# Measuring Geographical Regularities of Crowd Behaviors for Twitter-based Geo-social Event Detection

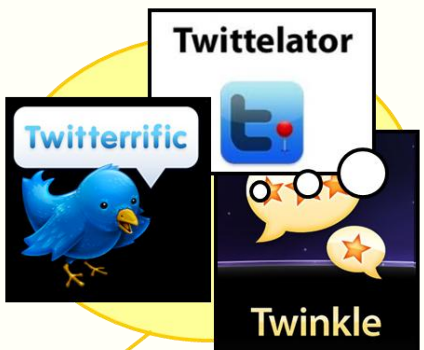
Ryong Lee and Kazutoshi Sumiya
University of Hyogo, Japan

# Outline

- Motivation: When Twitter Met LBS
- Geo-social Events Detection from Twitter
  - Geographic Regularity
  - Collecting Massive Geo-Tweets from Twitter
  - Social Geographic Boundary
  - Estimating Geographic Regularities
- Experimental Results
- Conclusions

# When Twitter Met LBS

- Micro-blogging sites ( Twitter, Jaiku, Prownce) as an important media to share info. about geo-social events
- Socio-geographic analysis using Social Network Services for discovery of social / natural events and urban characteristics




Twitter Apps for iPhone


Microblogs about Kobe Luminaria Festival

# Twitter as a Geo-social Database

- **Geo-Tweets are instant updates of people with whereabouts**

twitter +  ➔ 


**Geo-Tweets**
**(Tweets having geo-tags)**

Spatio-Temporal
Crowd Behavior Database

Example of Geo-Tweets

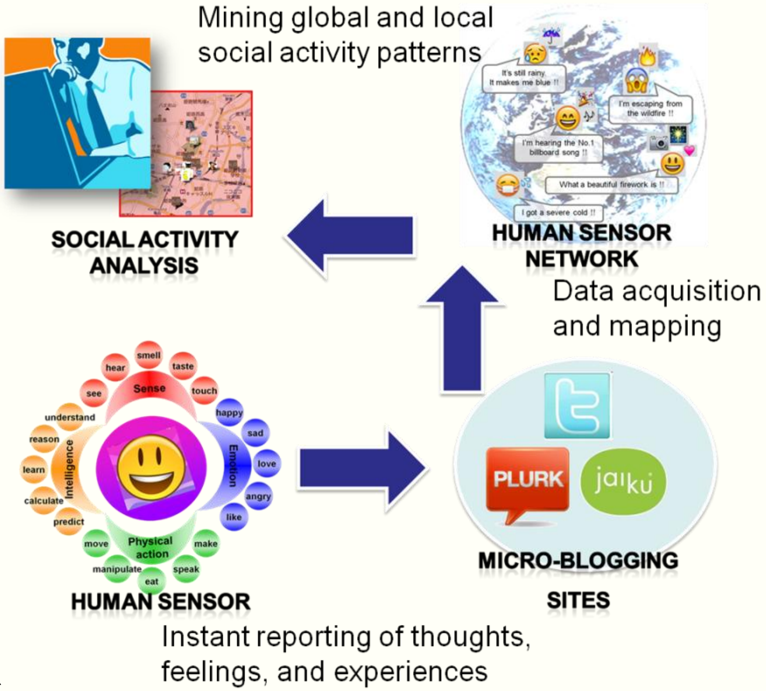
| user_id | created_at | loc_lat | loc_lng | texts |
|---|---|---|---|---|
| 1534****** | Thu, 03 Jun 2010 20:21:25 | 42.327873 | 142.4175637 | Kitami is 11 degrees now. By the way, Osaka seems to be 28 degrees. |
| 5235***** | Thu, 03 Jun 2010 21:16:13 | 42.7424814 | 143.6865067 | Good morning. It is heavy mist. |
| 7143**** | Thu 03 Jun 2010 15:59:41 | 41.939994 | 126.423587 | The soba of this shop seems to be delicious. http://twitpic.com/1tkoua |
| 1513**** | Fri 04 Jun 2010 00:20:5 | 44.0206319 | 144.2733983 | I passed Bihoro. |
| 1537**** | Fri, 04 Jun 2010 00:04:54 | 44.3045224 | 142.6389133 | The rain falls today. |

Who　　When　　Where　　What/Why/How/…

# Goal: Geo-social Events Detection from Twitter

- Challenging Issue: Can we exploit crowds' senses for various socio- geographic analyses?

- What crowds are sharing
  - Simple Types
    - Whereabouts
    - What s/he is doing
  - Complex Types (by analysis)
    - What they are doing together
    - What they are thinking or feeling
    - How the local/global societies work



# Detecting Local Events from Crowd Behavior

- The 1st idea: counting geo-tweets for every town
  →But, downtowns always have many tweets

- What Local Events we'd like to find out?
  →Not every-day activities such as commutes, but unusually occurring incidents like festivals

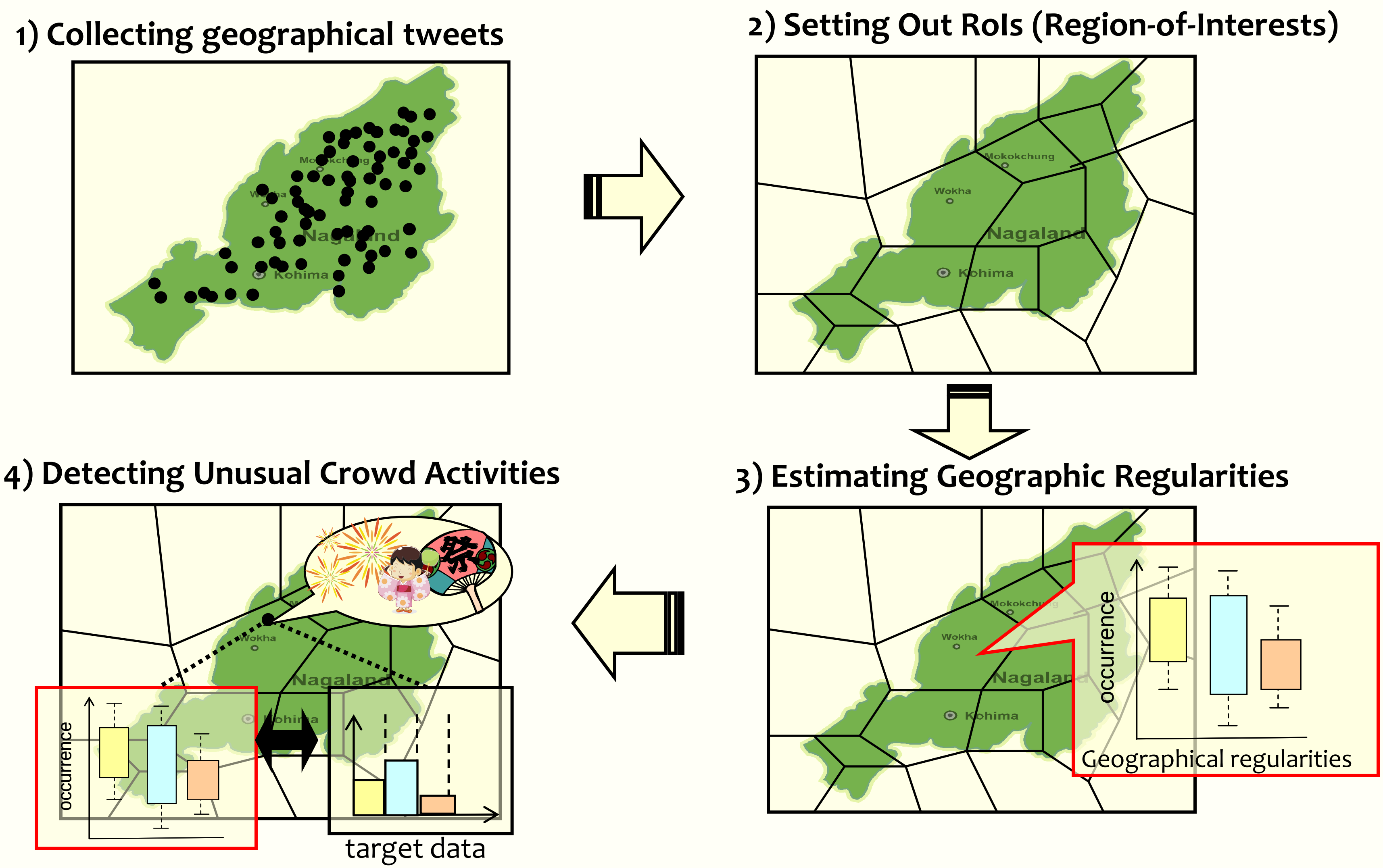
- Regularity vs. Irregularity would be a key to discover interesting and meaningful events, at first

# Geographic Regularity

- **How to define/measure Geographical Regularity** (usual status of crowd behaviors in a region)
    - How many tweets are posted
    - How many users are there
    - How active are the movements of the local crowd
- 'Regular/Irregular' may be a relative concept
    - We need to characterize each region's geographic regularities… (Stations, Office town, bed town, etc.)

