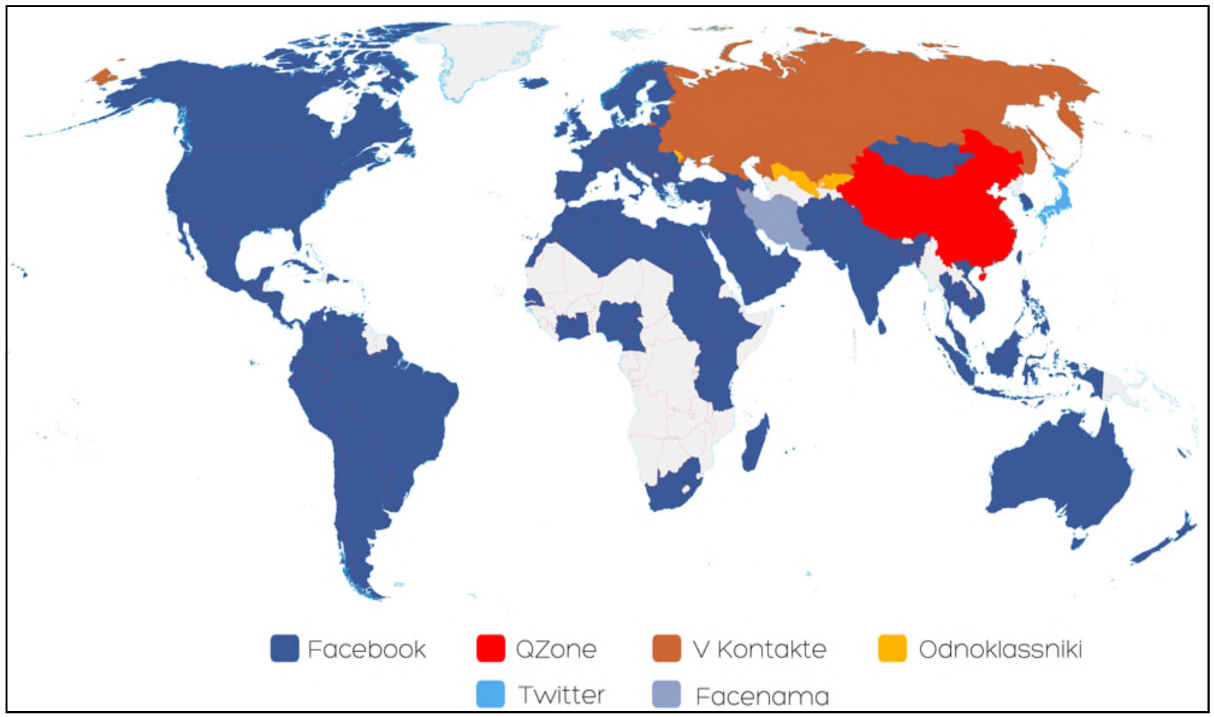
Instagram

Social Networks Popularity

- Facebook is the leading social network in 129 countries
- Around 1.6 billion active users
- Content is mostly private and not accessible for general browsing.



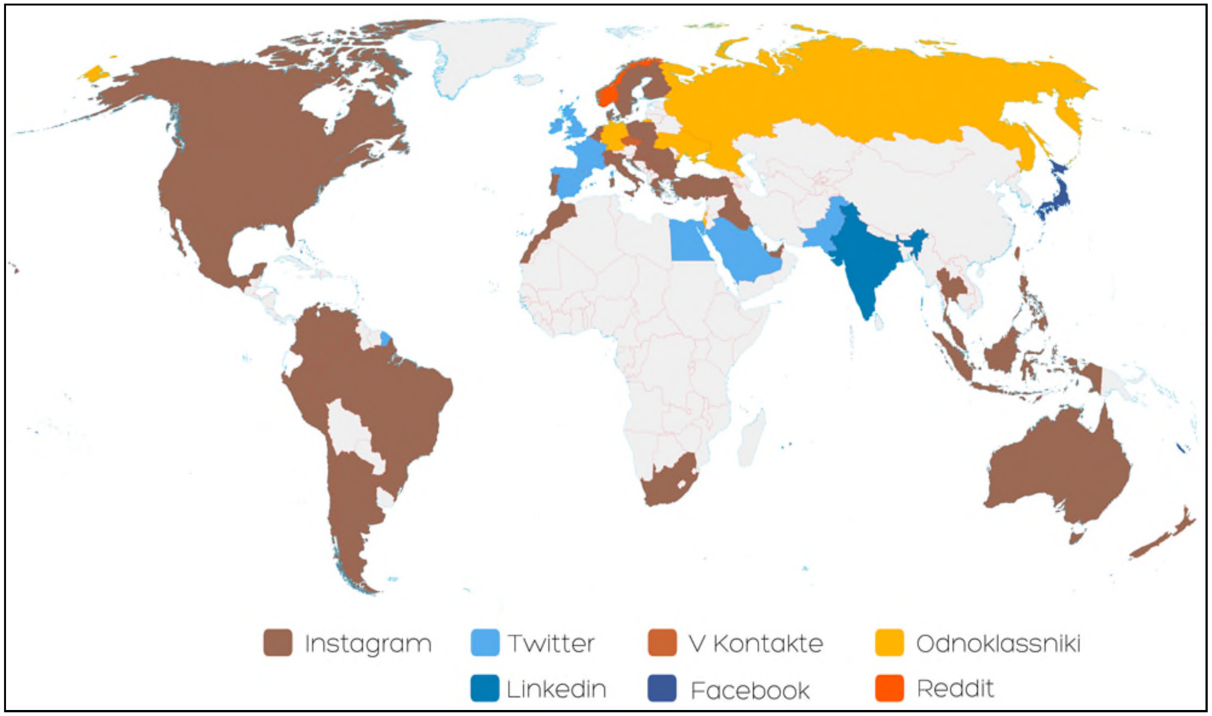
1st ranked social network in each country

