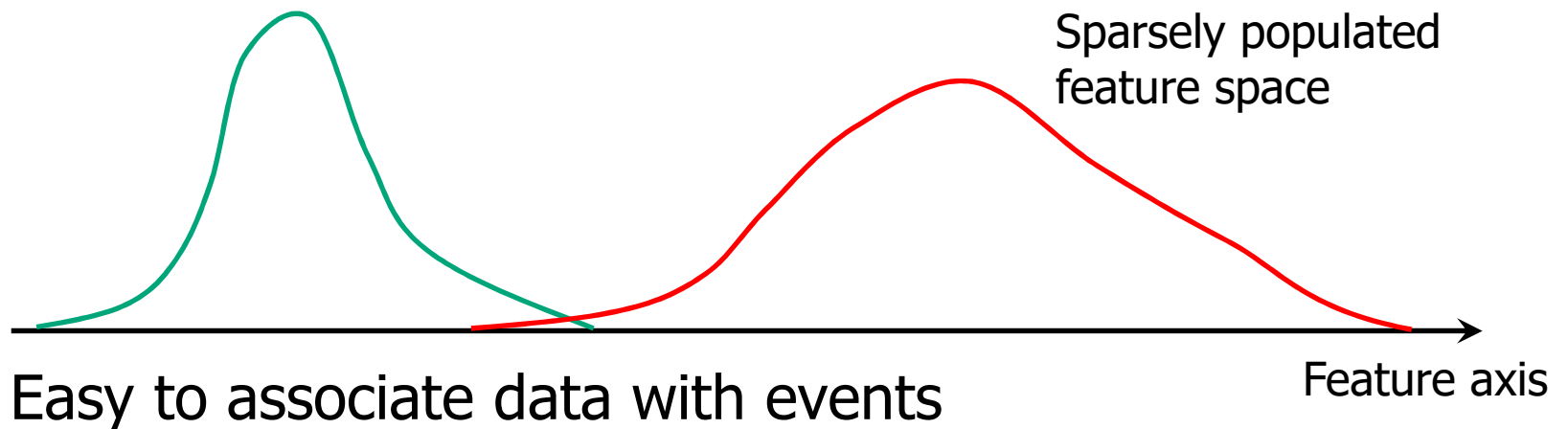
What's an Event?

The Keyword Frequency Domain

- *Question:* Can we detect and track events using frequency signatures only?
 - At first glance: text has complex semantics, so the ordering of keywords has great impact on meaning
 - "John killed Mary" versus "Mary killed John"
 - Do we need natural language processing to identify and track distinct events?

What's an Event?

A Data Association Problem





A Sparsity Observation

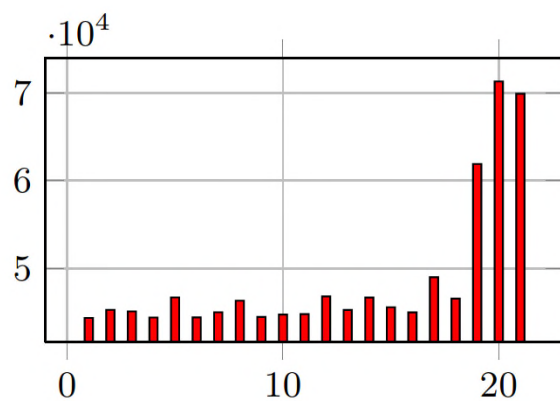
- Most languages have about 2000-3000 frequent words.
- Consider a 10-word event signature
 - There are at least 2000^{10} 10-word signatures
 - 1,000,000,000,000,000,000,000,000,000,000,000
- Tweets on an event are in the millions (at most)
 - 1,000,000
- The space of keyword signatures is vastly sparse:
 - Different events → Different signatures



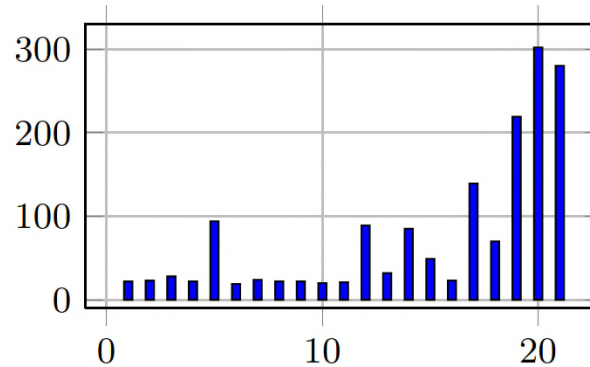
- Detect trending bigrams
- Identify similarity
 - Similar co-occurrence pattern in respective windows
 - Similar frequency pattern in respective windows
- Cluster tweets by bi-gram pattern similarity

Why Bi-grams?

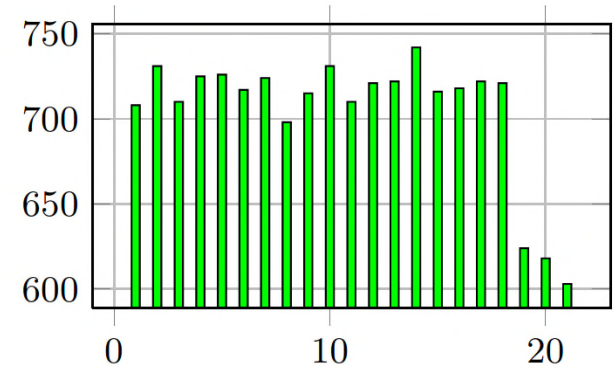
- Single keywords are not uniquely associated with event
- Trends in single keywords do not correspond to trends in events



(a) Tweets



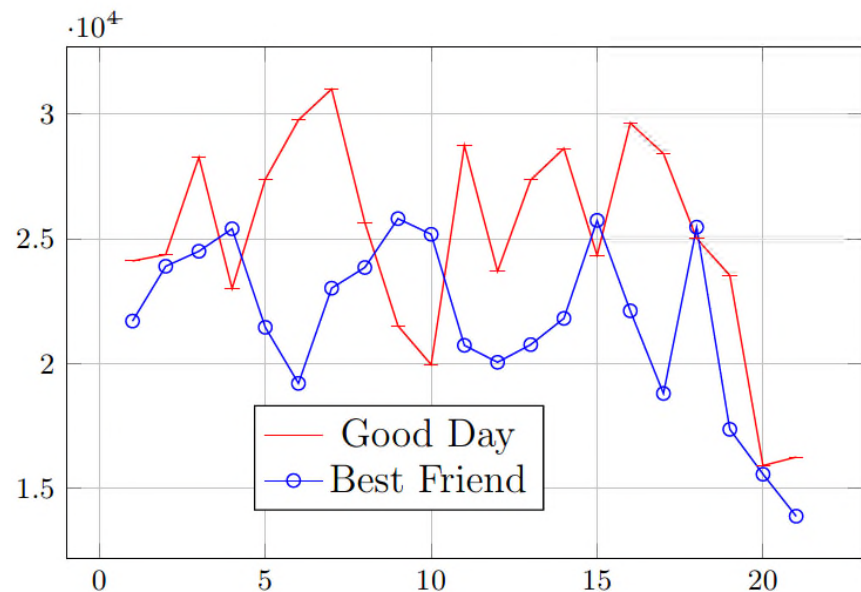
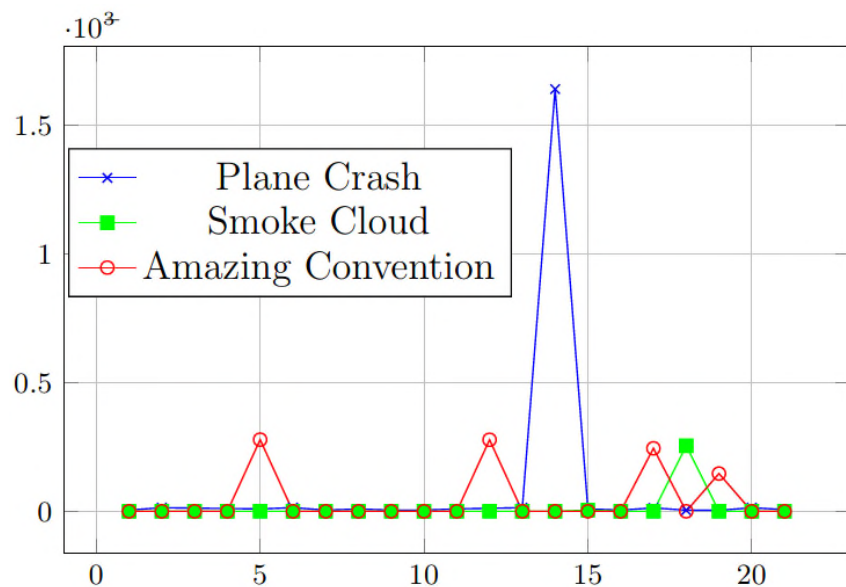
(b) Frequent Bigrams



(c) Frequent Unigrams

Representative Bigrams

- Representative bigrams clearly correspond to events
- Other random bigrams do not have temporal burstiness