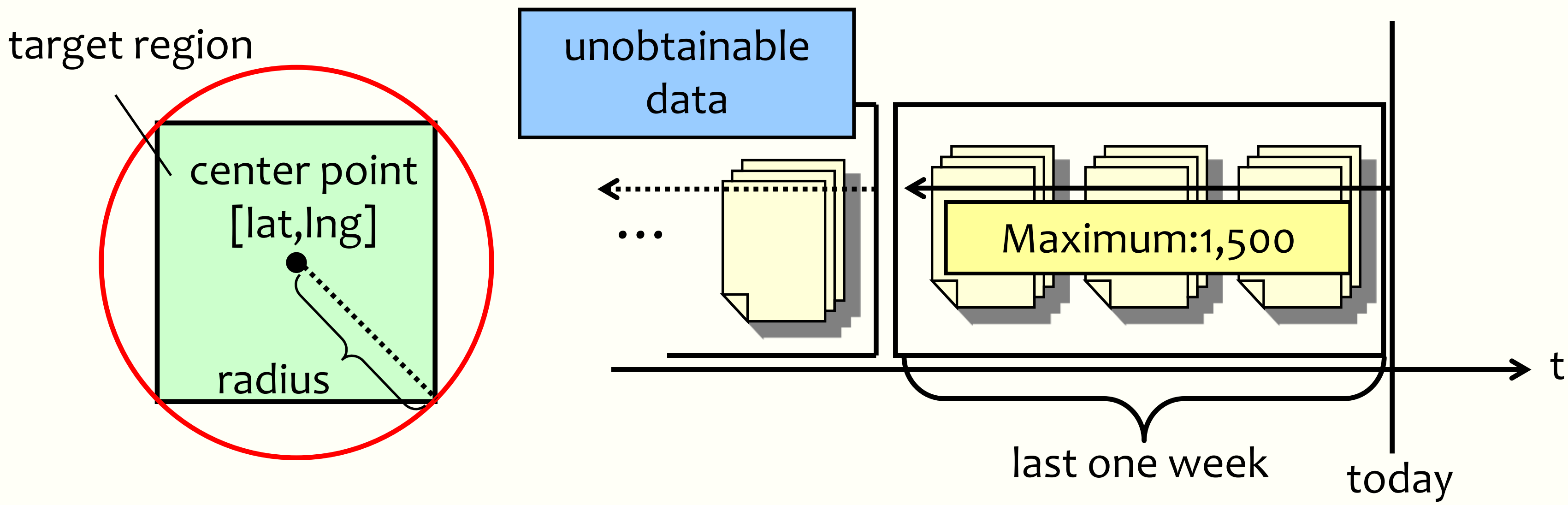
    Generally, a complicated sociological analysis task

# Process of Geo-social Event Detection



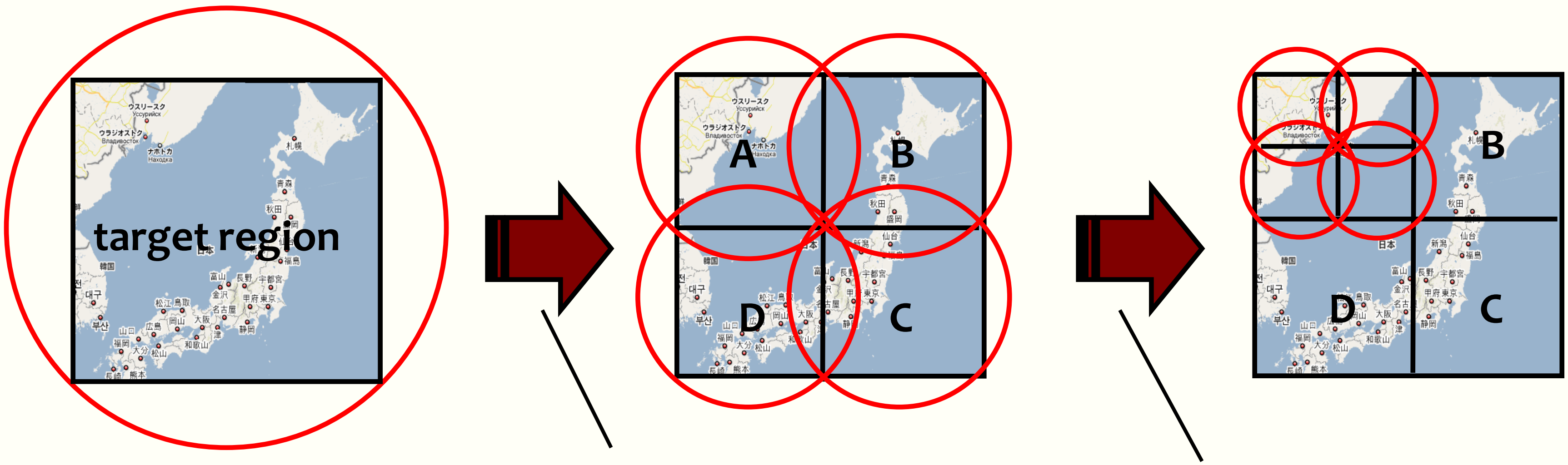
# Collecting Geo-Tweets from Twitter

- Collect Twitter messages (tweets) within a geographic region
- Twitter's GeoAPI supports only 'nearby' query by center location + searching radius
- Limitation: 1,500 tweets/query, up to the past week, 1km ≤ radius ≤ 500km or less



# Collecting Massive Geographical Tweets

- Deployment of querying locations by Quad-tree



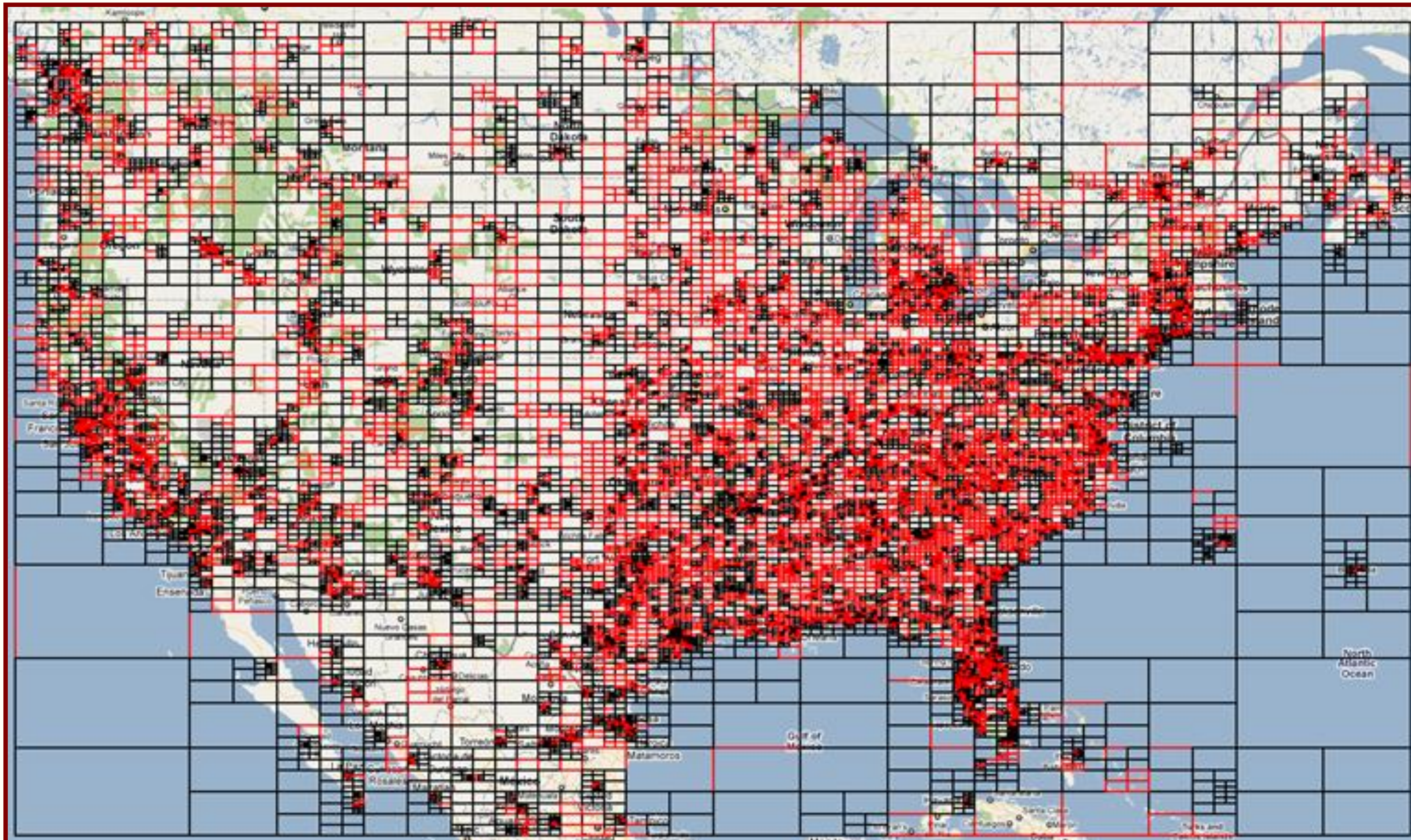

obtain data in the circle surrounding the region

If # message =Max_answer(=1,500)