Social Networks Popularity

•Most popular image sharing network

•600m monthly active users & 80m image uploads per day

•Content is publicly available via few web sources



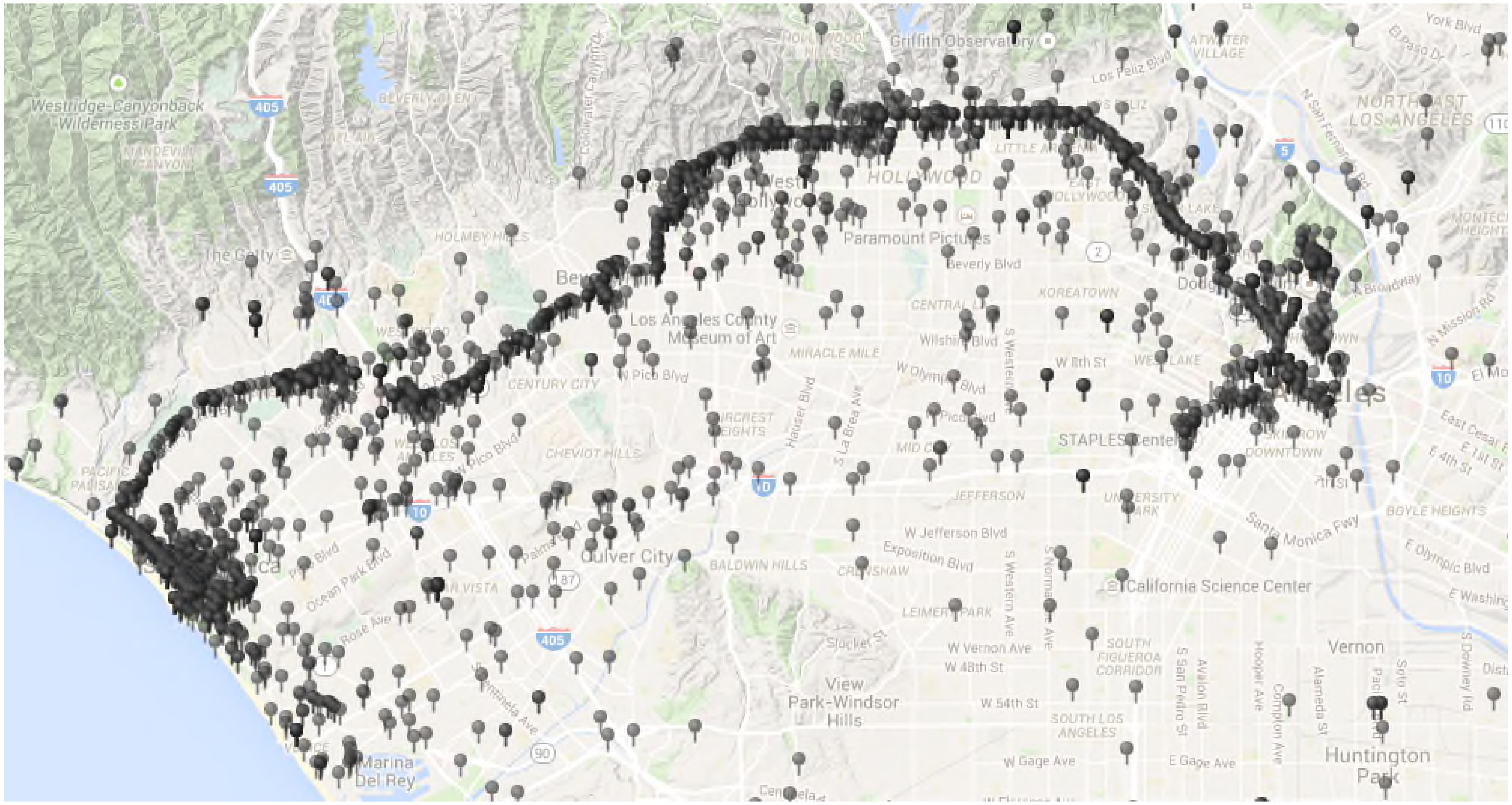
2nd ranked social network in each country

Detection, Localization and Tracking with Instagram

- *Motivation:* Why use Instagram for Event Localization?
  - Unlike text that can be written from anywhere, pictures of an event are generally taken at the event location.
  - Almost 25% of Instagram images are geotagged (but only 2% of tweets).
- *Analogy:* An Instagram picture = “binary sensor” signifying event “detection”
- *Method:* Leverage prior sensor network literature on localization and tracking with binary sensors