Examples

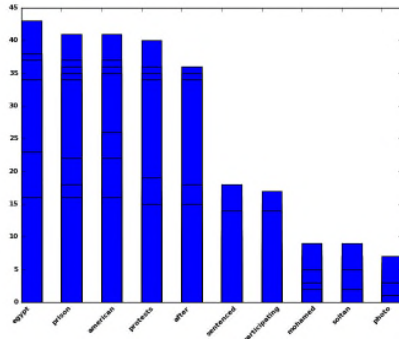
Event Representative Bigrams	Event	Time Interval(Day)
dense fog, advisory issued, rip current, statement issued, issued january, nws melbourne	Dense Fog Advisory and Rip Current Statement issued by NWS Melbourne	January 14, Interval-1
super bowl, tom brady, football game, playoff games, seahawks game, falcons seahawks, falcons game, dirty birds, nfl playoffs	Super bowl football nfl playoffs, Falcons (Dirty Birds) vs Seahawks game, People excited to watch Tom Brady	January 13, Interval-3
golden globes, globes tonight, downton abbey, red carpet,	Golden Globes Awards, Red Carpet, Maggi Smith of Downton Abbey (TV Series) wins at Golden Globes Awards.	January 14, Interval-0
season bad, bad girls, girls club, club atlanta	Bad Girls Club Atlanta Season 10 premiered on Jan 15	January 16, Interval-0, 1
patriots game, pats game, texans game	Patriots game Vs Texans game	January 13, Interval-3
sandy hook, hook conspiracy, hook shooting, conspiracy video	People watching Sandy Hook shooting conspiracy video.	January 14, Interval-1; January 16, Interval-1
music video, video macmerk, macmerk feat, feat yungking	Latest music video by Mac Merk Feat. Yung King	January 15, Interval-2
dark thirty, gangster squad, haunted house	Three movies released on same date: Zero Dark Thirty, Gangster Squad, and A Haunted House	January 13, Interval-3; January 14, Interval-0



StoryLine

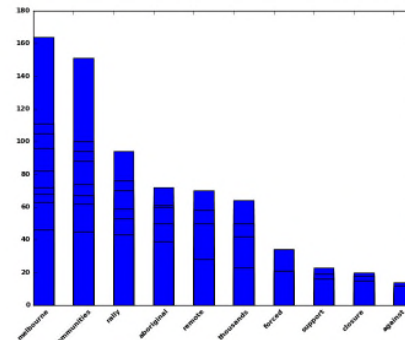
- Information gain to detect trending keywords pairs
- Consolidation of resulting tweet clusters

Consolidation



Event 1:
Top 10 keyword frequency distribution

Different events → different signatures



Event 2:
Top 10 keyword frequency distribution



Comparison

- Comparison of different detection techniques

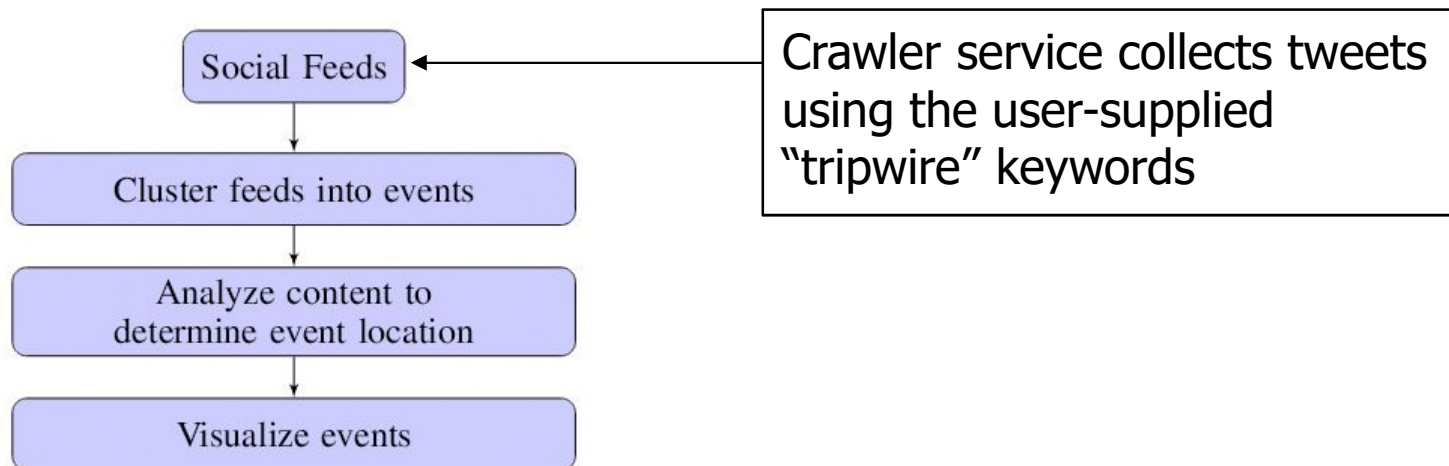
Algorithm	Traffic	Disaster	Protest	Armed Conflict
StoryLine	72.55%	76.92%	80.95%	88.24%
ET	57.14%	36.36%	86.36%	61.90%
TopicModel	55.10%	60.87%	65.22%	69.57%
GeoTag	66.67%	23.33%	47.37%	41.38%



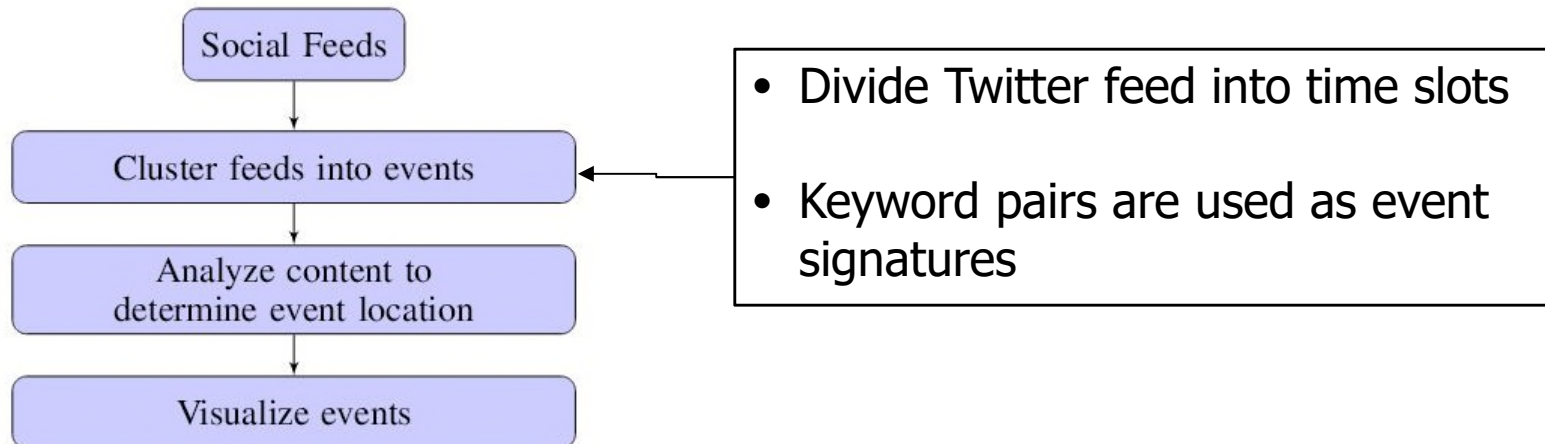
Earthquake Detection

Earthquake Location	Happened Time	Detection Time	Delay
Midoro, Philippines	10/19/2015 13:50	10/19/2015 18:26	4:16
Vanuatu	10/20/15 21:52	10/21/2015 02:40	4:48
Afghanistan	10/26/15 09:09	10/26/15 10:17	1:08
Molucca islands	01/11/16 16:38	01/11/16, 20:41	4:03
Afghanistan	01/12/16 20:05	01/12/16 22:59	2:54
Alberta, Canada	01/12/16 17:30	01/12/16 22:59	5:29
Urakawa, Japan	01/14/16 03:30	01/14/16 04:09	0:39
Alaska	01/24/16 10:30	01/24/16 11:37	1:07
Morocco	01/25/16 04:22	01/25/16 10:44	6:22
Taiwan	02/06/16 19:57	02/06/16 21:39	1:42
Fiji	02/06/16 01:39	02/06/16 02:41	1:02
Indonesia	02/12/16 10:02	02/12/16 13:28	3:26
Oklahoma	02/13/16 17:07	02/13/16 22:37	5:30
New Zealand	02/14/16 00:13	02/14/16 04:38	4:25
Wasco, CA	02/24/16 00:02	02/24/16 00:37	0:35
Antarctica	02/23/16 18:08	02/24/16 00:37	6:29
Cebu, Phillipine	03/01/16 14:52	03/01/16 17:14	2:22
Sumatra, Indonesia	03/02/16 12:49	03/02/16 14:19	1:30