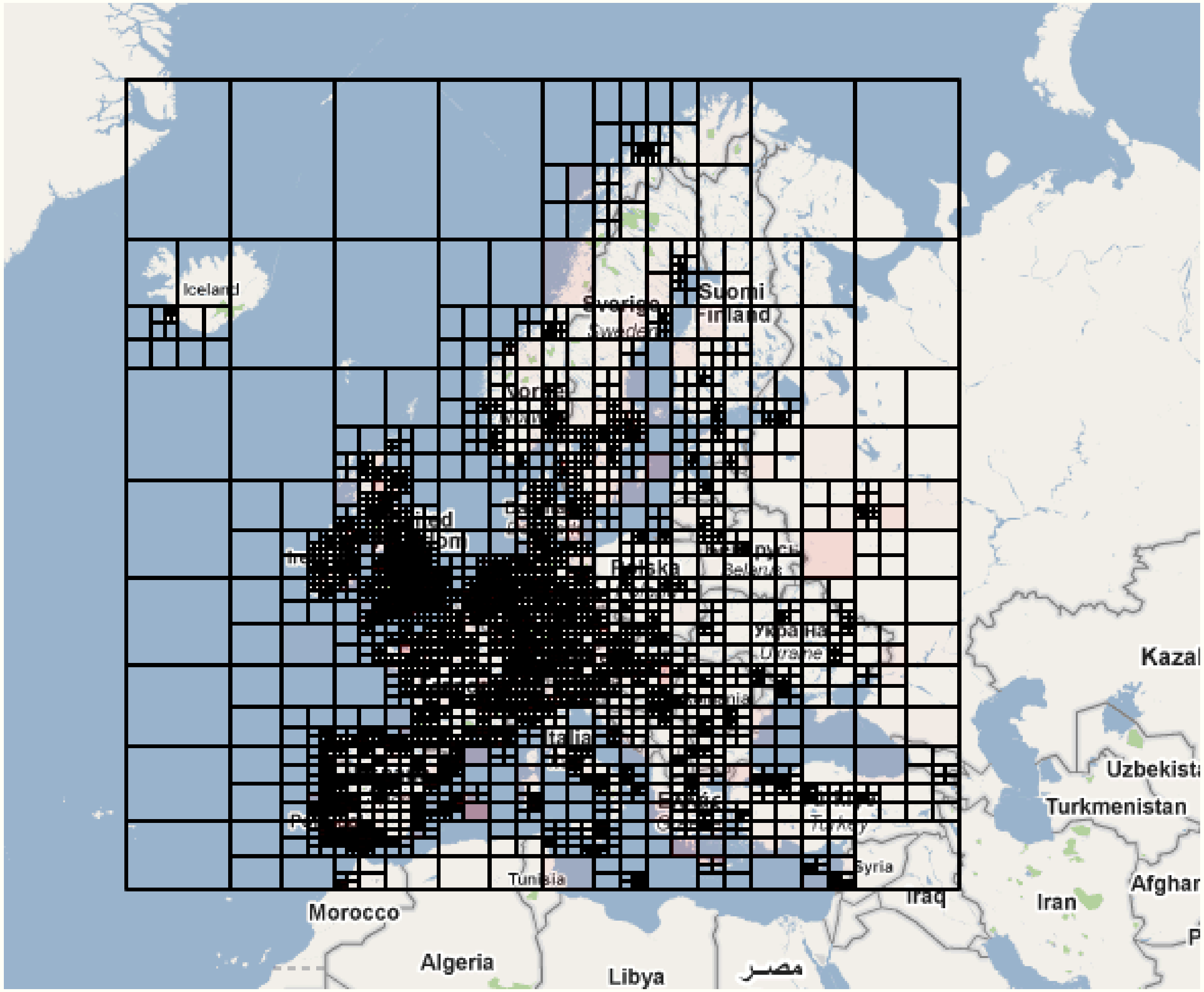
If # message in region A = Max_answer(= 1,500)

By repeating these operations recursively, we realized the <u>acquisition method which depends on quantity of the regional data</u>

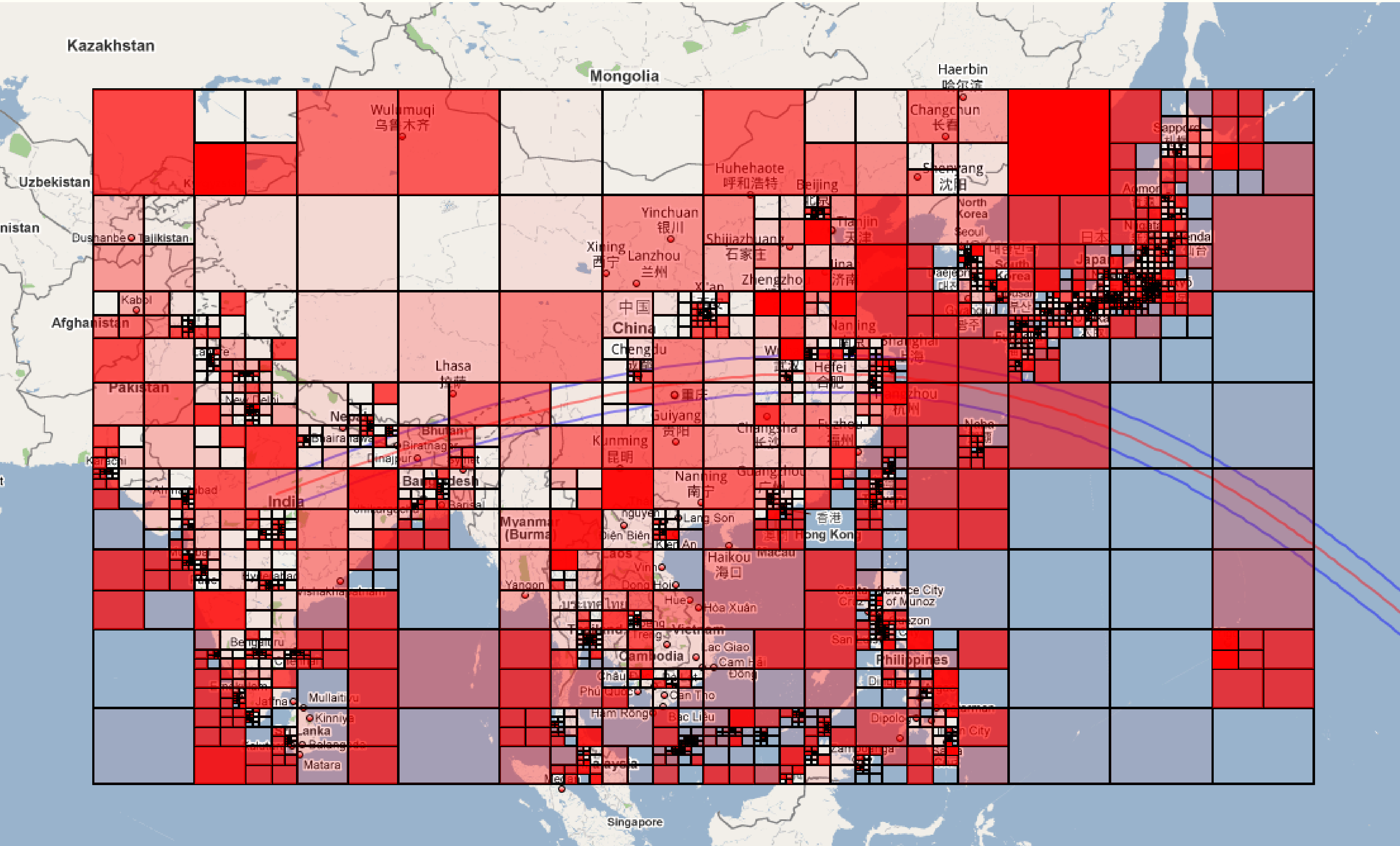
# Geographic Distribution of Geo-Tweets in the USA



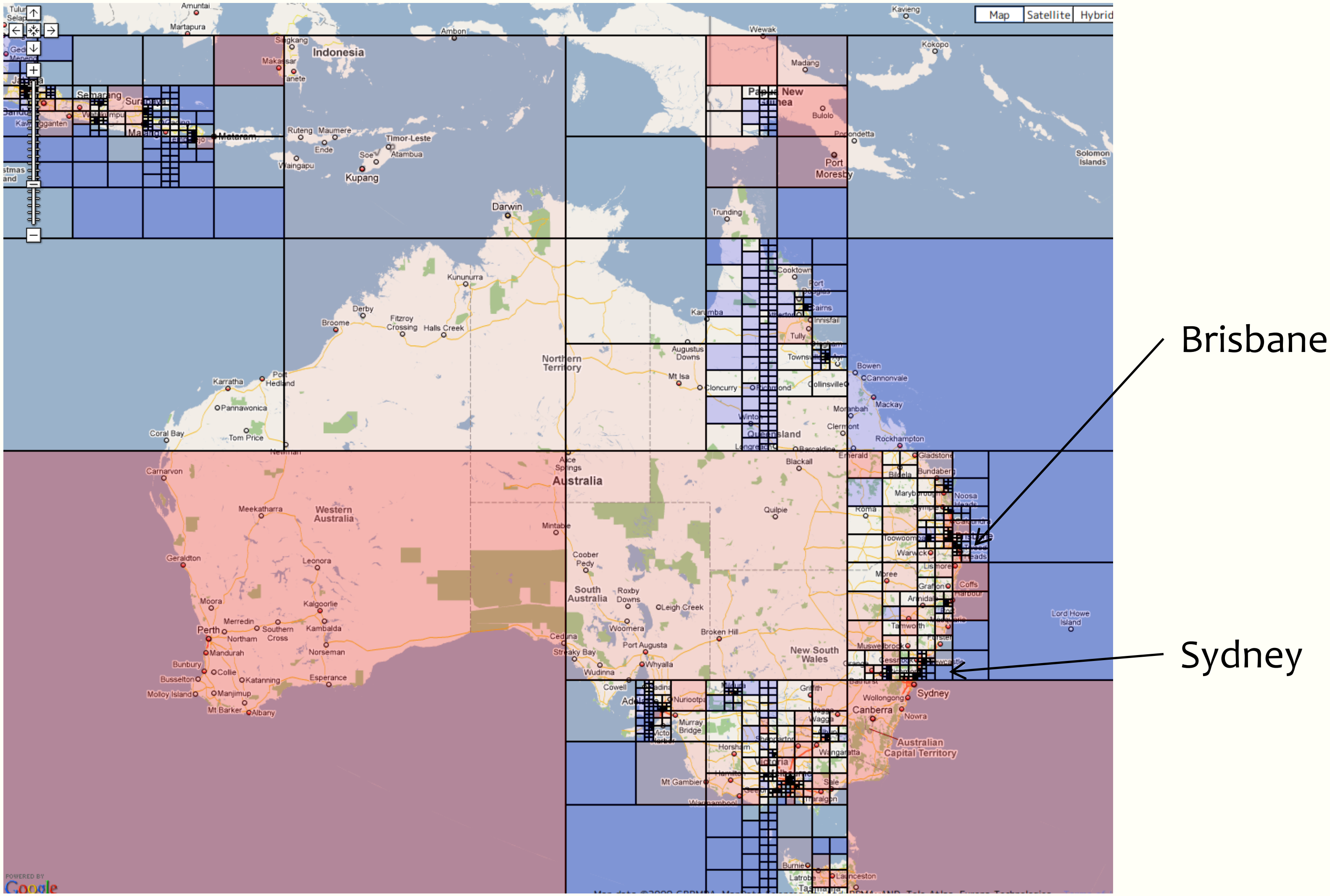
# EU


Iceland
Suomi
Finland
Sverige
Kazal
Uzbekist
Turkmenistan
Syria
Iraq
Iran
Afghar
Morocco
Algeria
Libya
Tunisia
مصر