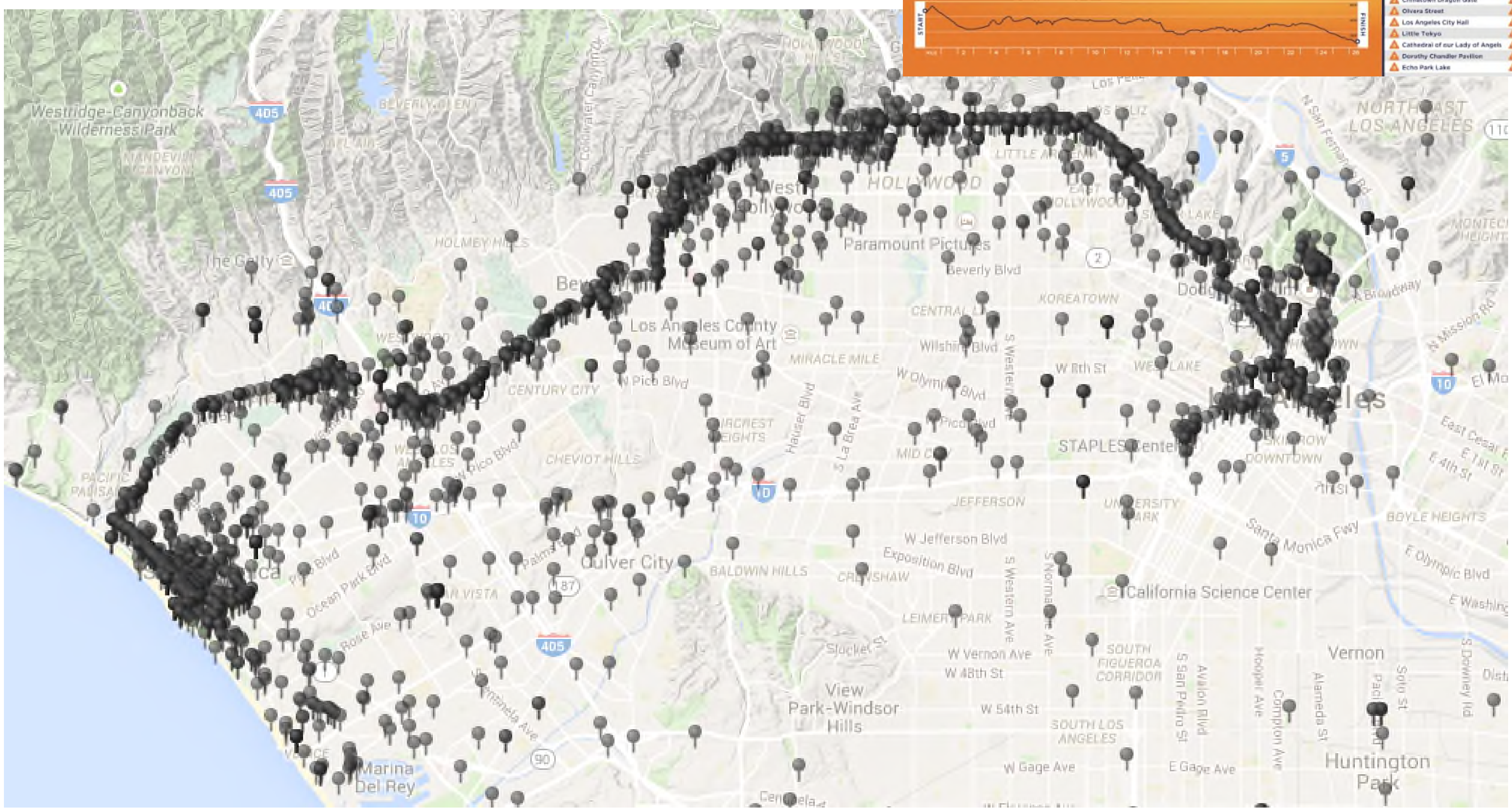
Feasibility of Instagram-based Localization

- Example: Tracking "LA Marathon" (2015)



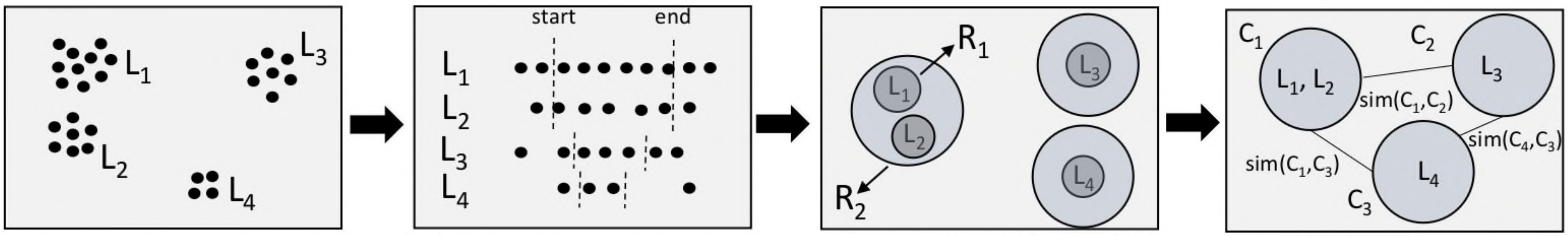
Instagram

- Tracking “LA Marathon”



Instagram Event Detection

- Paper accepted to Infocom 2017

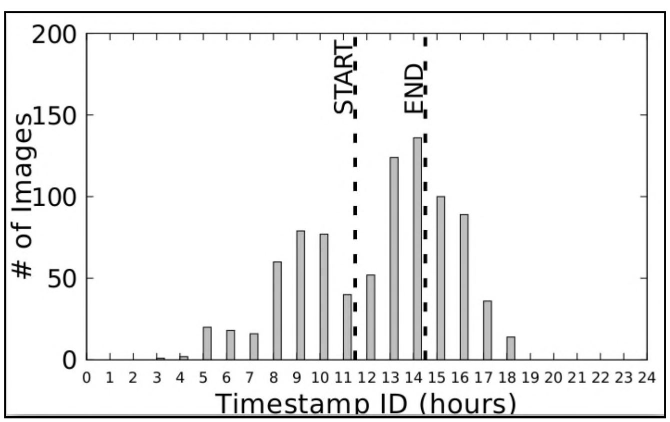


Step 1: Process the data from API and arrange locations in decreasing order of unique users and images.