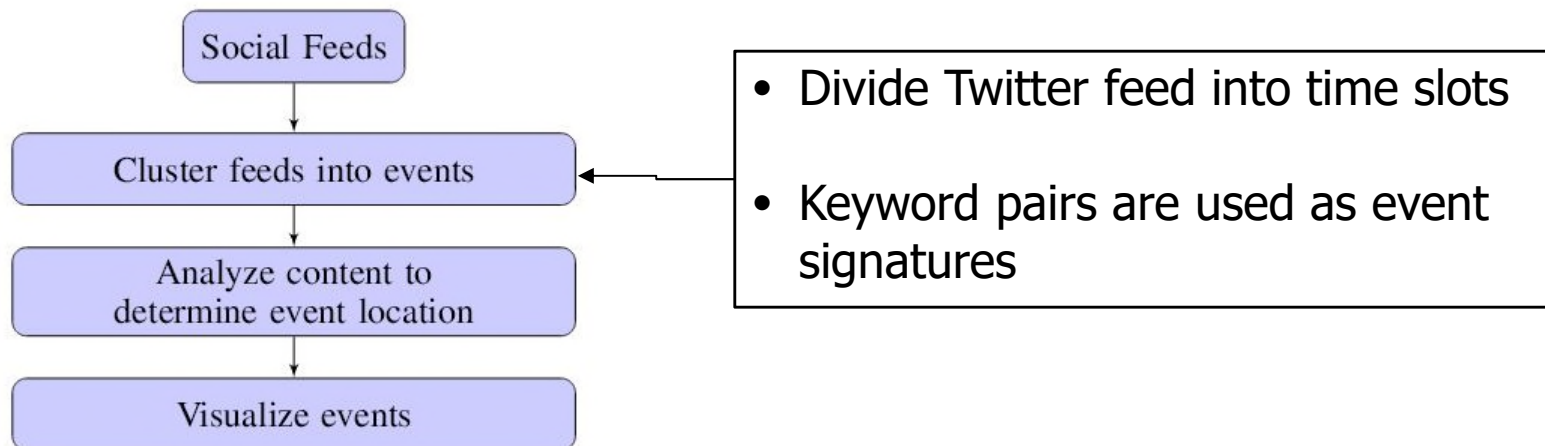
Unusual Event Detection



Unusual Event Detection



Unusual Event Detection



Detection:

- Find keyword pairs that occur disproportionately in the current window compared to the previous one (analytically: sort keyword pairs by information gain).
- All tweets with the same keyword pair form one cluster describing the event.

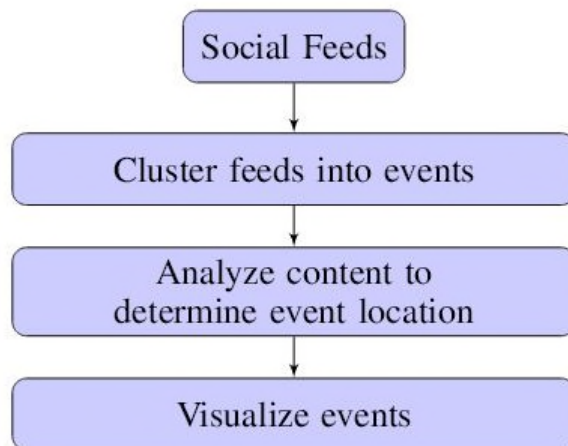


Unusual Event Detection

Example: Detecting Traffic Events

Keyword Pair	Event Cluster
(Crash, Rancho)	#BREAKING: Massive crash has traffic down to a trickle on SB15 at Via Rancho Parkway in Escondido. #BreakingNews #SigAlert Major traffic crash on the 15-southbound at Via Rancho Parkway. traffic backed up for miles. some lanes open now.
(Collision, North)	Traffic collision on SB I-5 just north of Encinitas Blvd. Vehicle hit center median. One lane blocked on SB I-5 just north of the San Diego-Coronado Bridge due to traffic collision.
(Lanes, Pasadena)	All lanes were closed on the westbound 210 Freeway in Pasadena because of the crash and rush hour traffic NBCLA: TRAFFIC ALERT: Big rig crash shuts down lanes of WB 210 Fwy in Pasadena. SIGALERT Pasadena - 210 W before Rosemead: The carpool & 2 left lanes are closed due to a crash. Traffic bad from Myrtle. E heavy from Lake.

Event Localization



- Find location tags
- Determine source location
- Find references to locations
- Find landmark names
- Vote on most likely location

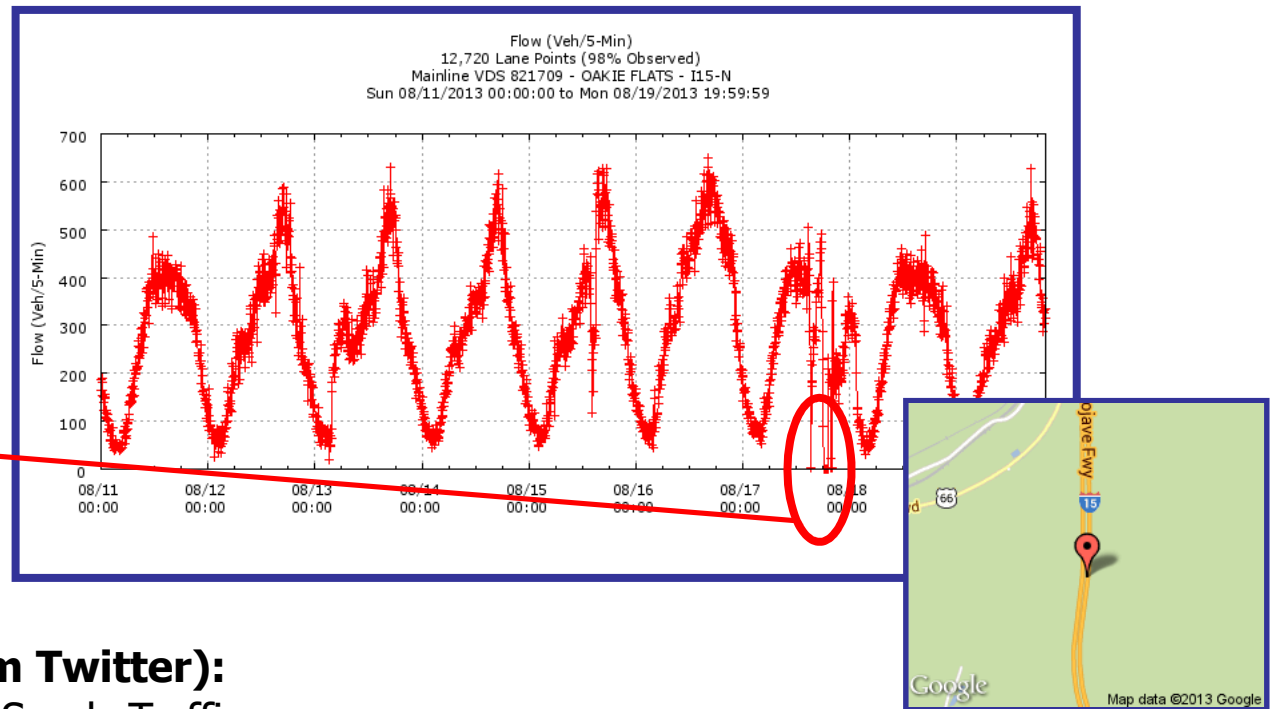
Example 1: Forest Fire

Time, Location:
6pm, Aug 17th, I-15 N, LA

Event Signature:
"Cajon", "Pass"

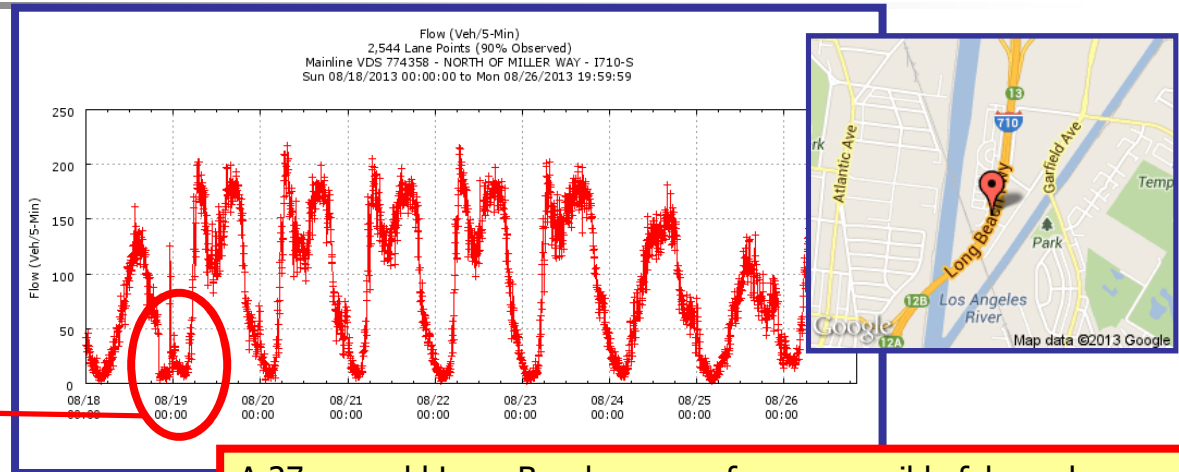
Possible Explanation (from Twitter):
Cleghorn Fire in **Cajon Pass** Snarls Traffic on I-15: (KTLA) One northbound lane on the 15 Freeway was reopened...
<http://t.co/nieqh4nsMX>

URL:
<http://t.co/nieqh4nsMX>



One northbound lane on the 15 Freeway was reopened Saturday evening in the Cajon Pass as firefighters continued to battle the so-called Cleghorn fire, authorities said. All southbound lanes on Interstate 15 were open, according to the U.S. Forest Service. State Road 138 remained closed from the 15 Freeway to Summit Valley Road.

Example 2: Pedestrian Death



Time, Location:
8pm, Aug 18th, SB I-710, LA

Event Signature:
"Drunk", "Kills"

Possible Explanation (from Twitter):
27-Year-Old Drunk Driver Hits, Kills Man
Trying To Stop Traffic On 710
<http://t.co/rMoI7DxFH4>

URL:
<http://t.co/rMoI7DxFH4>

A 27-year-old Long Beach woman faces a possible felony charge after fatally hitting a pedestrian on the 710 Freeway while she drove intoxicated.