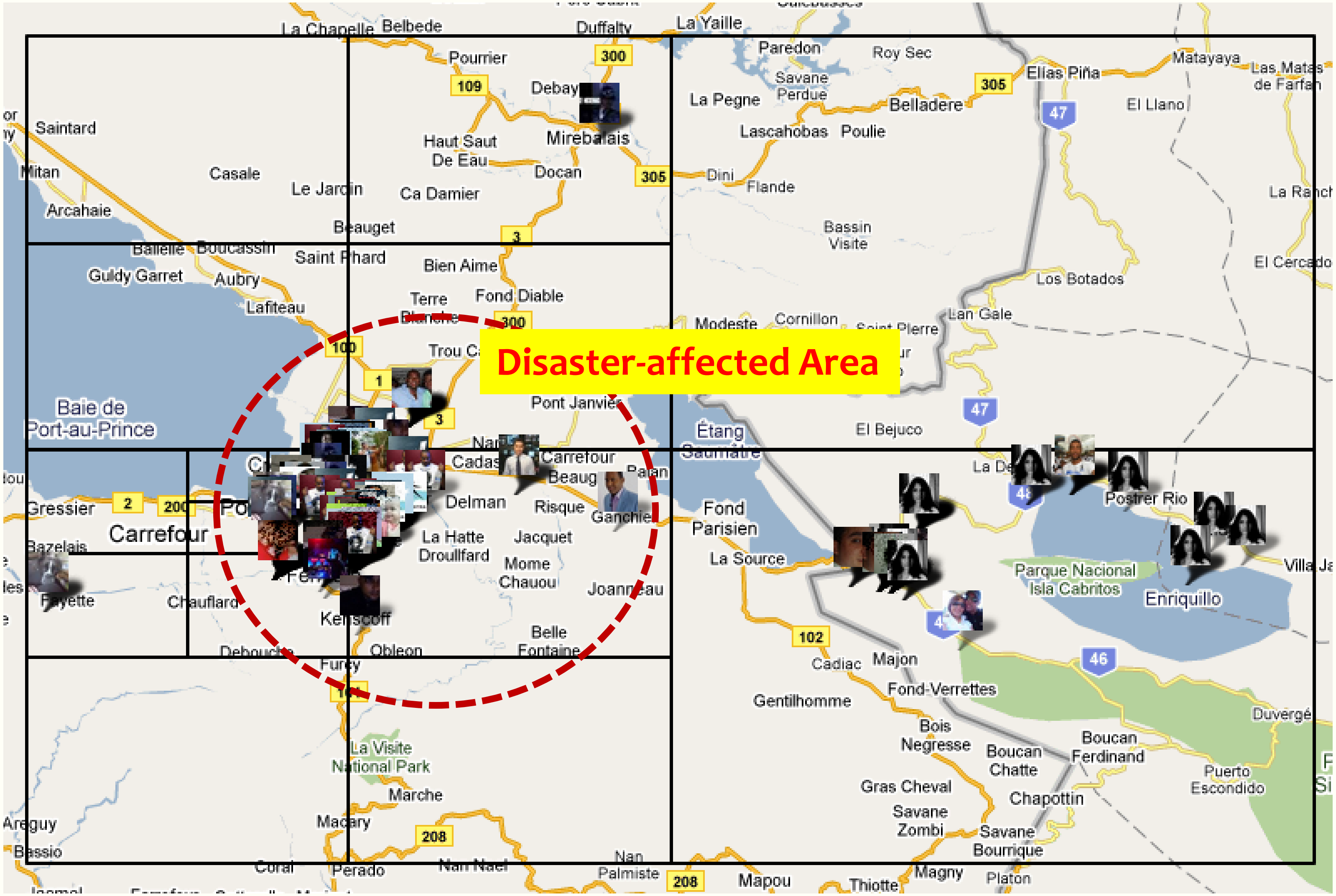
# Asia (before iPhones come to China)



# Australia


Map
Satellite
Hybrid
Indonesia
Papua New Guinea
Port Moresby
Darwin
Northern Territory
Western Australia
Australia
Queensland
South Australia
New South Wales
Perth
Canberra
Australian Capital Territory
Lord Howe Island
Brisbane
Sydney

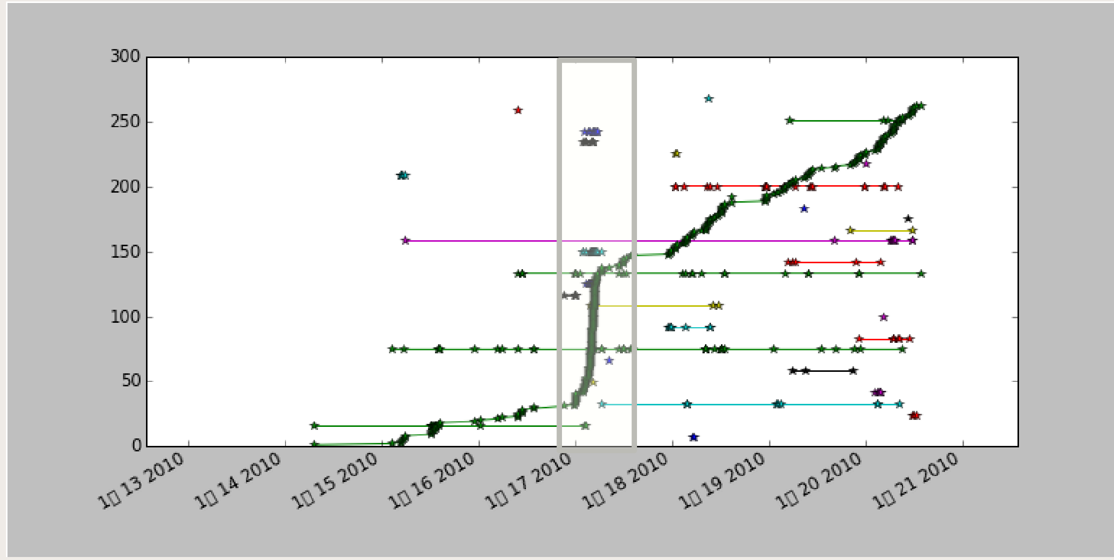
# In Haiti?

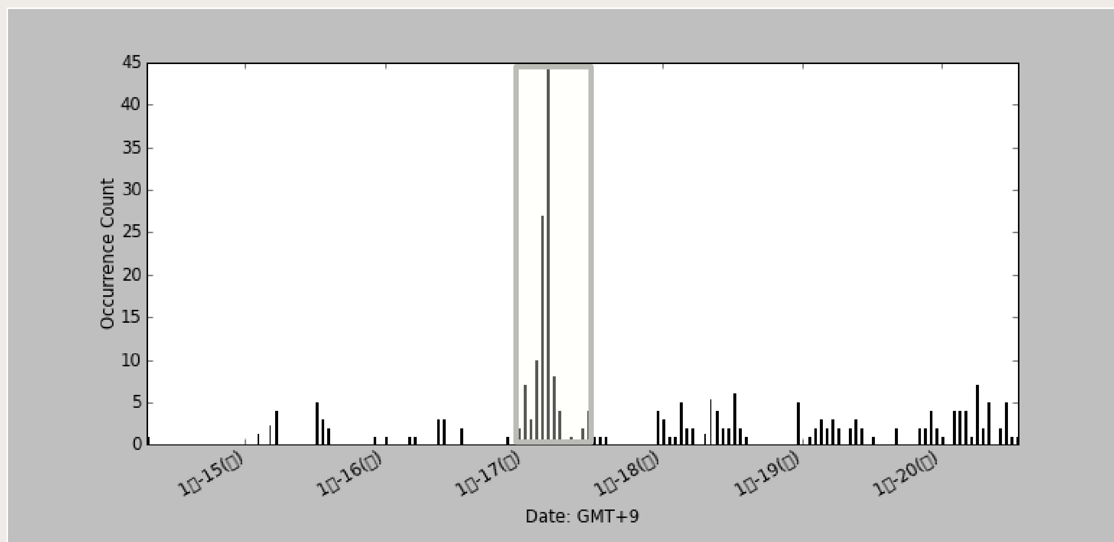



(*displayed only the most recent 300 tweets)