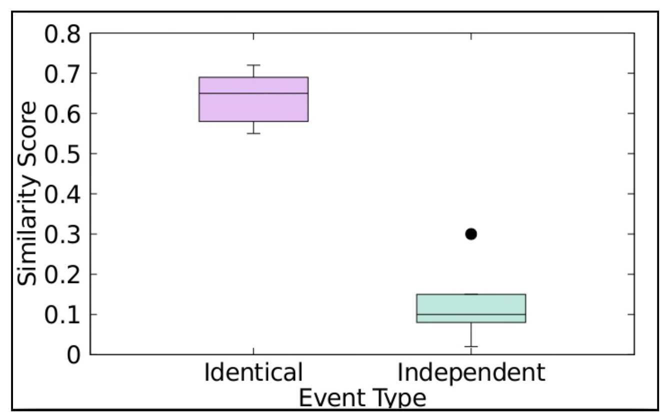
Step 2: Estimate the start and end times for each location using the time distribution of the posts.

Step 3: Determine the best clustering range using the silhouette score and group the locations.

Step 4: Eliminate the false positive clusters by finding the similarity between the top ten hash tags .



Estimated start and end times of an event



Similarity using hash tags between clusters

Event Detection Results

TABLE V. RECALL

Dataset	Our Localization Algorithm	Tag Similarity Localization				Geo Event Detection [23]				Points of Interest [5]
		$X =$ 20%	$X =$ 40%	$X =$ 60%	$X =$ 80%	$\theta = 0.2$	$\theta = 0.4$	$\theta = 0.6$	$\theta = 0.8$	
Taylor Swift	28/28	24/28	25/28	26/28	26/28	28/28	28/28	28/28	28/28	27/28
Maroon V	17/17	13/17	15/17	17/17	17/17	17/17	17/17	17/17	17/17	17/17
Marathon	5/5	3/5	4/5	4/5	4/5	5/5	5/5	5/5	5/5	5/5
Tornado	3/3	2/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3

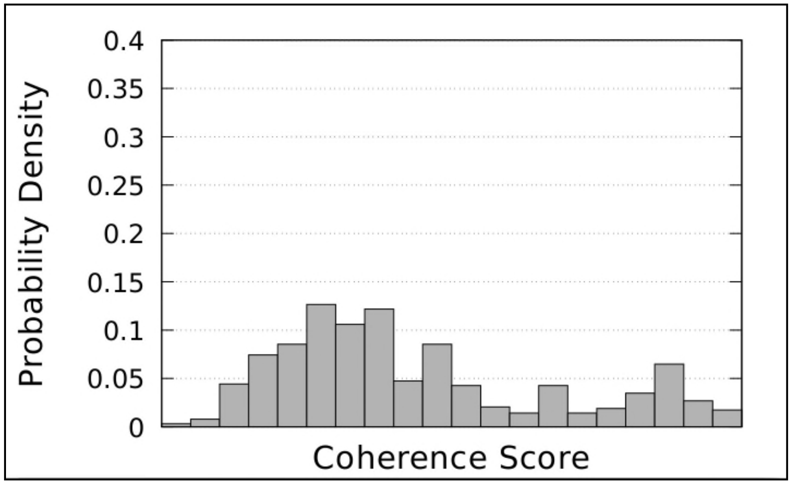
TABLE VI. FALSE POSITIVES

Dataset	Our Localization Algorithm	Tag Similarity Localization				Geo Event Detection [23]				Points of Interest [5]
		$X =$ 20%	$X =$ 40%	$X =$ 60%	$X =$ 80%	$\theta = 0.2$	$\theta = 0.4$	$\theta = 0.6$	$\theta = 0.8$	
Taylor Swift	2	18	10	5	4	35	16	9	9	26
Maroon V	0	5	4	4	2	19	8	8	8	14
Marathon	0	16	10	7	6	17	11	11	11	15
Tornado	1	3	3	3	2	6	6	6	6	19

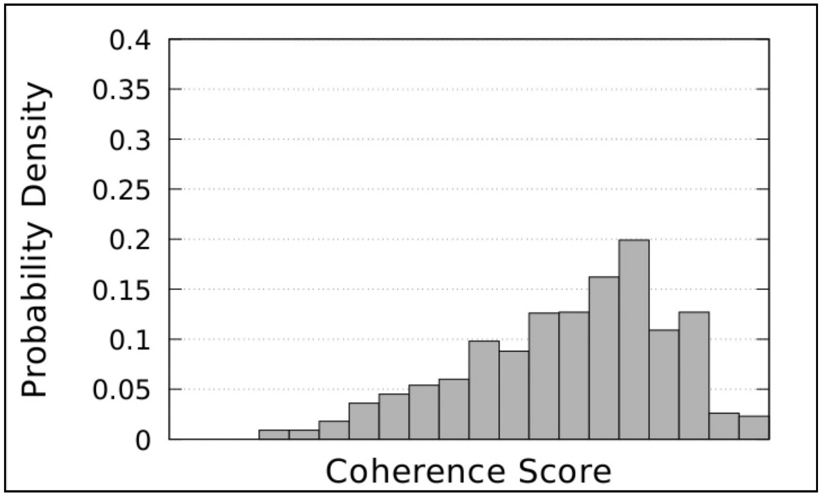
Combined Detection from: Twitter + Instagram