Melanie Gosch struck a man in his 20s at about 8 p.m. on Sunday near Imperial Highway on the southbound 710 in South Gate, according to City News Service. The man, whose name has been withheld, was allegedly trying to stop traffic in the number three freeway lane.

Police responded to the scene after a caller reported a "long-haired man in dark clothing" on the freeway. Within a minute, the California Highway Patrol received a call that a pedestrian was lying in the freeway. The man was pronounced dead at the scene. Gosch stopped her 2007 Nissan Sentra after striking the man, and she was arrested and booked on suspicion of causing injury or death while driving under the influence of alcohol or drugs.

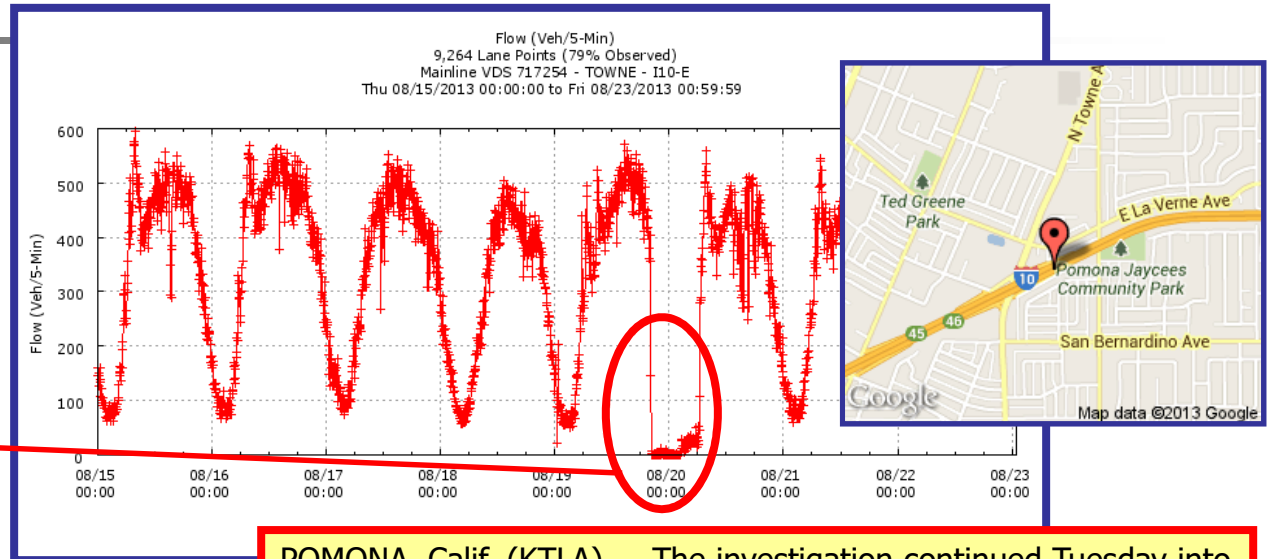
Example 3: Vehicle Crash

Time, Location:
8pm, Aug 19th, I-10E, LA

Event Signature:
"Crashes", "Divider"

Possible Explanation (from Twitter):
1 dead, 9 hurt in fiery **crashes** on **10** Fwy in Pomona; EB 10 closed, traffic allowed to pass along center **divider**
<http://t.co/cCu8xLzx1U>

URL:
<http://t.co/cCu8xLzx1U>



POMONA, Calif. (KTLA) — The investigation continued Tuesday into a pair of chain-reaction crashes on the 10 Freeway in Pomona that left one person dead and eight others injured. One killed, eight hurt in pair of chain-reaction crashes on 10 Freeway in Pomona. It all happened around 8 p.m. on Monday on the eastbound 10 Freeway near Towne Avenue, according to the California Highway Patrol. The first crash involved four cars and created a traffic back-up, CHP officials said. That's when a second crash occurred involving a big rig with a full tank of diesel fuel and three other vehicles. The fuel tank of the big rig ruptured, causing it to burst into flames, authorities said. The driver of the semi was able to get out safely. However, the driver of a red BMW that became trapped under the big rig was not able to escape. That person, who was not immediately identified, died at the scene.

Example 4: Vehicle Explosion

Time, Location:
8pm, Sept 1st, 405N, LA

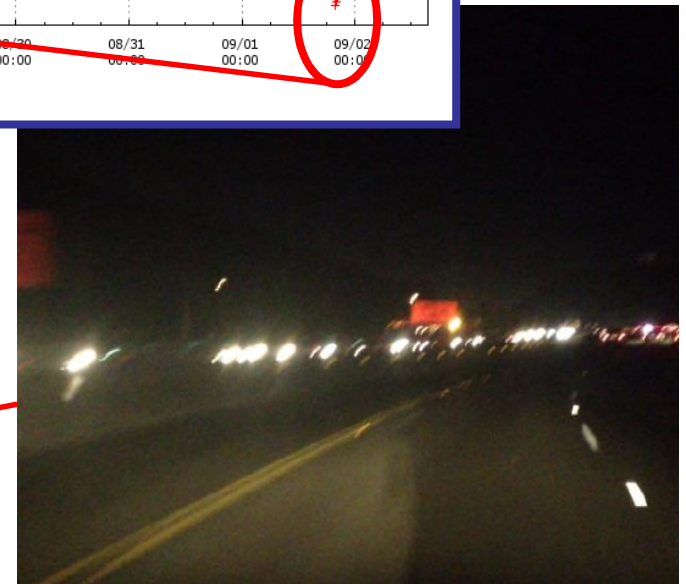
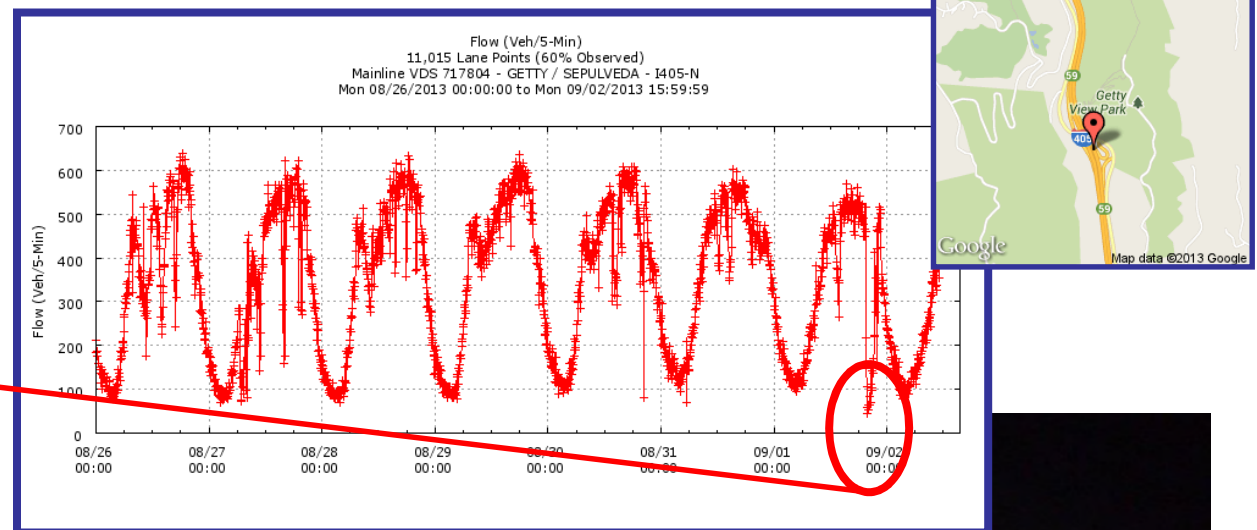
Event Signature:
"Explosion", "lifeinLA"

Possible Explanation (from Twitter):

Heading to LAX. Traffic backed UP on 405
N due to an SUV explosion... #lifeinLA
#405parkinglot <http://t.co/UBxUYMGrkS>

URL:

<http://t.co/UBxUYMGrkS>





Localization Accuracy

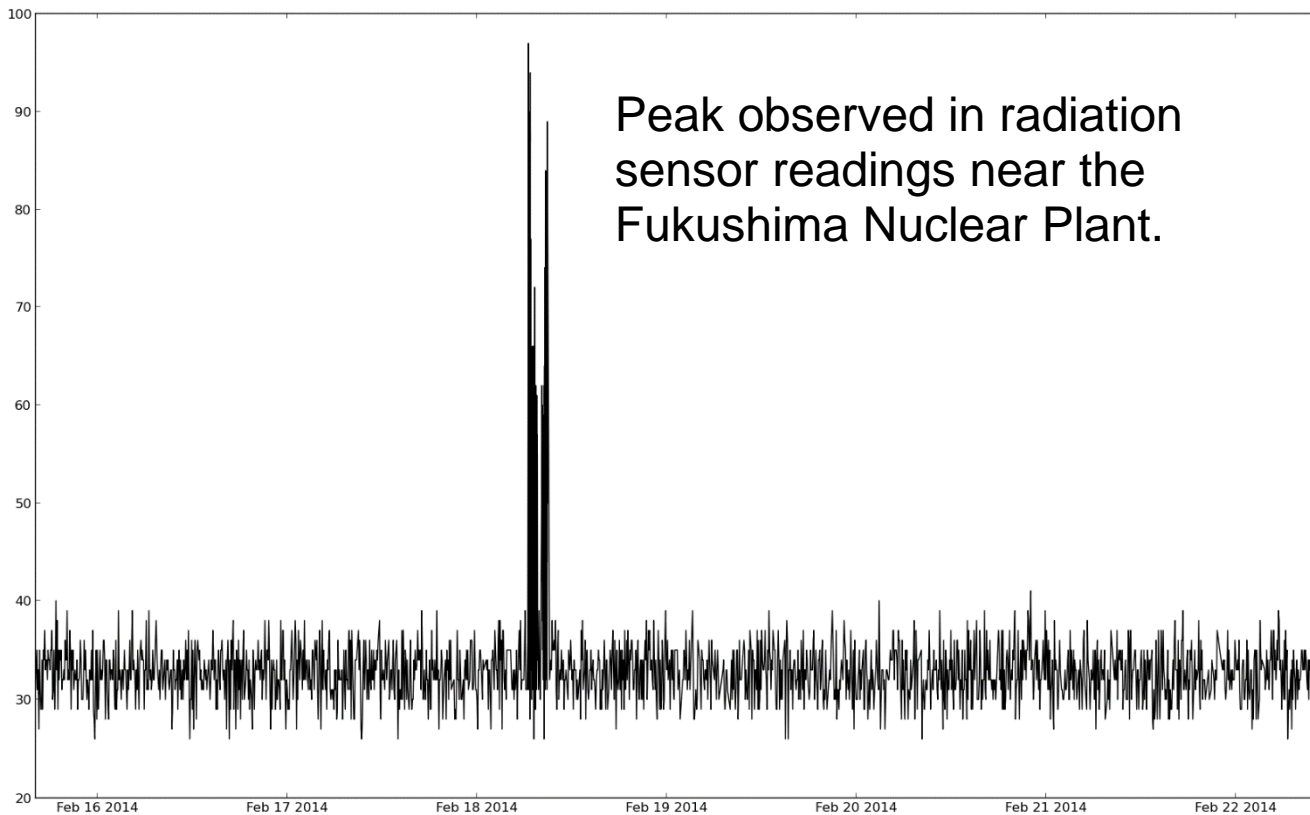
Percentage of successfully localized events

IG Range	100%			50%			25%			10%		
	SF	SD	LA	SF	SD	LA	SF	SD	LA	SF	SD	LA
[0 – 0.005)	76.7 %	80.9 %	75.3 %	51.6 %	71.4 %	52.8 %	48.8 %	66.6 %	43.7 %	44.1 %	66.6 %	39.1 %
[0.005 – 0.01)	68.7 %	100 %	100 %	56.2 %	100 %	100 %	50% %	100 %	100 %	43.7 %	100 %	100 %
[>0.01]	85.7 %	100 %	100 %	85.7 %	100 %	100 %	57.1 %	100 %	100 %	57.1 %	100 %	100 %

- Randomly removed location metadata for 50%,25%, and 10% sources
- SD, LA still had 100% localization for high IG events for 25%(ideal) source metadata availability

Example: Anomaly Explanation

Fukushima leak (Feb 2014)



Example: Anomaly Explanation

Fukushima leak (Feb 2014)

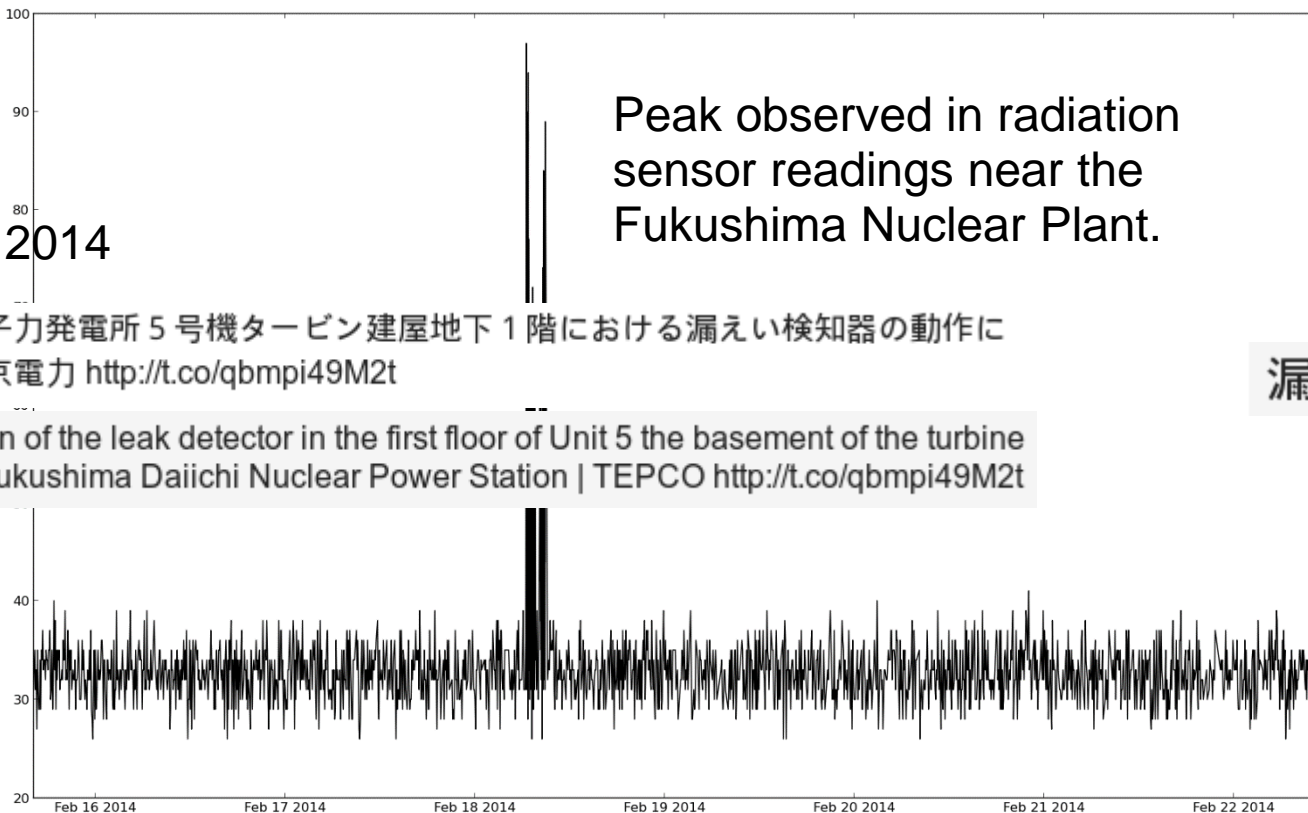
Feb 19, 2014

Peak observed in radiation sensor readings near the Fukushima Nuclear Plant.

福島第一原子力発電所 5号機タービン建屋地下1階における漏えい検知器の動作について | 東京電力 <http://t.co/qbmpi49M2t>


The operation of the leak detector in the first floor of Unit 5 the basement of the turbine building of Fukushima Daiichi Nuclear Power Station | TEPCO <http://t.co/qbmpi49M2t>

漏れ、建物



Results:

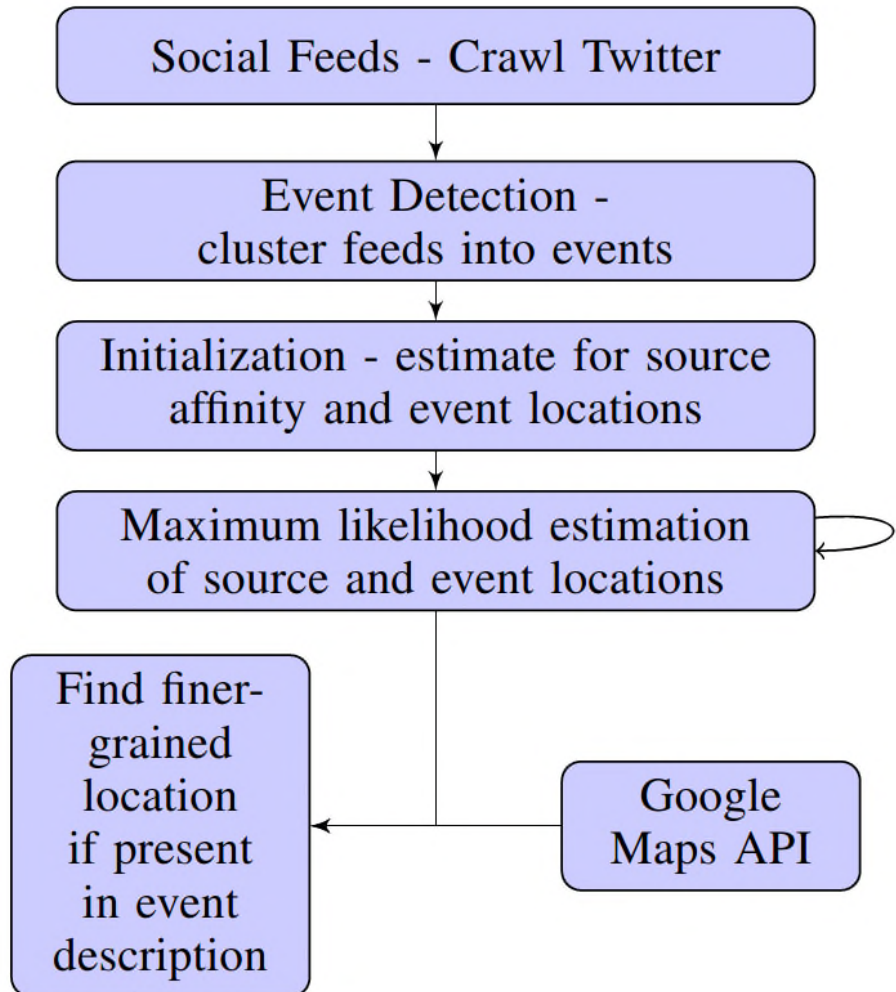
Anomaly Explanation Study

- Recall = Percentage of events explained in top 5 tweets.
 - Los Angeles: 83%
 - San Diego: 100%
 - San Francisco: 100%(for Hazard related events)
- Precision: Average rank of correct explanation of found traffic anomalies (after sorting and filtering). 