# Crowd Interests by Term Frequency

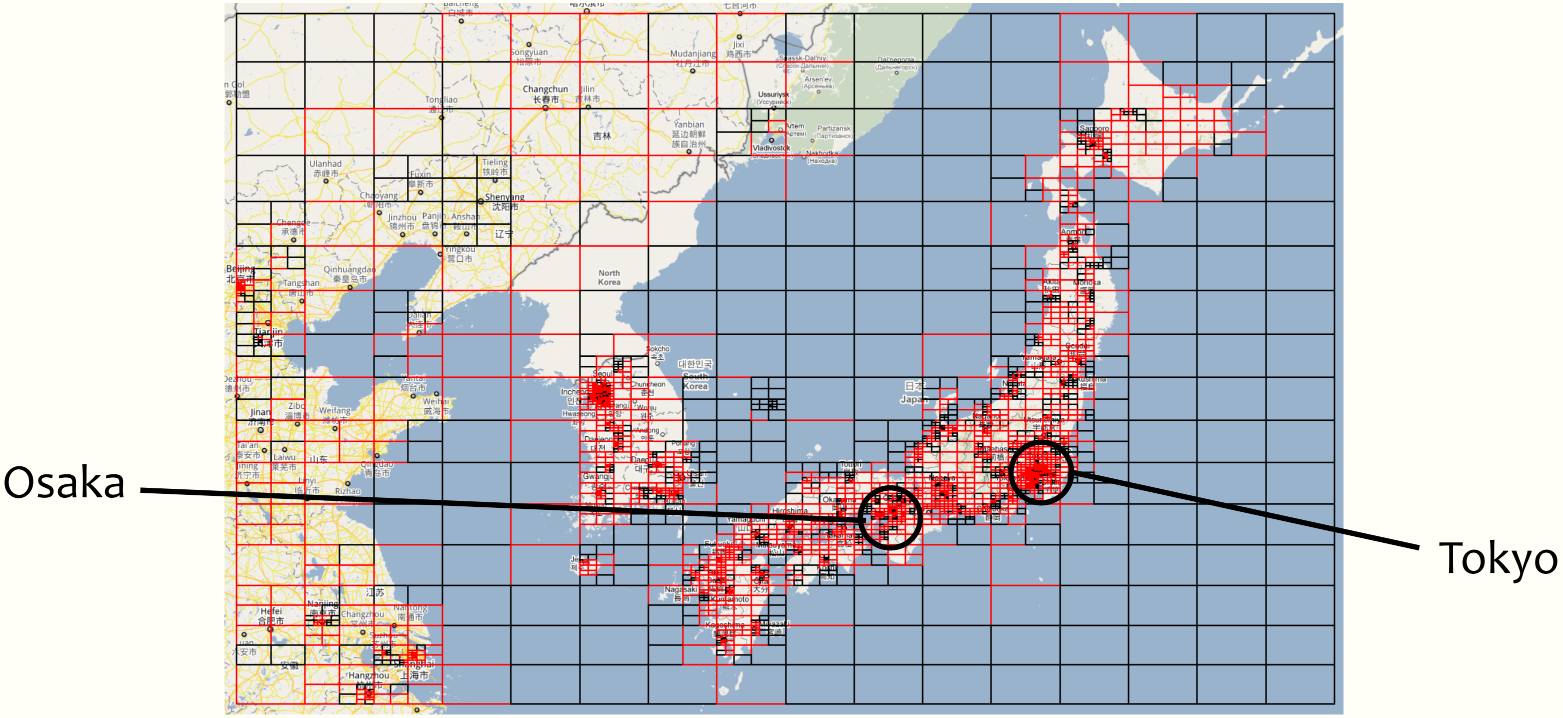
| Jan./14-21 | | Jan. 17th (0:00-12:00) | |
|---|---|---|---|
| haiti | 11.5 | petionville | 2 |
| pap | 3.5 | smoothed | 1 |
| petionville | 3 | glory | 1 |
| paup | 3 | aftershock | 1 |
| jacmel | 3 | god | 0.6 |
| border | 3 | words | 0.5 |
| aftershock | 3 | water | 0.5 |
| tia | 2.66667 | reconnect | 0.5 |
| home | 2.6 | pierre | 0.5 |
| help | 2.6 | orphanage | 0.5 |
| embassy | 2.5 | haiti | 0.5 |
| unicef | 2 | firemen | 0.5 |
| tentes | 2 | distribute | 0.5 |
| supplies | 2 | thanks | 0.4 |
| sabine | 2 | glad | 0.4 |
| preval | 2 | weapons | 0.33333 |
| people | 2 | weapon | 0.33333 |
| padf | 2 | tia | 0.33333 |
| oas | 2 | semi | 0.33333 |
| luckner | 2 | patrol | 0.33333 |
| jimani | 2 | papa | 0.33333 |
| haitian | 2 | mountains | 0.33333 |
| food | 2 | heavily | 0.33333 |
| delievered | 2 | encouragement | 0.33333 |
| bourdon | 2 | describe | 0.33333 |
| dr | 1.75 | bucket | 0.33333 |
| earthquake | 1.66667 | blessings | 0.33333 |
| water | 1.5 | automatic | 0.33333 |
| praying | 1.5 | auto | 0.33333 |
| need | 1.4 | armed | 0.33333 |



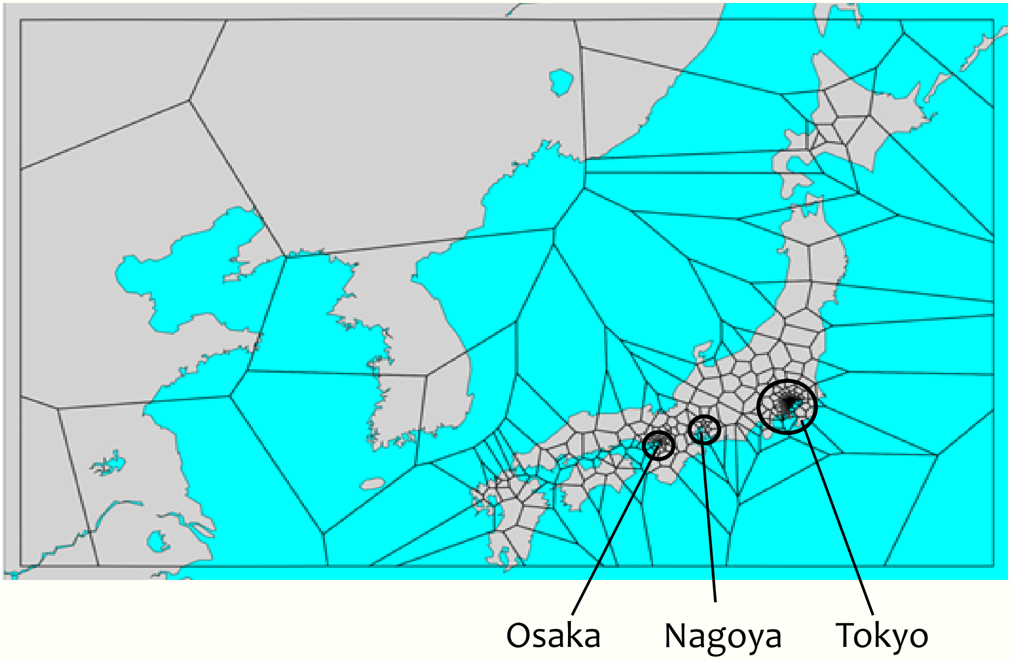
# Experimental Data

- Geo-Tweets found around Japan
  - Date: 2010/06/04-2010/07/24
  - Geographical tweets : 21,623,947 (geo-tagged)
  - Users : 366,556



# Social Geographic Boundary

- Set out RoIs (Region-of-interests) for estimating geographical regularities
  - RoIs: Partitioned sub-areas are used for monitoring
- Space partitioning for estimating geographical regularities of each region

