

# Analysing Tweets: During natural disasters in Europe and Asia

Presenter: Kiran Zahra



Universität  
Zürich <sup>UZH</sup>

VGI Analytics - AGILE 2017



**Structural @Nalk** · 24 Aug 2016

6.2-magnitude earthquake crumbles central Italy [sbne.ws/r/vq6x](https://www.sbn.com/news/2016/08/24/earthquake-crumbles-central-italy)



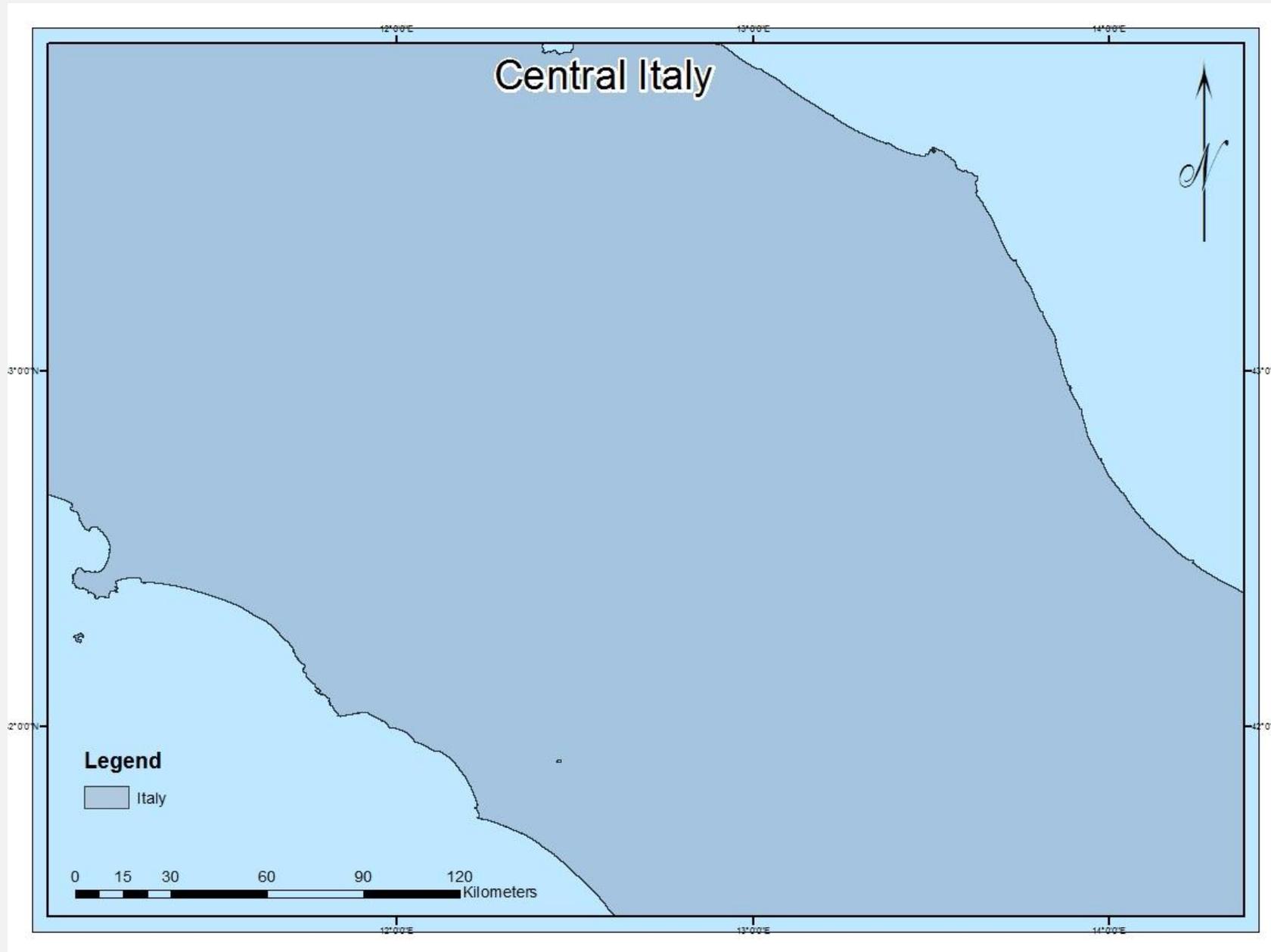


**Structural @Nalk** · 24 Aug 2016

6.2-magnitude earthquake crumbles central Italy [sbne.ws/r/vg6x](https://www.sbn.com/news/2016/08/24/earthquake-crumbles-central-italy)