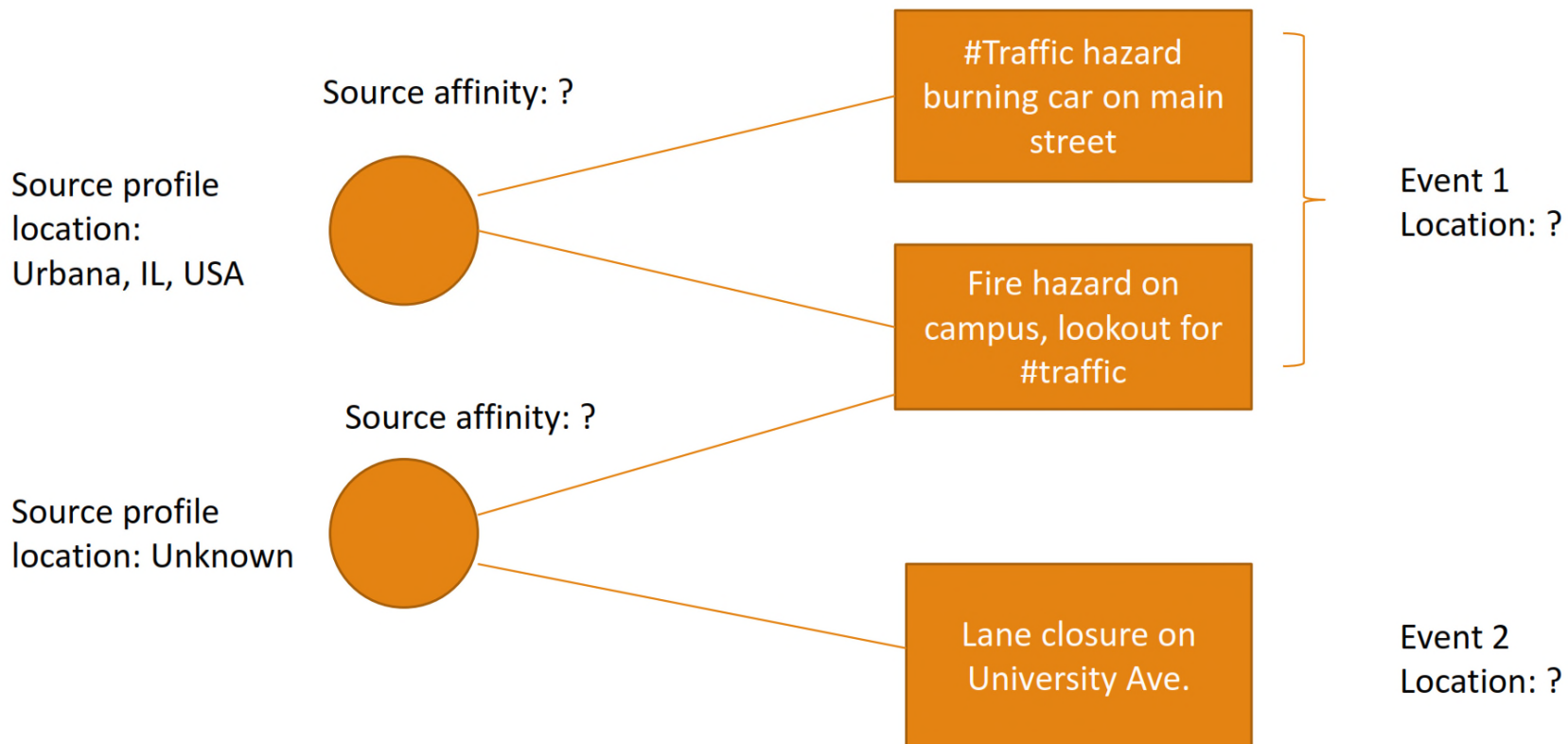
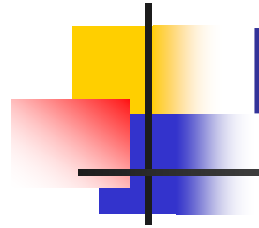
	Information Gain (IG)	Spatial filter + IG	Credibility filter + IG	Spatial + Cred. + IG
Los Angeles	1.43	1.36	1.36	1.29
San Francisco	1.6	1.6	1.2	1
San Diego	1.14	1.07	1.21	1.2

Joint Event and Source Localization

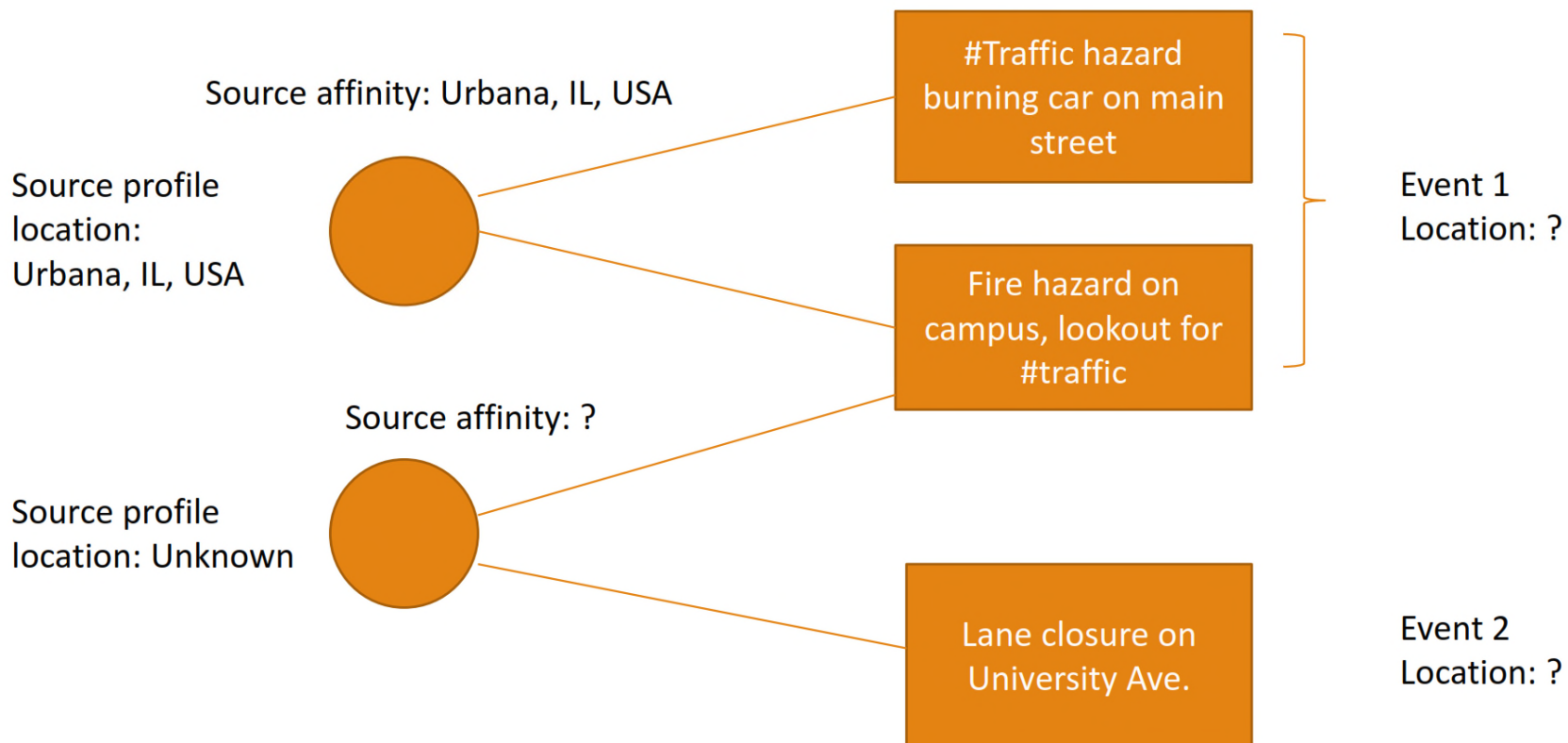
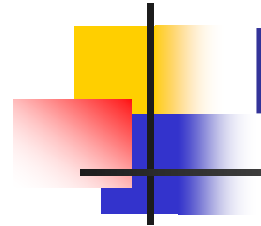
- Sources have affinity to locations. Determine them and propagate to location of tweets they emit.