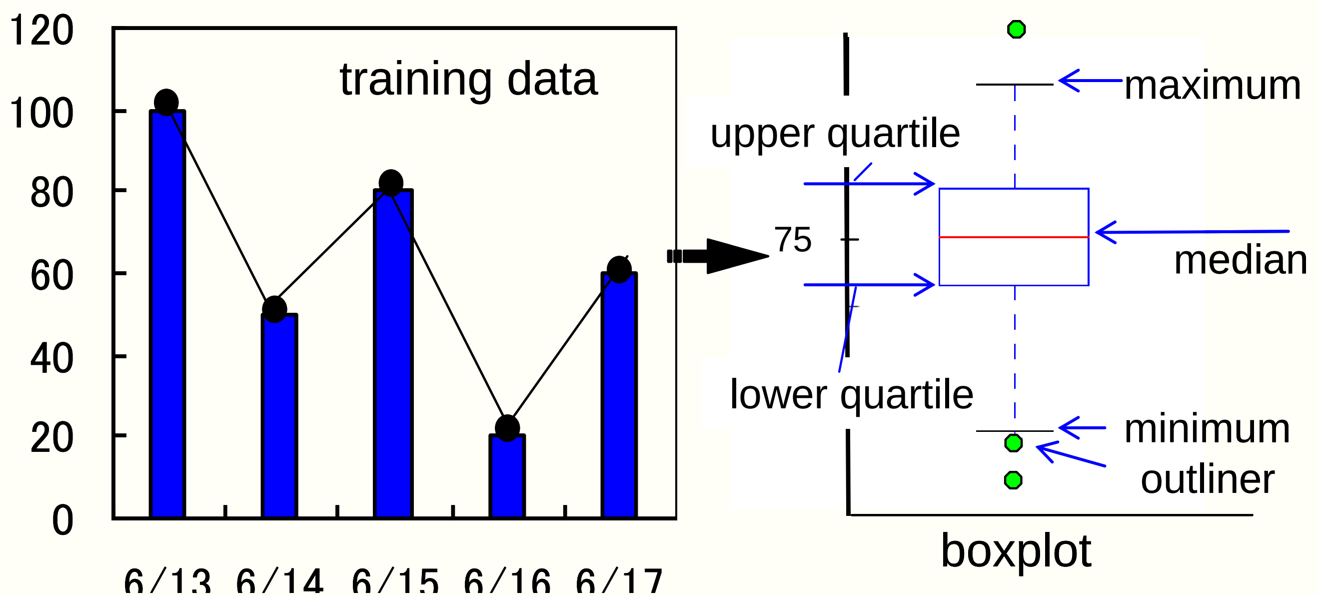



Voronoi-based Space Partition (after K-Means Clustering),

K=300

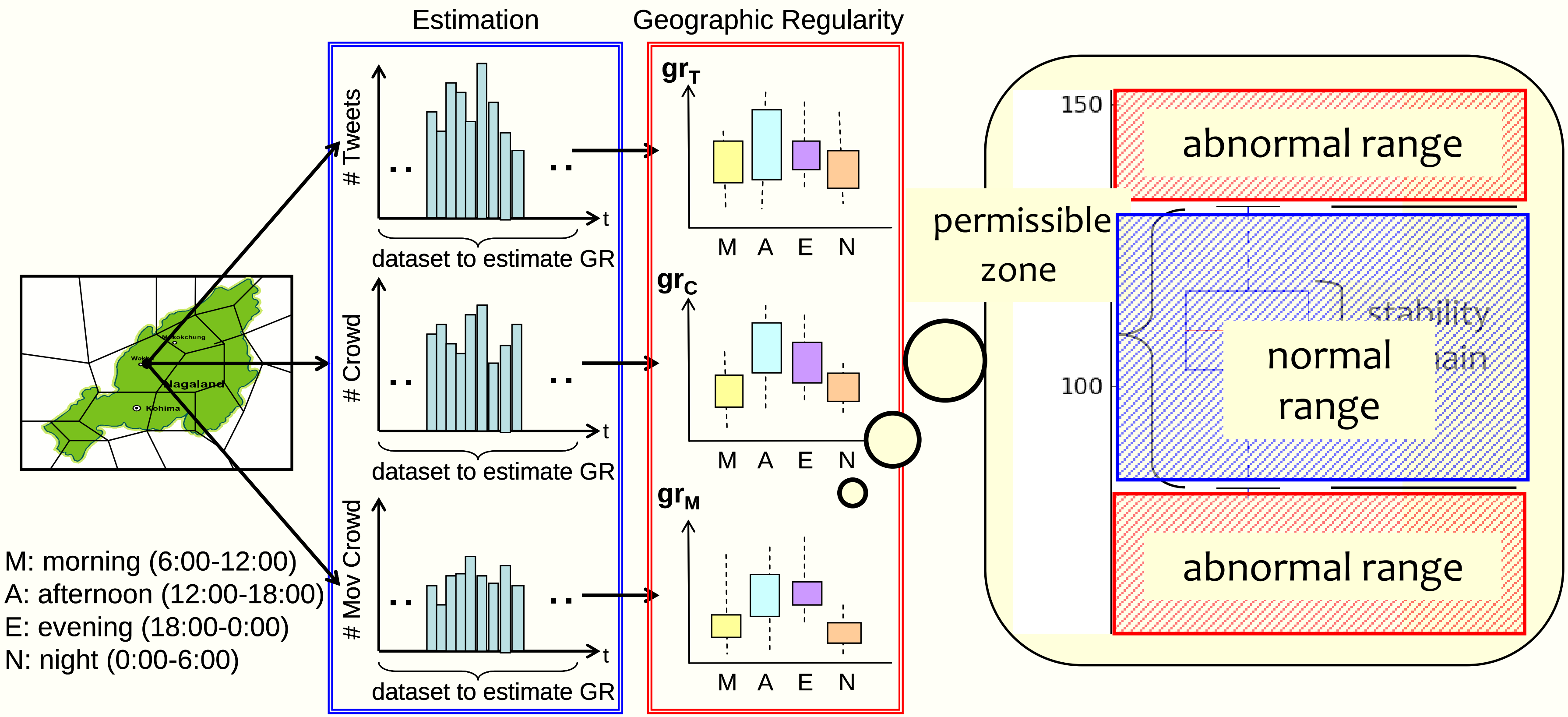
# Estimating Geographical Regularities (1/2)

- RoI's geographical regularity (gr) based on <u>three indicators</u>
  - **#Tweets** : the number of tweets that were written in an RoI
  - **#Crowd** : the number of Twitter users found in an RoI
  - **#MovCrowd** : the number of moving users related to an RoI
- Estimating by a statistical manner using boxplot
  - a boxplot: presentation of five sample statistics (the minimum, the lower quartile, the median, the upper quartile, and the maximum)



# Estimating Geographical Regularities (2/2)

Decide a normal/abnormal range using boxplots




Detection of unusually crowded region by using combinational cases of #Tweets, #Crowd, and #MovCrowd

# Decision of Detection Conditions

- Combination of three indicators

| | #Tweets | #Crowds | #MovCrowds | Final Decision (F) |
|---|---|---|---|---|
| (a) | N | N | N | N |
| (b) | | | A | N |
| (c) | | A | N | N |
| (d) | | | A | A |
| (e) | A | N | N | N |
| (f) | | | A | A |
| (g) | | A | N | N |
| (h) | | | A | A |

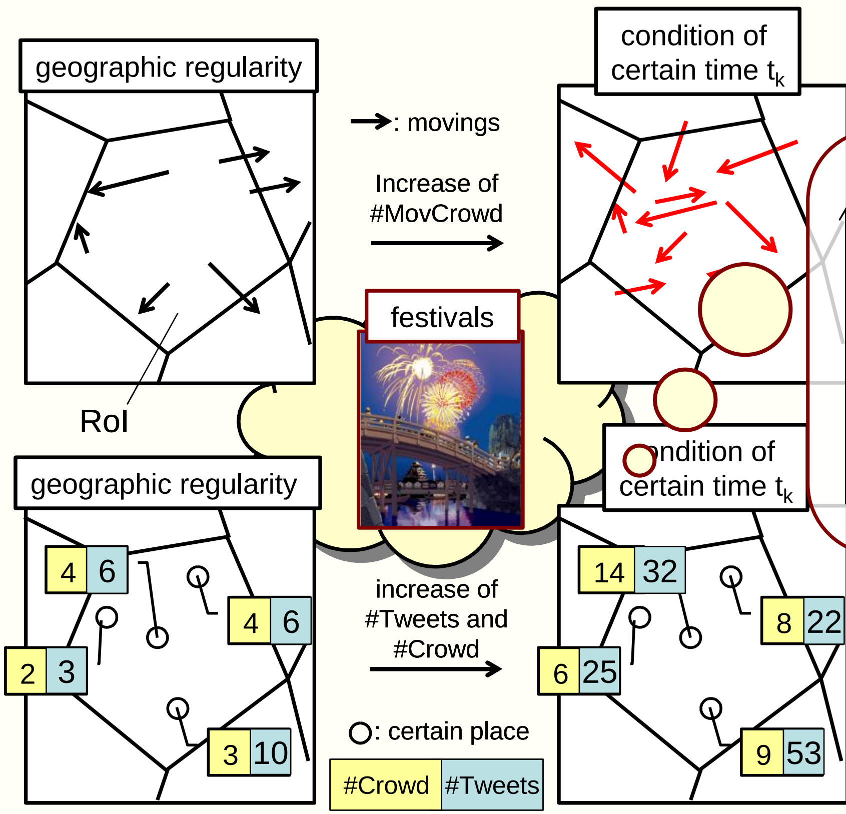
N: normal A: abnormal

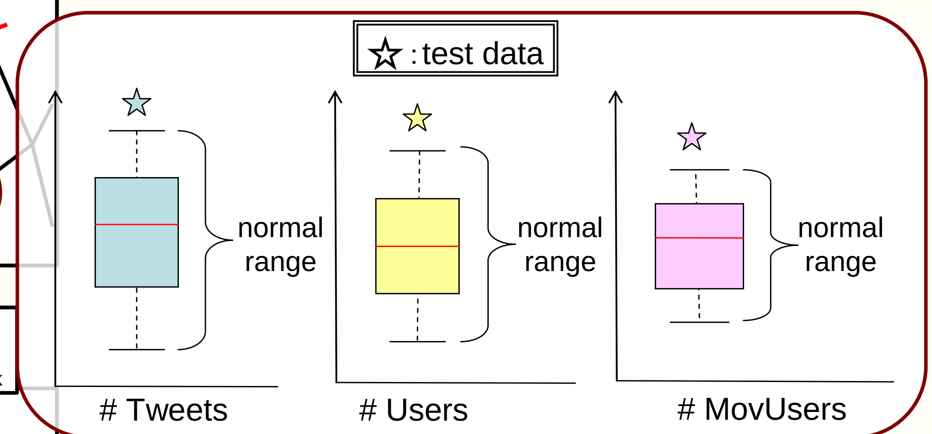
- Combinational cases of abnormal status
  - (h): all the indicators show an abnormality
  - (f): #Tweets and #MovCrowd only show an abnormality
  - (d): #Crowd and #MovCrowd only show an abnormality