Central  
Italy





**SweetOne** @alanacto2009 · 24 Aug 2016



Article: 'It was so strong, so sudden,' says survivor of deadly **#Italy**  
**#earthquake**



**SweetOne** @alanacto2009 · 24 Aug 2016

Article: 'It was so strong, so sudden,' says survivor of deadly **#Italy**  
**#earthquake**



Strong



**SweetOne** @alanacto2009 · 24 Aug 2016



Article: 'It was so strong, so sudden,' says survivor of deadly **#Italy**  
**#earthquake**

Strong

Sudden



**P J Anthony** @JMoota · 24 Aug 2016



**Earthquake in Myanmar Kills 3 and Damages Over 170 Temples**

[nyti.ms/2bFtKyO](https://nyti.ms/2bFtKyO)





**P J Anthony** @JMoota · 24 Aug 2016

**Earthquake in Myanmar Kills 3 and Damages Over 170 Temples**

[nyti.ms/2bFtKyO](https://nyti.ms/2bFtKyO)



3 Killed



**P J Anthony** @JMoota · 24 Aug 2016

**Earthquake in Myanmar Kills 3 and Damages Over 170 Temples**

[nyti.ms/2bFtKyO](https://nyti.ms/2bFtKyO)



3 Killed

170 Temples  
damaged



**CJ Lopez** @kickdblaj · May 1

That wasn't an [#earthquake](#). That was my responsibilities crashing down around me



5



212



500





**CJ Lopez** @kickdblclj · May 1

That wasn't an [#earthquake](#). That was my responsibilities crashing down around me

↩ 5

↻ 212

♥ 500

Earthquake

# Aim

- How geographic features are reported from two different continents during disasters
- The credibility of users reporting from two different continents
- Performance of machine learning algorithm to filter noise