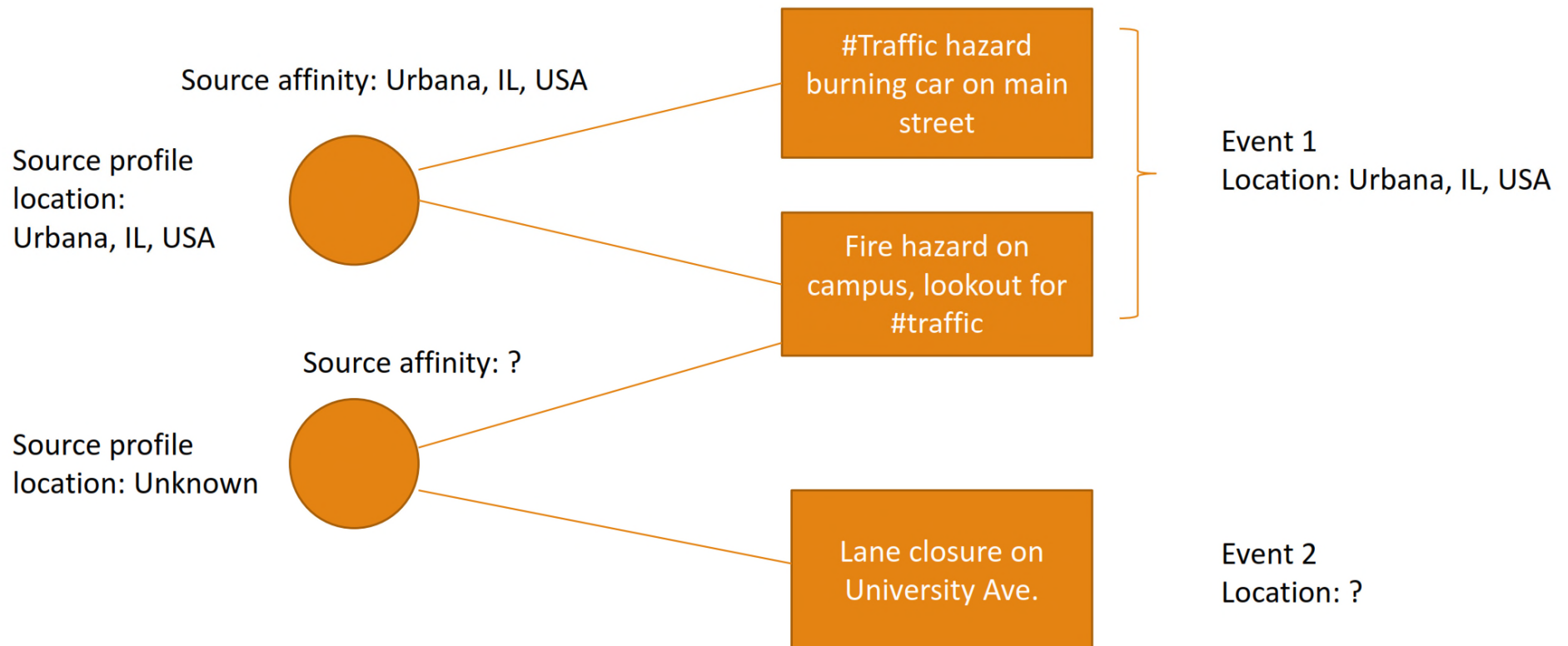
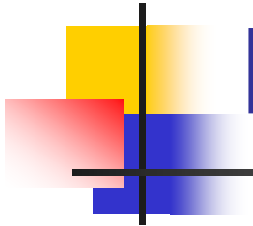
Joint Event and Source Localization



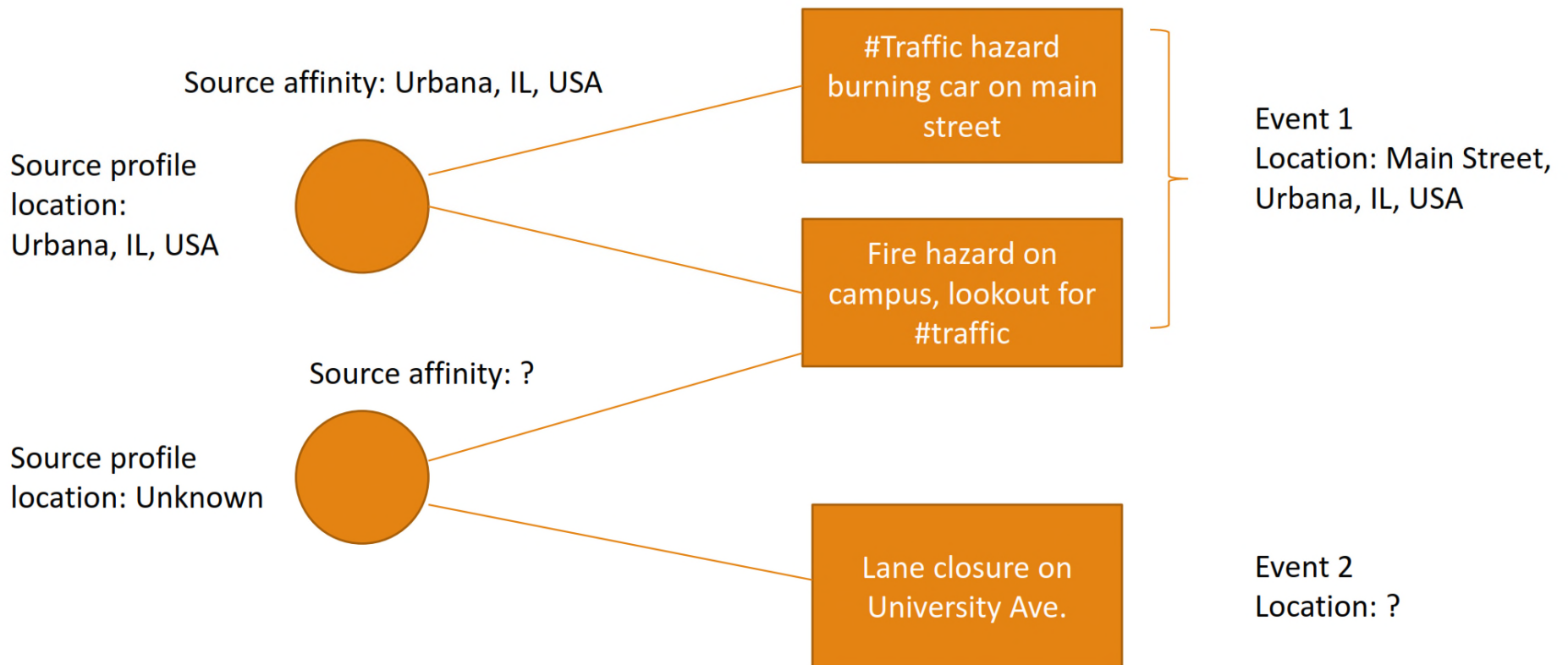
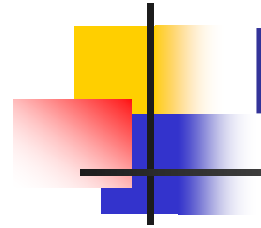
Joint Event and Source Localization



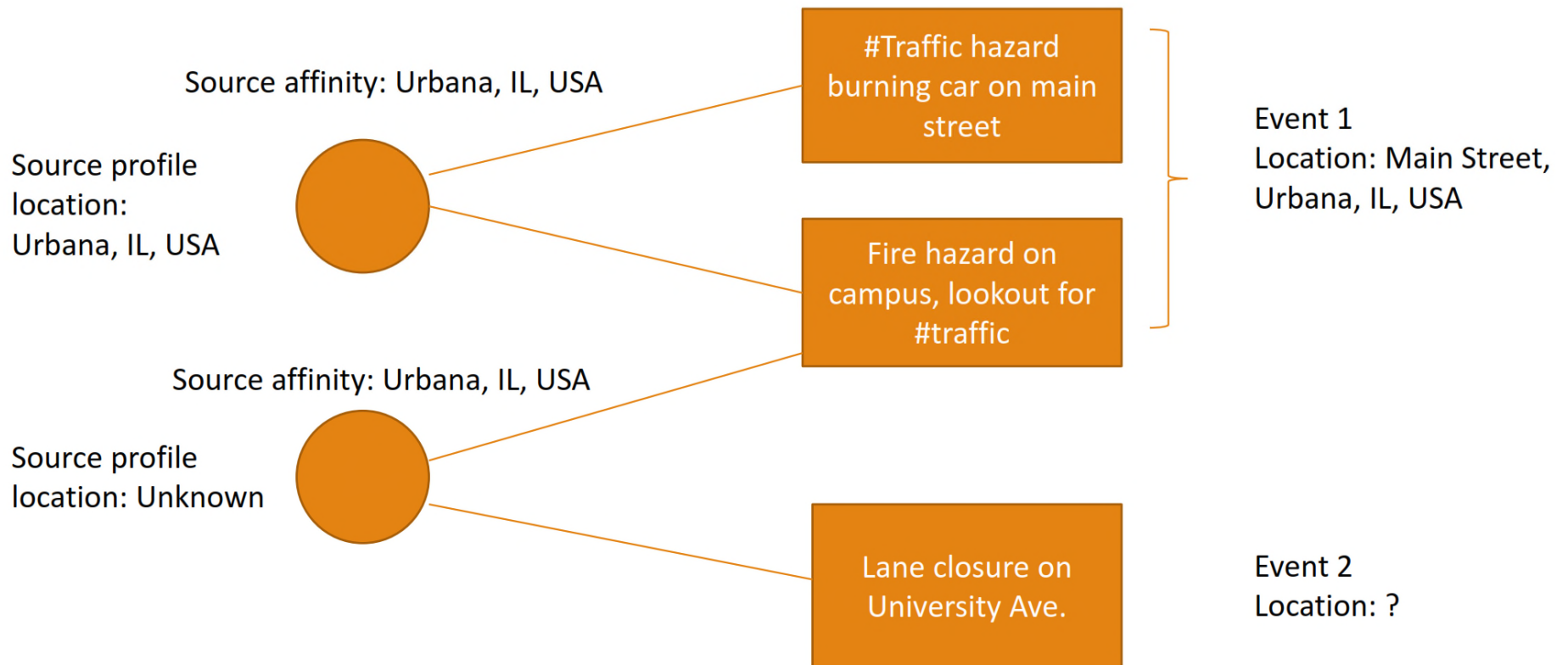
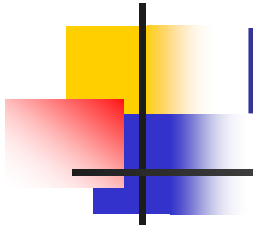
Joint Event and Source Localization