# Detection of Unusual Events (1/3)

h) **all the indicators →** abnormal

ex. Tendency of big event or famous festival

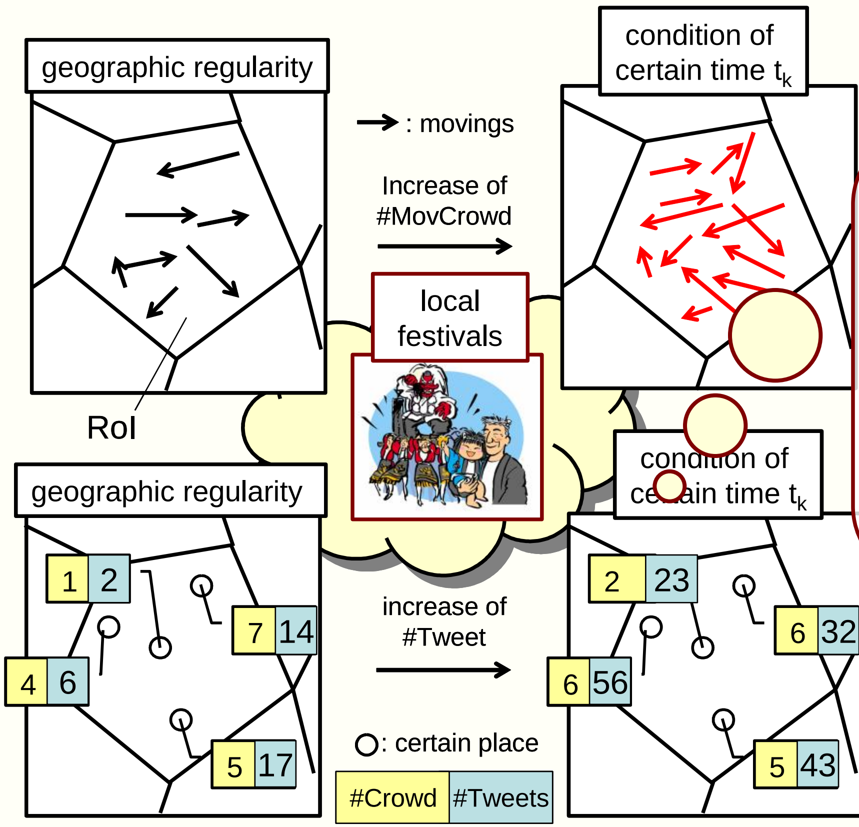


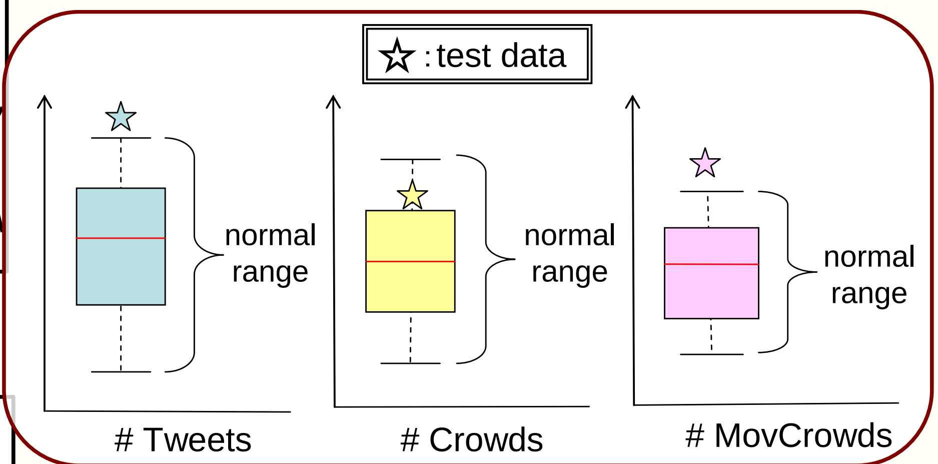




| #Tweets | #Crowds | #MovCrowds | Final Decision (F) |
|---|---|---|---|
| A | A | A | A |

# Detection of Unusual Events (2/3)

f) **#Tweets** and **#MovCrowds** → abnormal
ex. : Local small festival



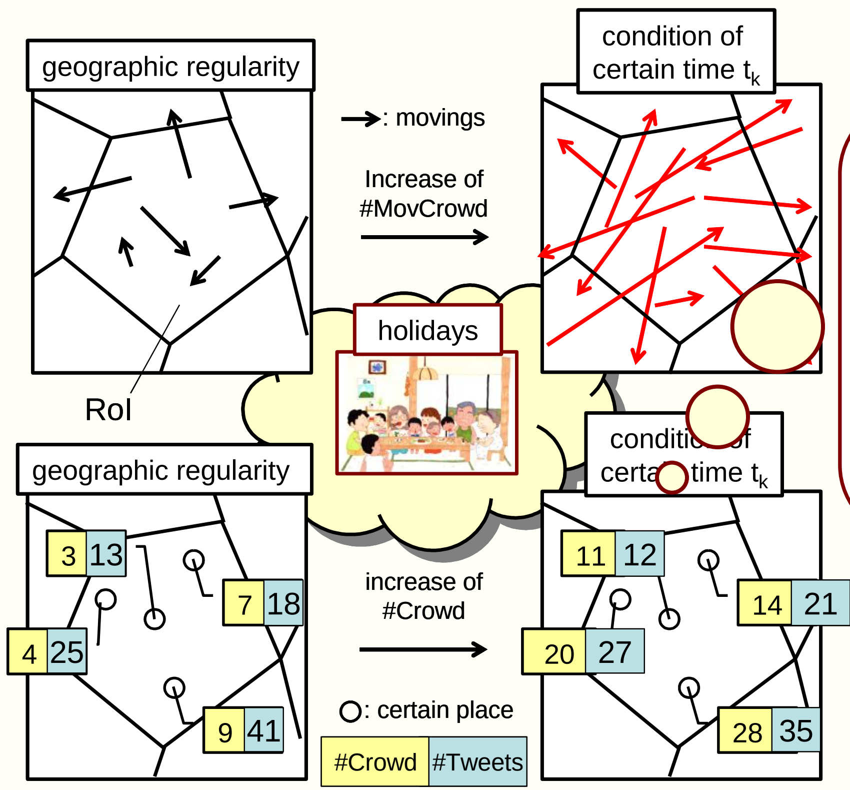



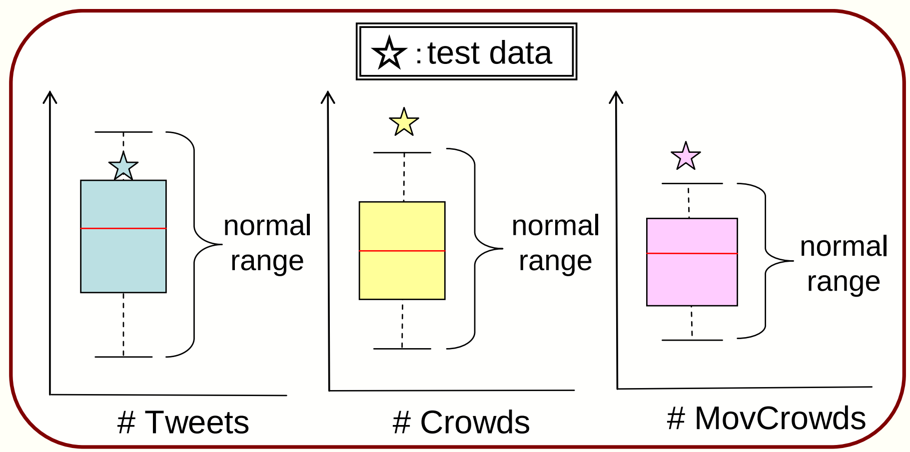

| #Tweets | #Crowds | #MovCrowds | Final Decision (F) |
|---|---|---|---|
| A | N | A | A |

# Detection of Unusual Events (3/3)

g) **#Crowd** and **#MovCrowds** → abnormal

ex. Long holidays




| #Tweets | #Crowds | #MovCrowds | Final Decision (F) |
|---|---|---|---|
| N | A | A | A |

# Experiment

- Experimental purpose
  - Validate our geo-social event detection method
  - Test how many town festivals in Japan would be found by our proposed method

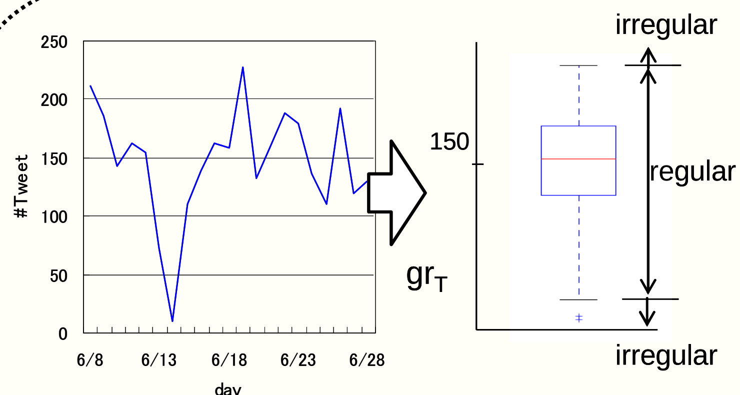
**Town festivals held in Japan for 7/17–7/19, 2010**

| No. | Event name | Place | Event day(s) |
|---|---|---|---|
| 1 | Kyoto Gion Festival | Sakyo, Kyoto, Kyoto | 7/17 |
| 2 | Shishikui Gion Festival | Shisikui, Kaiyoumachi, Tokushima | 7/17, 7/18 |
| 3 | Towada Kosui Festival | Towada, Aomori | 7/17 |
| 4 | Tamamura Firework | Tamura, Gunmu | 7/17 |
| 5 | Ise Firework | Nakajima, Ise, Mie | 7/17 |
| 6 | Akiyoshi Firework | Syoho, Mine, Yanaguchi | 7/17 |
| 7 | Kanonji Festival | Kannonji, Kagawa | 7/17, 7/18 |
| 8 | Muroto Festival | Muroto, Kouchi | 7/18 |
| 9 | Sanoyoi Carnival | Arao, Kumamoto | 7/18 |
| 10 | Uminohi Festival in Nagoya | Nagoya, Aichi | 7/19 |
| 11 | Housui Festival | Noboribetsu, Hokkaido | 7/17 |
| 12 | Shinmatsudo Festival | Matudo, Chiba | 7/17, 7/18 |
| 13 | Nanao Festival | Nanao, Ishikawa | 7/17 |
| 14 | Oota Festival | Oota, Gunma | 7/17 |
| 15 | Toukou Festival | Arita, Wakayama | 7/18 |