# Data Collection

# Case-study

- Italy (Europe) 6.2 Magnitude
- Myanmar (Asia) 6.8 Magnitude

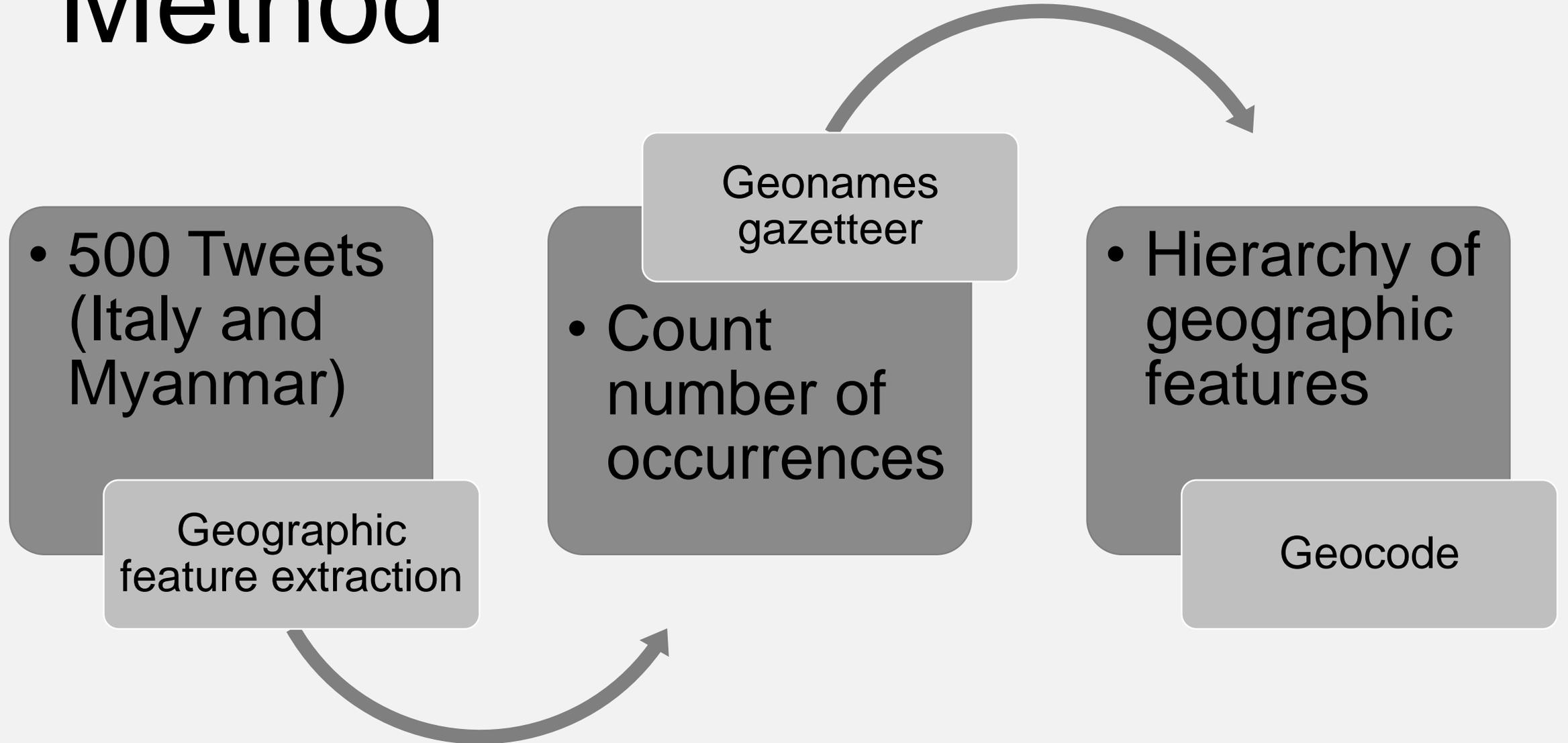
**Wed 24 Aug,  
16**

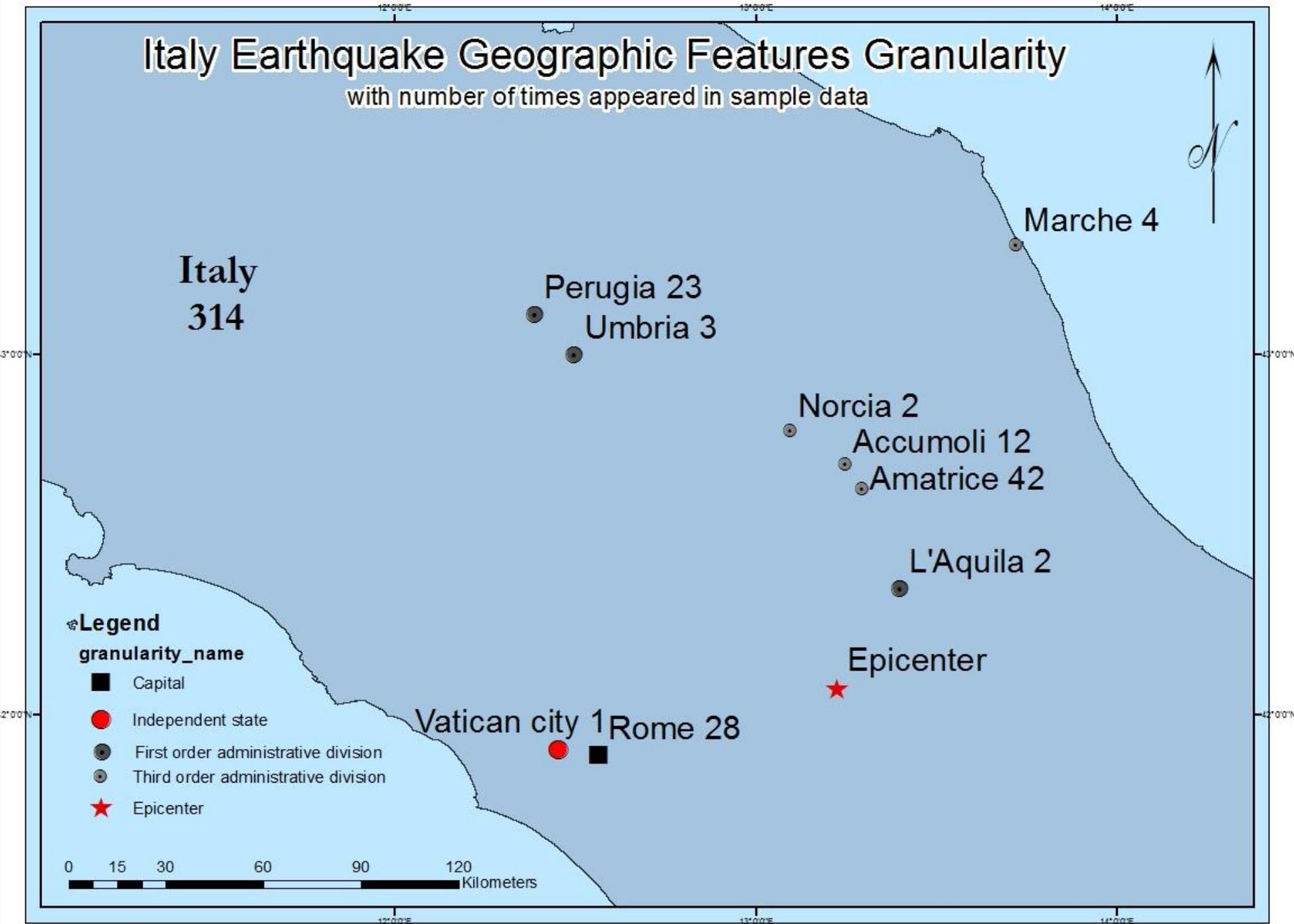
**488,238  
Tweets**

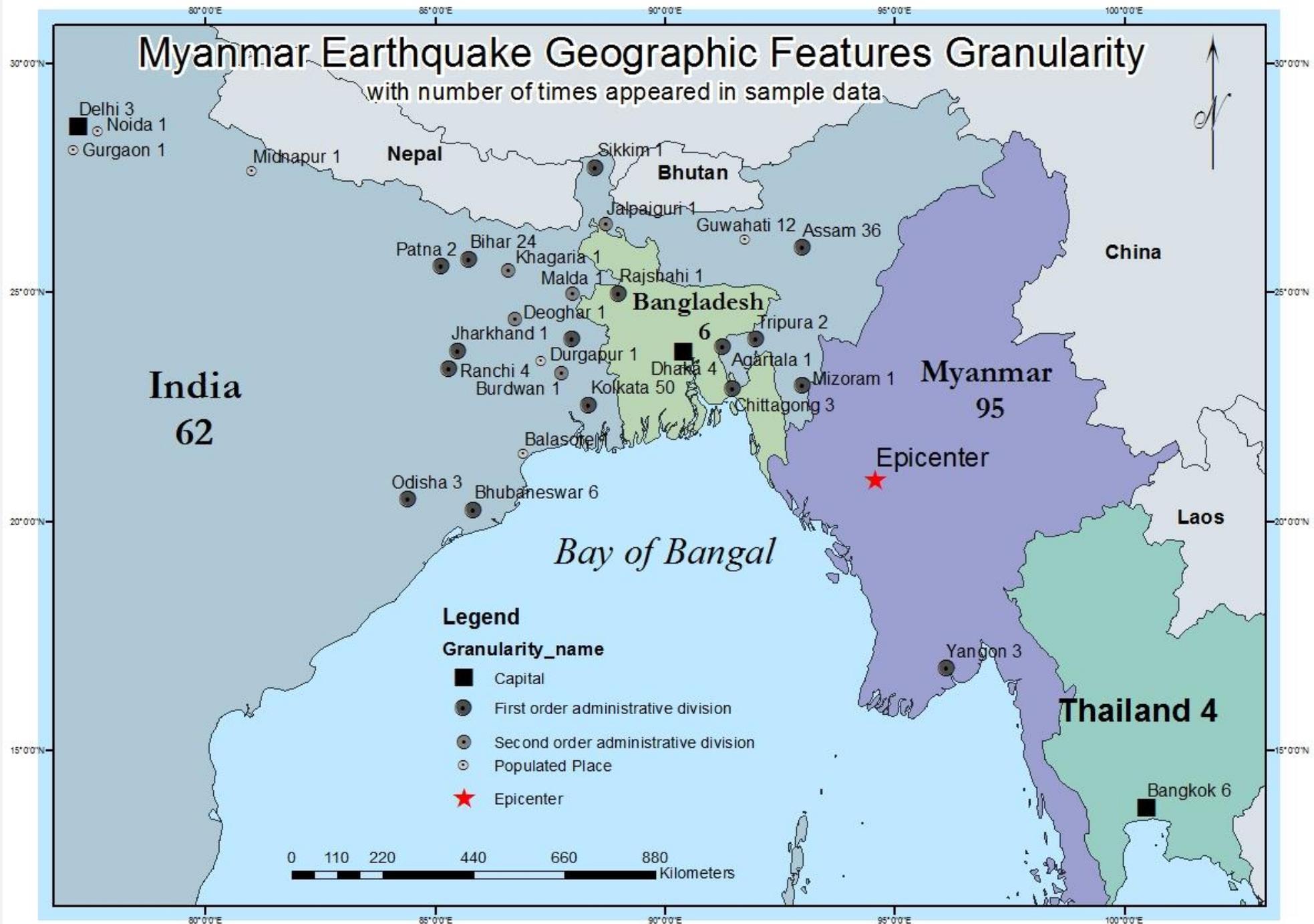
# Research Question 1

- How reported geographic feature granularity vary from two continents?