Instagram Location	Instagram Tags	Tweets	Event Signature
(39.045417, -95.721562)	['picket', 'brainwashed', 'westboro', 'protest', 'important', 'wbc', 'truth', 'spreadtheword', 'westborobaptistchurch', 'true', 'dontworrybehappy']	(1) you realize christians protest westboro baptists right is wrong (2) westboro baptist church really protest gunderson production laramie project put years ago (3) fisher westboro protest offers gunderson students opportunity show grizzly pride	(westboro, protest)
(37.7870288, -122.407553)	['protest', 'themission', 'gentrification', 'valenciacorridor', 'googlebus', 'displacement']	(1) laylamrazavi el desalojo ya basta protest googlebus displacement gentrification valenciacorridor (2) video tech workers displaced googlebus protest catch another bus (3) tech buses blocked 45 minutes 2 yrs amp 2 months 1st googlebus protest sfbos sfmayor sb50	(googlebus, protest)

Combined Detection from: Twitter + Instagram



Non events



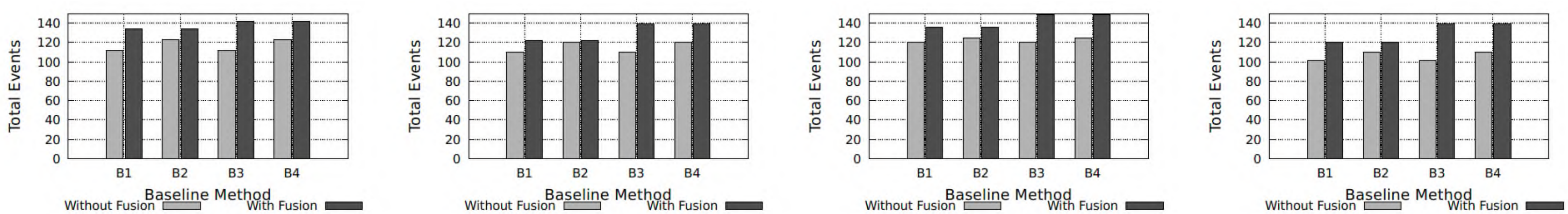
Events

$$P(C|R=1) = B(\alpha 1, \beta 1, C)$$
$$P(C|R=0) = B(\alpha 2, \beta 2, C)$$

$$P(R=1|C,L) = \frac{P(L|R=1)P(C|R=1)P(R=1)}{\sum P(L|R)P(C|R)P(R)}$$

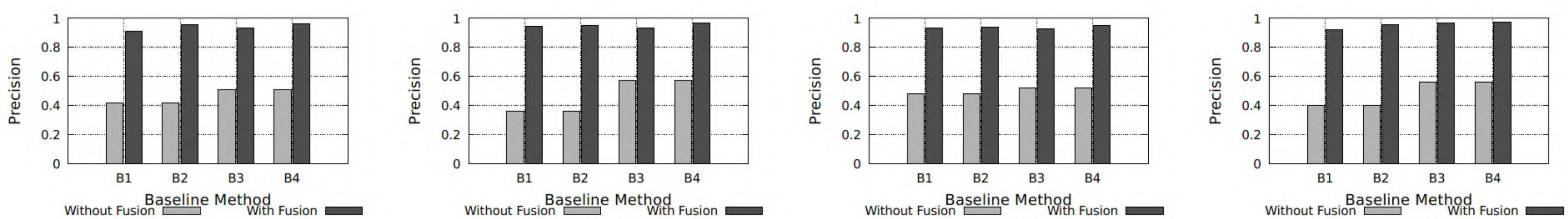
Combined Detection from: Twitter + Instagram

- Fewer false-positives than Twitter
- Better recall than Instagram



(a) Week 1 (b) Week 2 (c) Week 3 (d) Week 4

Instagram detection improvement - comparison of different baseline methods with and without fusion method