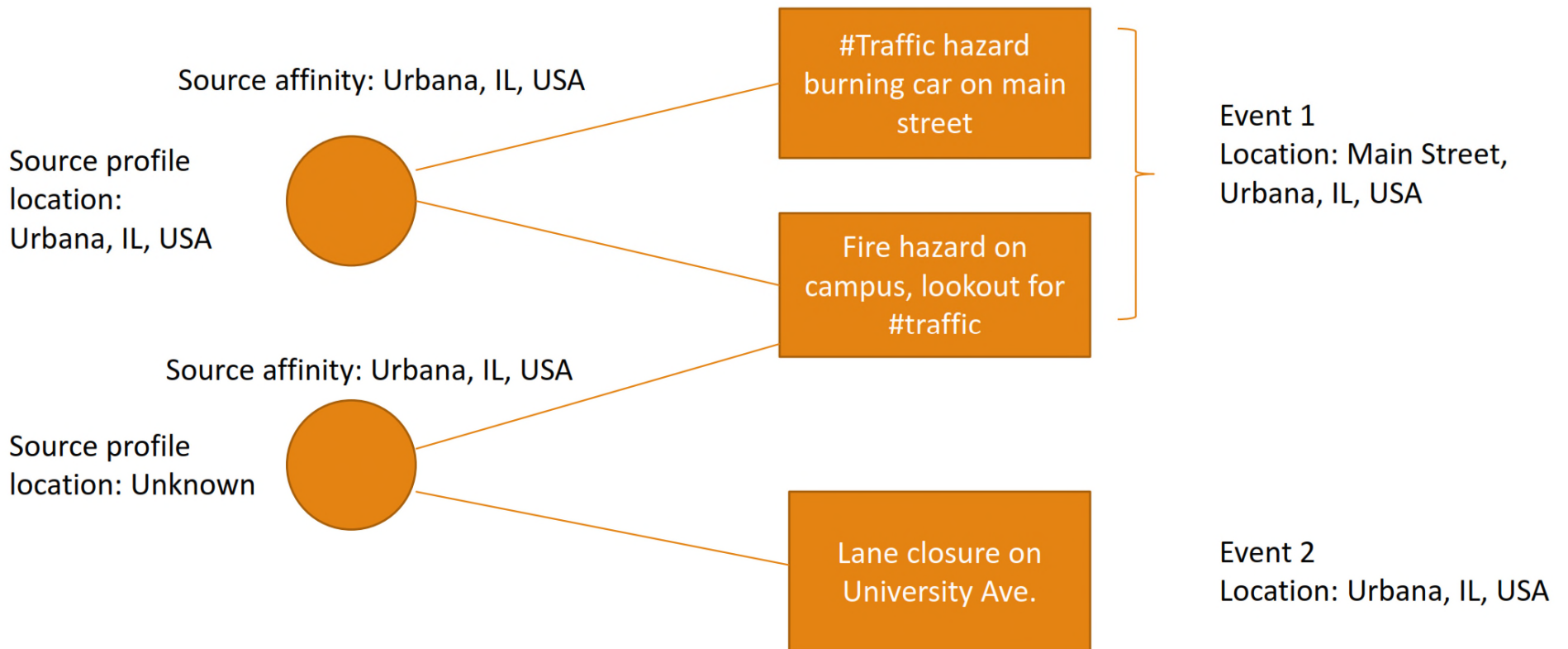
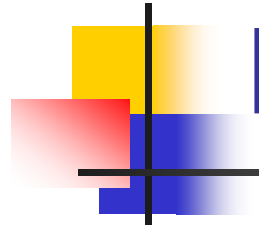
Joint Event and Source Localization



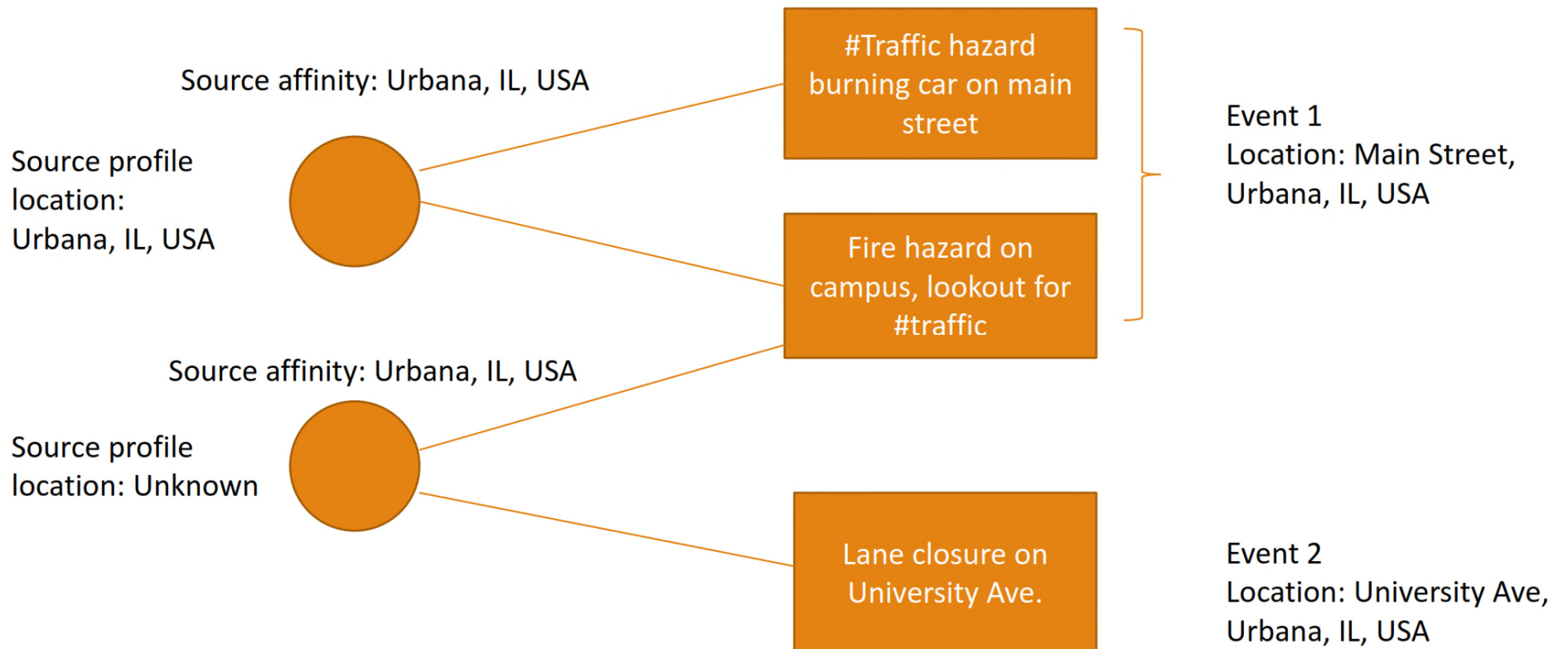
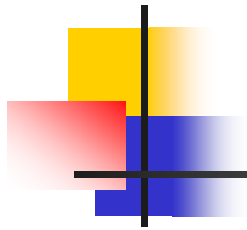
Joint Event and Source Localization



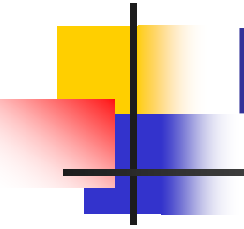
Joint Event and Source Localization



Joint Event and Source Localization



Joint Event and Source Localization



Fraction non-localized	EM	Baseline 2	
		word length ≥ 5	word length ≥ 6
0.2	50.14%	41.85%	38.92%
0.3	48.63%	40.8%	38.21%
0.4	47.72%	40.24%	38.14%
0.5	46.59%	40.13%	37.86%
0.6	46.05%	39.53%	37.43%
0.7	44.89%	38.88%	36.44%



Conclusions

- Social networks as sensor networks
- Goal: accurately detect and localize unusual events (matching a user's interest) without semantic inspection of text
- A new modality for sensing (acoustic, magnetic, optical, and now social)
 - Physical events impact a medium which responds with a signature
 - Analyze signature to measure the event

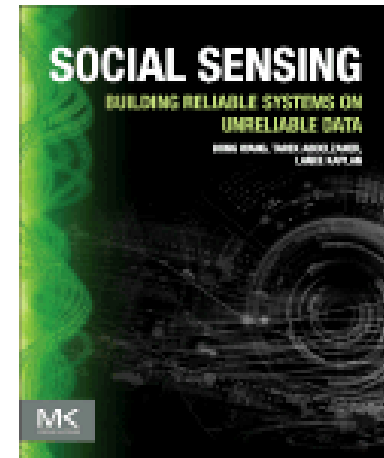


Tool Available Online

Link:

apollofactfinder.net

Book:



Social Sensing: Building Reliable Systems on Unreliable Data

Dong Wang, Tarek Abdelzaher, Lance Kaplan, *Morgan Kaufmann*,
1st Edition, April, 2015