# Method







# Research Question 2

- What is the quality of tweets in terms of credibility?

# Twitter Metadata

- Retweet count
- Friends count
- Location
- Vrified
- Screen name etc.

# User-based Features

Account age (AG)	The time passed since the author registered their account
Statuses count (SC)	The number of tweet sent by the user
Followers count (FoC)	Number of people following this user
Friends count (FrC)	Number of people user is following
Verified (V)	If the account has been verified
Has description (D)	A non-empty bio
Has URL (U)	A non-empty homepage URL

(Castillo, Mendoza & Poblete 2011)

# Credibility

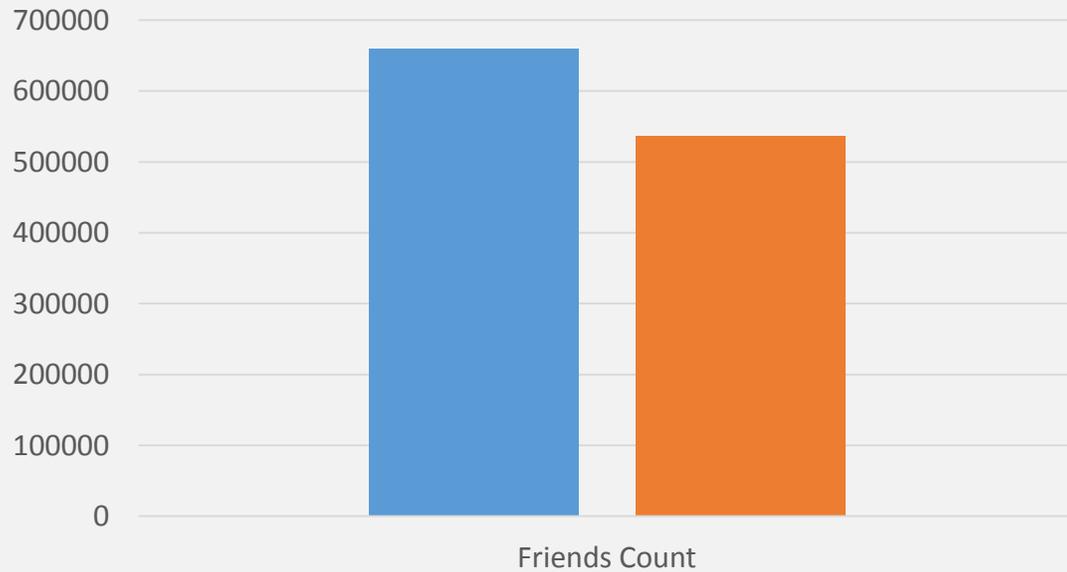
- Credibility is the function of

$$C = f(\text{FrC}, \text{SC}, \text{FoC}, \text{AG}, \text{U}, \text{D}, \text{V})$$