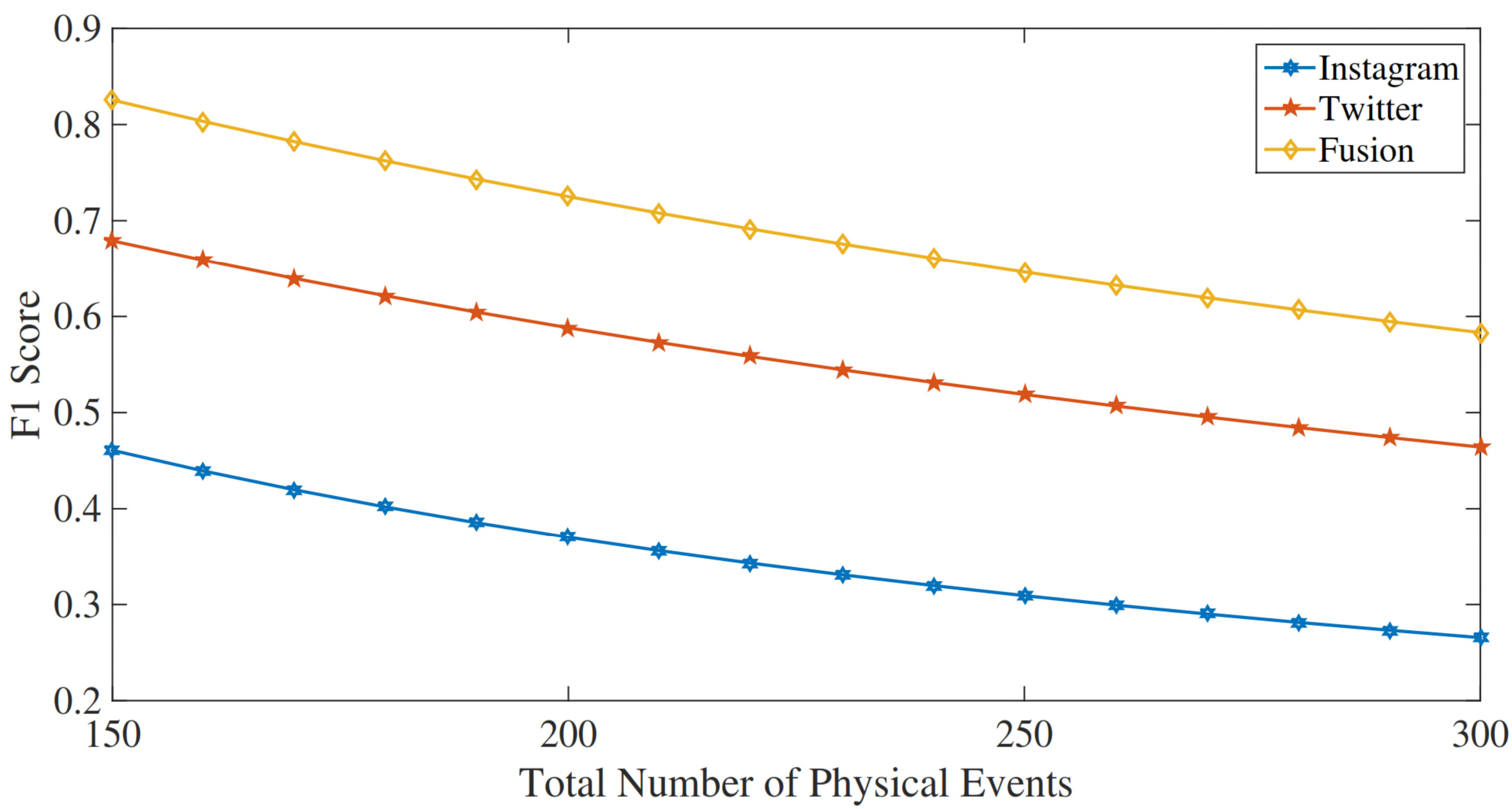


(a) Week 1 (b) Week 2 (c) Week 3 (d) Week 4

Twitter detection precision - comparison of different baseline methods with and without fusion method

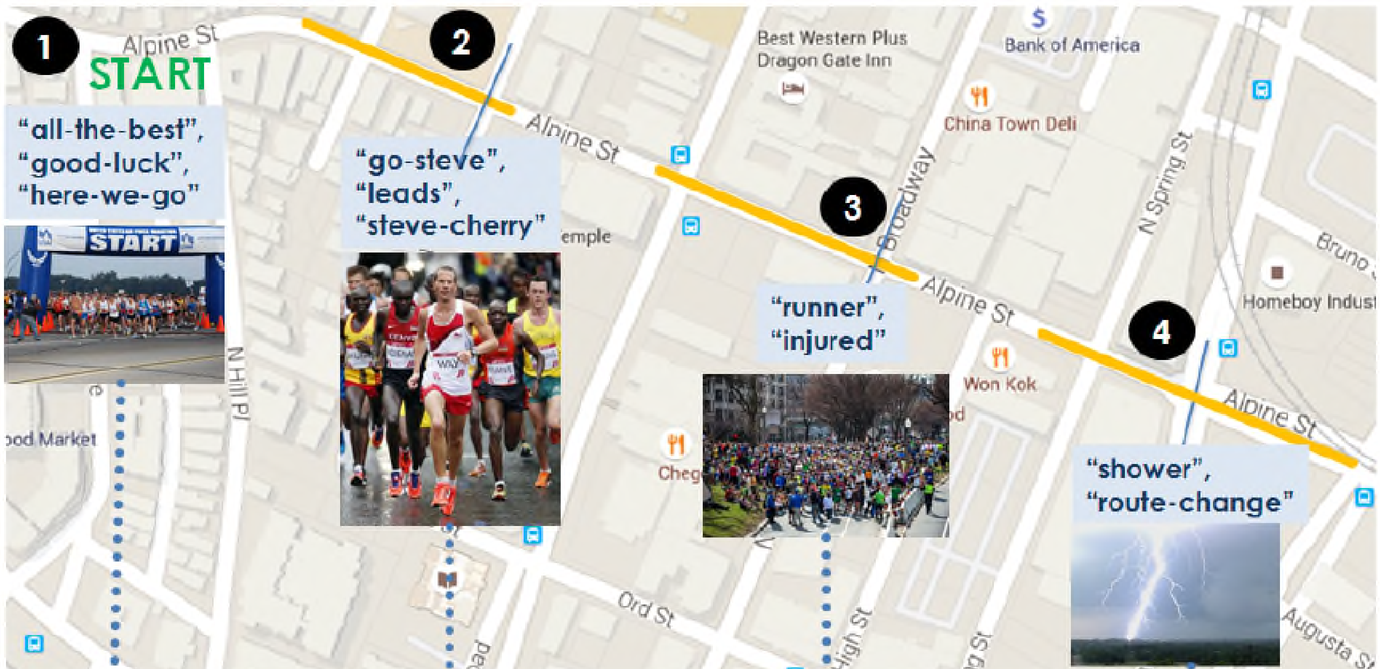
Better Trade-off between Precision and Recall