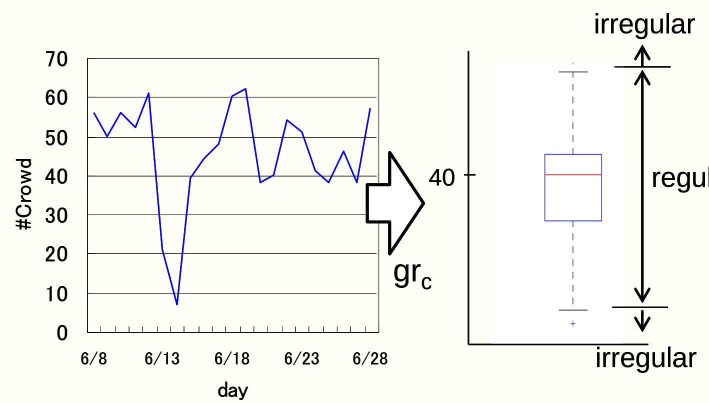
# Estimating Geographical Regularities

Estimation about three indicators (#Tweets, #Crowds, #MovCrowds) for <u>every fixed time slot</u>

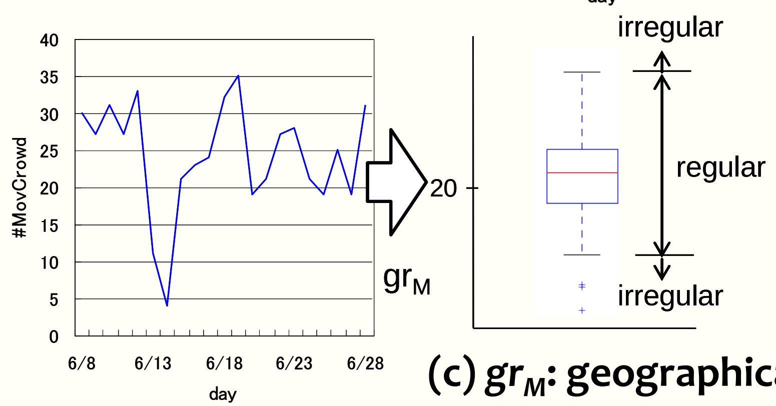
- split a day into four equal time slots—morning, afternoon, evening, and night— for the period of 6/8-6/30




(a) $gr_T$ : geographical regularity of #Tweet




(b) $gr_C$: geographical regularity of #Crowd

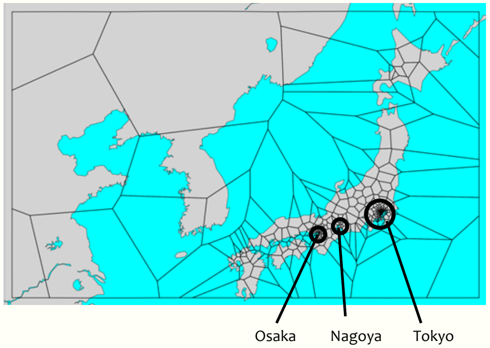



(c) $gr_M$: geographical regularity of #MovCrowd

Geographical regularities based on #Tweets, #Crowds, and # MovCrowds of Kyoto

# Evaluation

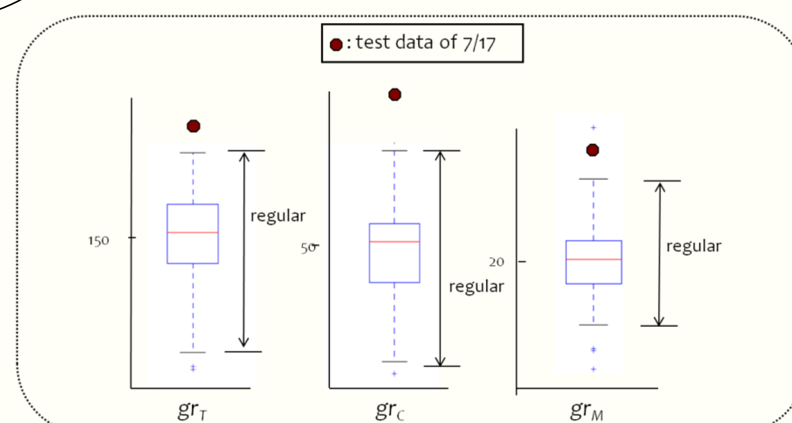
- confirmed unusually crowded regions from 3,600
  (4*3*300, for 4 time slots during 3 test days (7/17–7/19) in 300 RoIs)
  - the total number of RoIs that were evaluated as Unusual
    - 138(morning: 53, afternoon: 64, evening: 9, night: 12)
    - 3.8% (=138/3,600) of all the test time slots were answered as unusually crowded regions
  - could find 9 festivals among the prepared event list
    - recall = 9/15 = 60%
  - precision = 9/138 = 6.5%
    - We couldn't expect all the events
    - We found many other unexpected events



# Detection of Expected Events

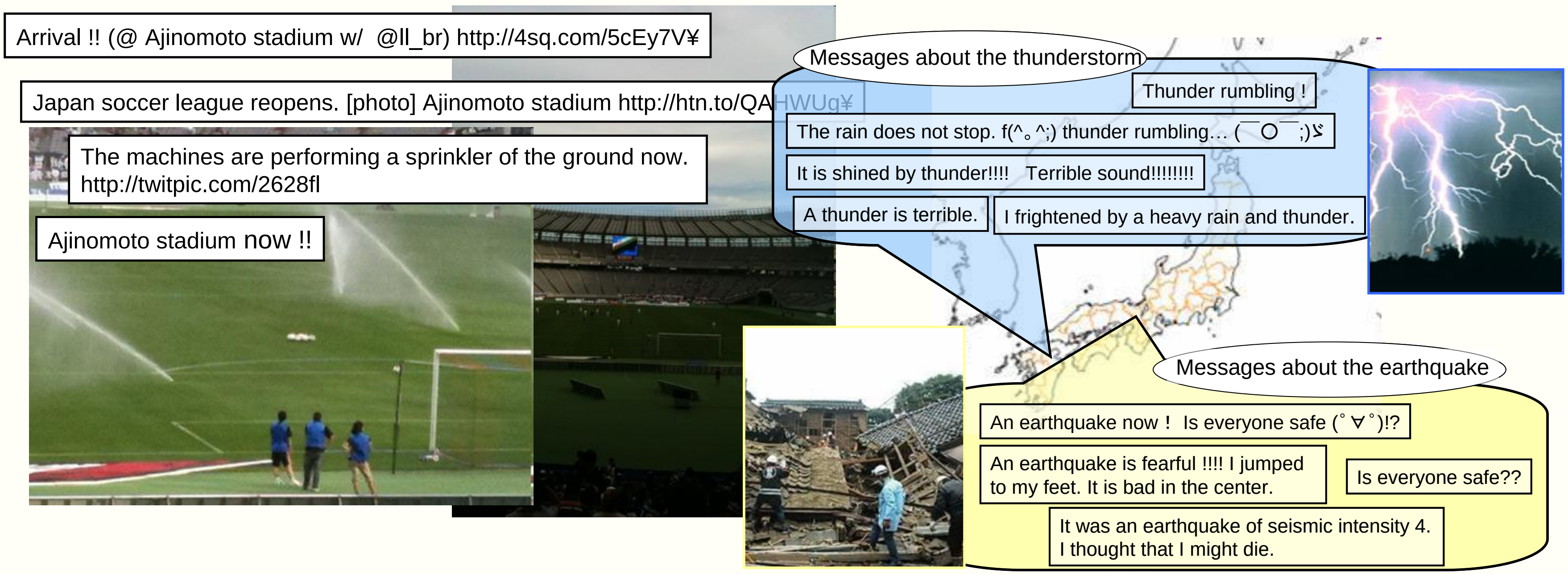

***An expected Event: Gion-Festival was found***

**$gr_T$, $gr_C$ and $gr_M$ of Kyoto, and the tendency in the afternoon of 7/17**

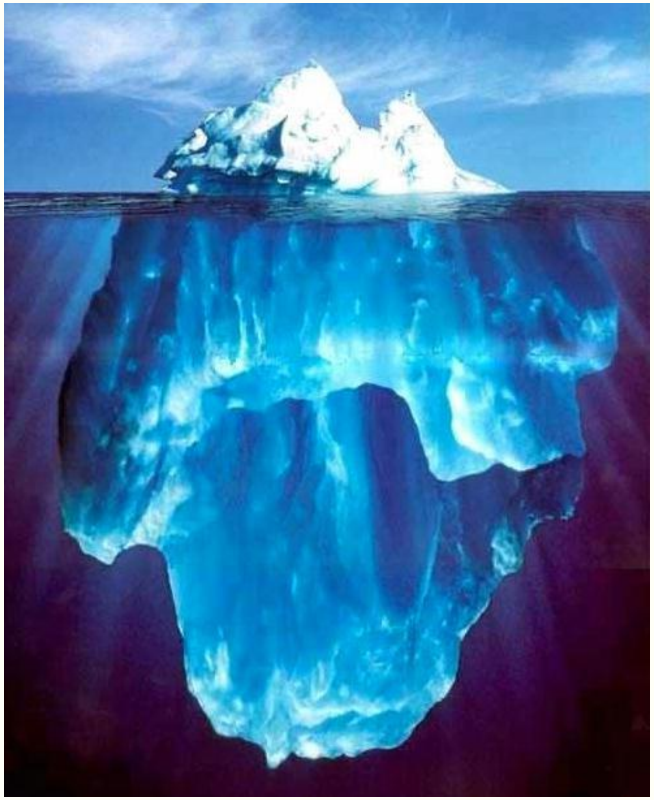


# Detection of Unexpected Events

- Newly discovered geo-social and natural events
  - Aggregation with supporters of a soccer team to watch the soccer game in a stadium
  - An earthquake / a heavy thunder with lightning (Natural Incidents)
- Precision = 52/138 = <u>38%</u>



# Conclusions



- We just touched a tip of Uncharted Iceberg by LBS + Micro-blog
  - We proposed Geo-social Event Detection Method with Twitter
  - We presented a concept of Geographic Regularity to detect usual statuses for social-geographic boundary

- Future Work:
  - Exploring Crowd Behavior Patterns for Various Event Types
  - Considering Diverse Granularities: Temporal /Spatial dims

Thank you very much for attention !!