# Credibility

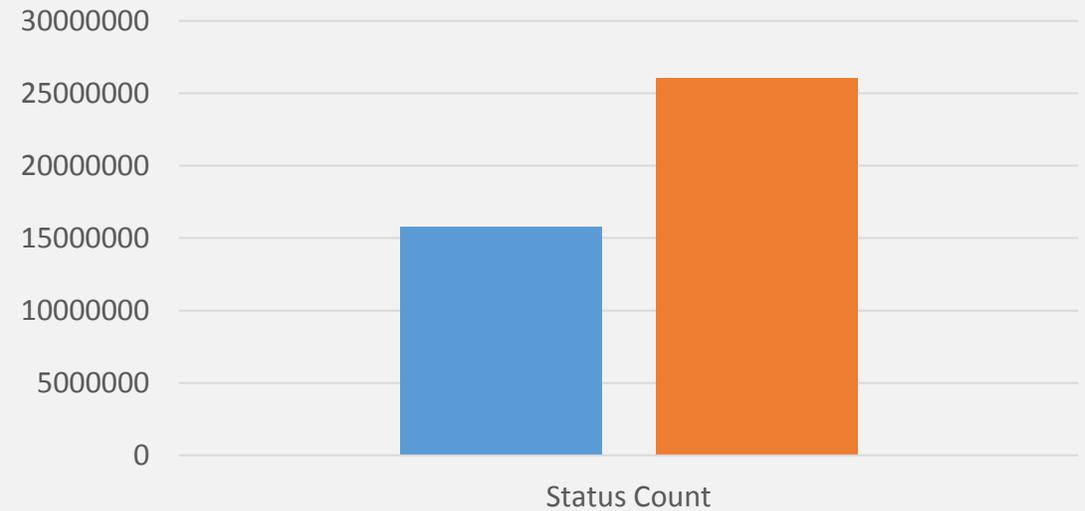
Friends Count

Italy Myanmar



Status Count

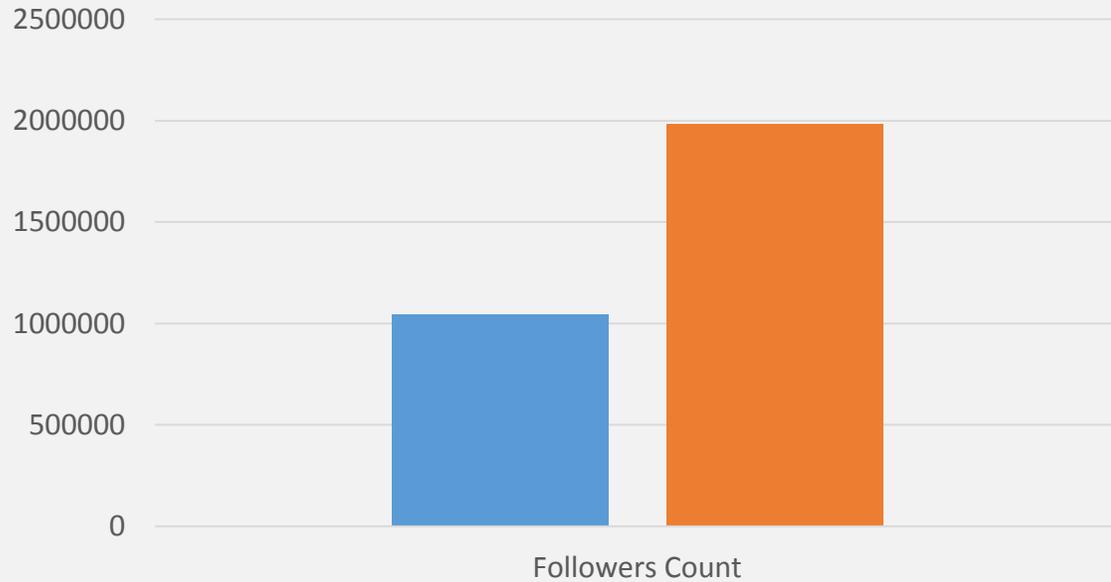
Italy Myanmar



# Credibility

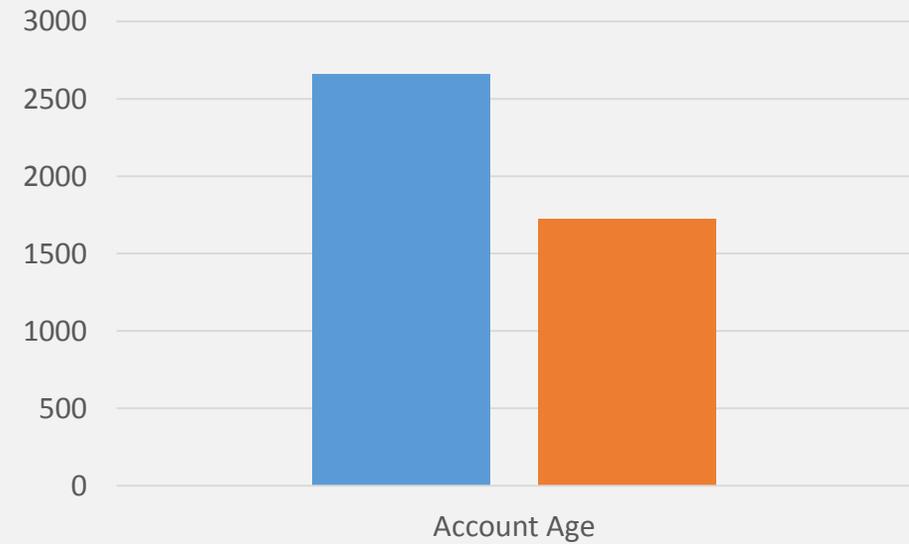
Followers Count

Italy Myanmar



Account Age

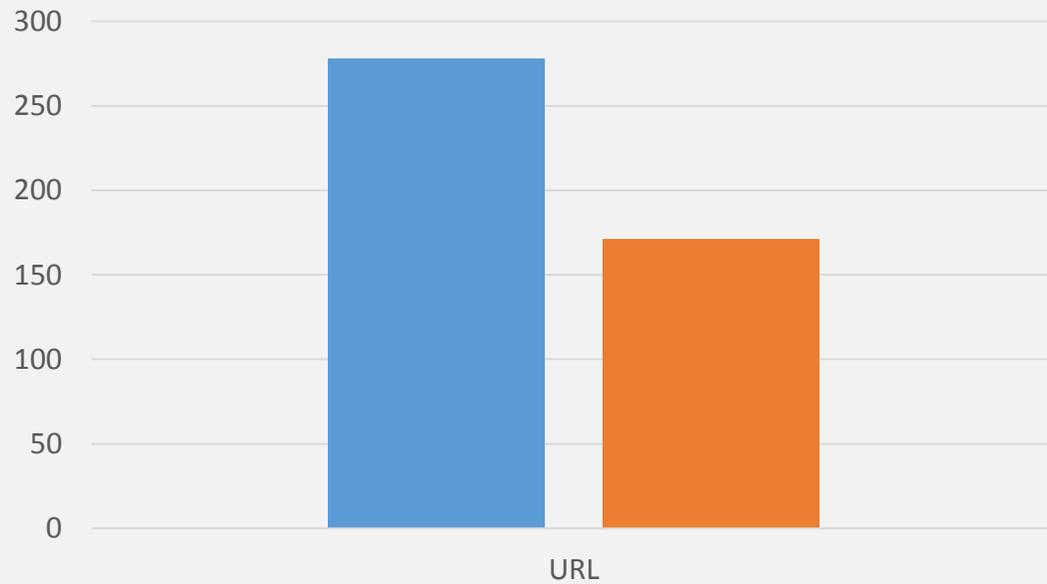
Italy Myanmar