- F1 score



Detection of Micro Events

MULTI-MODAL INFORMATION



INFERRED TRAJECTORY

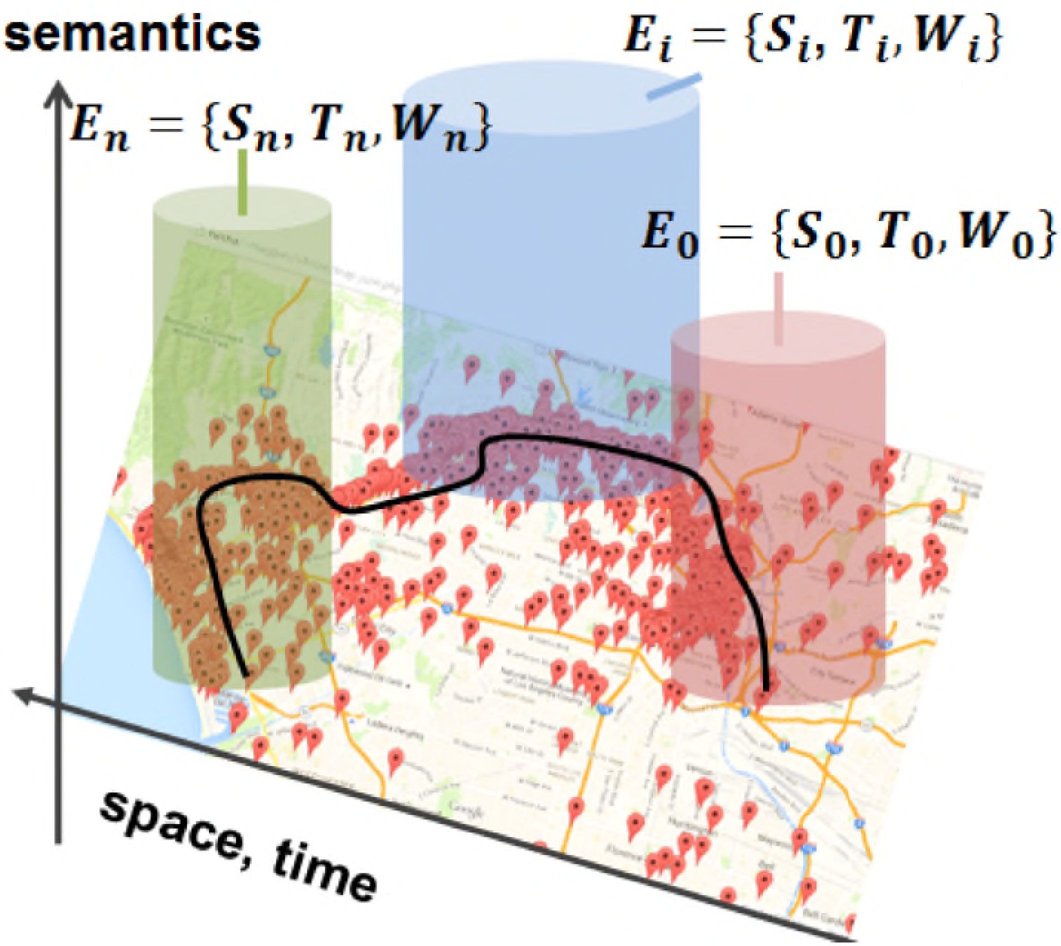
Race starts at N Grand – Alpine at 6 AM → Steve takes lead at Alpine – Yale at 8 AM → Congestion at intersection due to injury at 10 AM → Thunderstorm floods path – change of route at 11 AM

EXPECTED TRAJECTORY

Race starts at N Grand – Alpine at 6 AM → Race progresses (minor incidents are expected) → Race progresses (minor incidents are expected) → Race progresses till completion at the Finish Line

Event Representation

- A series of posts, each with location, time, and semantic tags



What are Key Challenges?

What are Key Challenges?