# Credibility

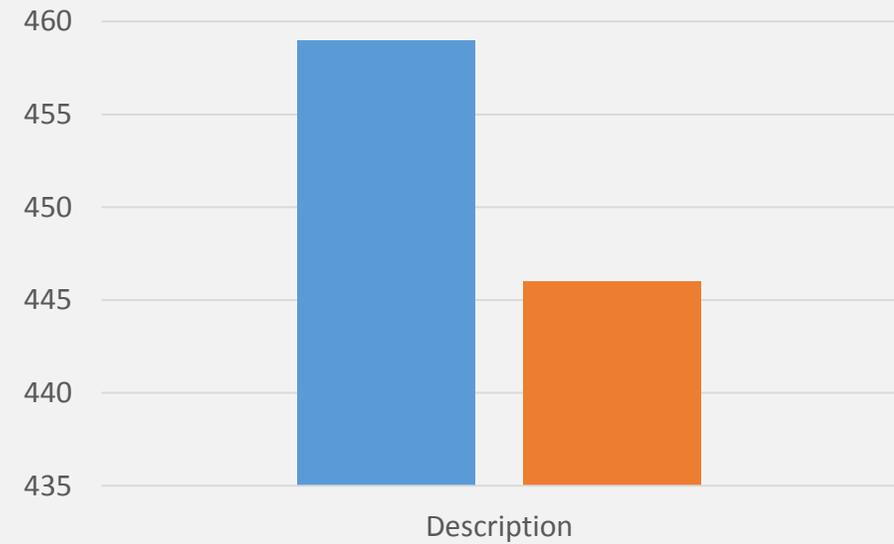
URL

Italy Myanmar



Description

Italy Myanmar





Classifiers trained for one event,  
work good for the same kind of  
event without training data

(Verma et al. 2011)

# Research Question 3

- How well Naïve Bayes perform when the classifier is trained for one event and tested on another event of same nature?

# Tweet Classification

- **Information:** Tweet text about disaster event and its location
- **Not Information:** Everything else falls in this category

# Information



**Twickelerveld INTEL** @LowlandsSN · 24 Aug 2016

**UPDATE: 73 people now known to have died in #Italy earthquake, Italy's civil protection unit says @BBCBreaking**



# Information



**Twickelerveld INTEL** @LowlandsSN · 24 Aug 2016

**UPDATE: 73 people now known to have died in #Italy earthquake**, Italy's civil protection unit says @BBCBreaking



Earthquake



Italy

# Information ☹️



**I Love My Kids** @MyKidsAreAholes · 24 Aug 2016

**Told my wife I need to go to Italy to help the earthquake victims but she thinks I only want to go to drink red wine and eat pizza.**



3



34



130



# Information ☹️



I Love My Kids @MyKidsAreAholes · 24 Aug 2016

Told my wife I need to go to Italy to help the earthquake victims but she thinks I only want to go to drink red wine and eat pizza.



3



34



130

Earthquake



Italy

# Not Information



**jeandiata** @jeandiata · 24 Aug 2016

Replying to @HillaryClinton

**Two major earthquakes. Death toll climbing.** You want to have your skills flood our timelines with this crap? #BadTiming



2



12



31



# Not Information



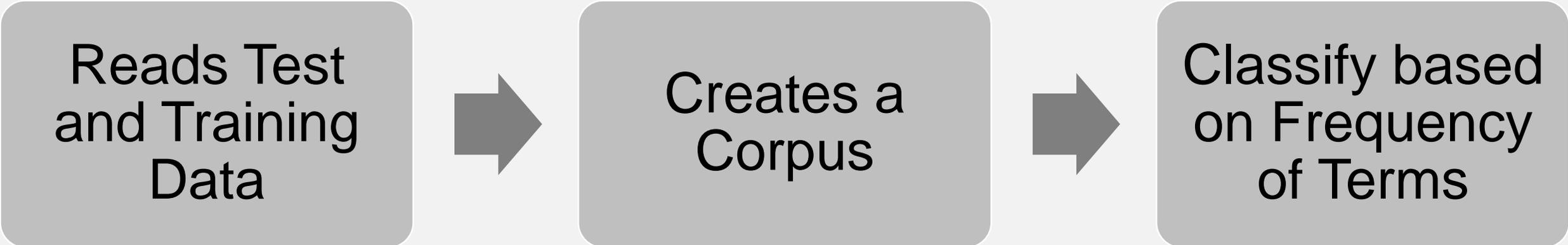
**Mae** ✨ @maepryor · May 4

Ok idk if **that was an earthquake** or I'm crazy but Ive never been so scared 😂



# Naïve Bayes Workflow

Reads Test  
and Training  
Data

A flowchart illustrating the Naïve Bayes workflow. It consists of three gray rounded rectangular boxes connected by right-pointing arrows. The first box contains the text 'Reads Test and Training Data'. An arrow points from this box to the second box, which contains 'Creates a Corpus'. Another arrow points from the second box to the third box, which contains 'Classify based on Frequency of Terms'.

Creates a  
Corpus

Classify based  
on Frequency  
of Terms

# Precision & Recall for Italy

- **Information:** 98% Precision , 93% Recall
- **Not Information:** 92% Precision, 97% Recall

# Naïve Bayes Workflow

Reads Test  
Data



Creates a  
Corpus



Classify based  
on Frequency  
of Terms

# Precision & Recall for Myanmar

- **Information:** 88% Precision , 91% Recall
- **Not Information:** 92% Precision, 89% Recall

# Key Points

- Granularity of geographic location is very important when reporting disasters
- Geographic locations can be differently reported from different regions of the world
- User-based credibility features reflect different user characteristics

# Key Points

- Difficult to judge credibility based only on one set of features
- Naïve Bayes trained on one event performed well on the same kind of event in a different geographic region

# Limitations

- Local administrative divisions do not match Geonames gazetteer
- Use of English language to collect data in non-native English speaking countries
- Twitter streaming API
- Sampling

Understanding the properties of social media are important before using this data during disasters