- Boundaries vary by event (different temporal and spatial scales)
- Many posts are not relevant
- Image semantics are important
- Credibility varies

Approach

Approach

- Apply relevance filter
- Extract semantic labels for images
- Represent the bag of word by a vector
- Cluster features (time, location, and word vector) by similarity

Data

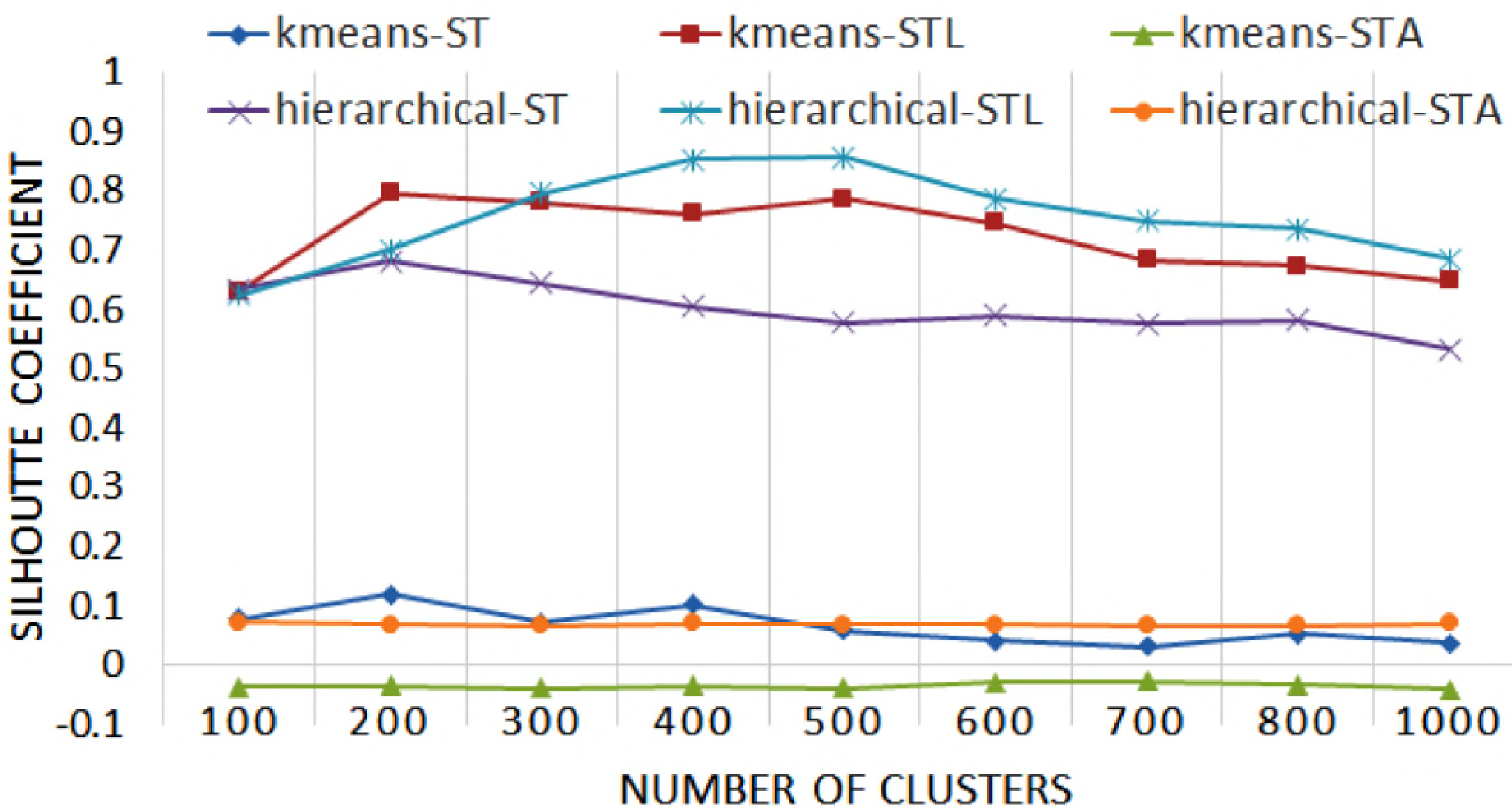
Category	Observation Period	Keywords Used for Filtering
Marathon Events		
Boston	April 2015	“bostonmarathon”, “baa”
LA	March 2015	“lamarathon”
London	April 2015	“londonmarathon”, “vmlm”
Other Event		
F1 SGP	September 2015	“f1 ”,“race(s) ”,“racing ”, “formula”
Non-Events		
Food	September 2015	“food”,“yummy”, “recipe”, “delicious”
Pets	September 2015	“dog(s) ”, “cat(s) ”,“puppy ”,“doggy”
Fashion	September 2015	“fashion”, “style”,“trend”,“outfit”
Selfies	September 2015	“selfie”,“friend”,“fun”

Image Relevance

- How to tell if an image is pertinent to the event?

Clustering

- Quality of clustering



Results

- Error in time and location for known micro-events

Micro-Event	Apprx. Location	Apprx. Time	Location Error in *km* *ST+LDA*	*STA*	Time Error in *mins* *ST+LDA*	*STA*
Boston						
Cheering	Hopkinton	8:50 - 11:15	5.22 (0.066)	13.03 (0.151)	66.93	68
Winners	Public Library	11.56 - 12:09	5.49 (0.062)	3.89 (0.047)	50.03	332.18
London						
Start of race	Greenwich Park	9:00 - 10:10	0.79 (0.027)	5.77 (0.079)	31.37	185.55
Winners	The Mall	11:43 - 12:14	2.70 (0.007)	5.44 (0.071)	57.03	56.46
LA						
Cheering	Dodger Stadium	6:30 - 6:55	3.62 (0.039)	3.88 (0.035)	93.01	99.53
Winners	Santa Monica	9:05 - 9:17	10.48 (0.094)	10.65 (0.098))	333.85	34.24

Quiz

- Which paper(s) on the reading list evaluate the efficacy of using *unlabeled data* to enhance learning/classification accuracy from smartphone sensor data?

- Paper:
- Purpose (Application):