Social Data for Social Good
& a Biased Perspective on Research Impact

Alexandra Olteanu
Social Computing +
Computational Social Science

IBM Science for Social Good
Data Science Department
IBM TJ Watson Research Center

April 23, 2018 — WinDS, The Web Conference

Social good applications that leverage social data

Identify several data and methodological challenges

Highlight several ways in which you can have impact

Online Social Data

Flixster
ebay
LinkedIn
Quora
twitter
agoda
Booking.com
amazon.com
NETFLIX
facebook
flickr
KICKSTARTER
stackoverflow
goodreads
Meet your next favorite book.
WIKIPEDIA
The Free Encyclopedia
github
SOCIAL CODING
Beeradvocate
bing
YouTube
digg
foursquare
豆瓣 douban
Medium
Google
Instagram
GROUPON
Epinions.com
Pinterest
fitbit

Location traces

Play lists

Pageviews

Crowd-funding

Bookmarks

Recommendations

Collaborative editing

Activity tracking

Collaborative coding

Content generation

The Underlying Idea

We can analyze such user, behavioral traces to learn about the world.

Social Good

A Biased Set of Application Domains

Humanitarian crises/Social media use

Are we collecting the right data?
Can we generalize observations from one dataset to other seemingly similar datasets?

[ICWSM'14, CSCW'15]

Climate change/Media coverage bias

Is social media a good proxy for some phenomena of interest?

[ICWSM'15]

Minority issues/BlackLivesMatter movement

Is the user sample representative?

[AAAI SSS'16]

Health/Distilling outcomes from self-reports

Can we extract causal relations among personal events from social media?

[ICWSM'16, CSCW'17]

Hate speech/The effect of external events and of user traits

How do we evaluate systems that work with "subjective" concepts?
How do external events impact online phenomena?

[WebSci'17, ICWSM'18]

Humanitarian crises/Social media use

How can we retrieve comprehensive collections of crisis related messages?

Are we collecting the right data?

What kind of information is posted on social media during different type of crises?

Can we generalize findings from one dataset to other seemingly similar datasets?

Humanitarian crises/Social media use

How can we retrieve comprehensive collections of crisis related messages?
Are we collecting the right data?

What kind of information is posted on social media during different type of crises?
Can we generalize findings from one dataset to other seemingly similar datasets?

with **Carlos Castillo**, **Fernando Diaz**, and **Sarah Vieweg** [ICWSM'14]

Data Collection: How Is It Done?

Tweets are queried by

Maximum 1% of all tweets

Content

#prayforwest

#abflood

Low recall: 33%

Not everyone uses the keywords.

Maximum 400 terms.

Location

longitude: [-97.5, -96.5]

& latitude: [31.5, 32]

Low precision: 12%

Not everyone on the ground talks about the event.

Maximum 25 geo-rectangles.

Data Collection: How Is It Done?

Tweets are queried by

Maximum 1% of all

Firehose access!

Content

#prayforwest

#abflood

Low recall: 33%

Not everyone uses the keywords.

Maximum 400 terms.

Location

longitude: [-97.5, -96.5]
& latitude: [31.5, 32]

Low precision: 12%

Not everyone on the ground talks about the event.

Maximum 25 geo-rectangles.

Data Collection: How Is It Done?

Tweets are queried by

Maximum 1% of all

Firehose access!

Content

#prayforwest

#abflood

Low recall: 33%

Not everyone uses the keywords.

Maximum 400 terms.

Location

longitude: [-97.5, -96.5]
& latitude: [31.5, 32]

Low precision: 12%

Not everyone on the ground talks about the event.

Maximum 25 geo-rectangles.

We need better data collection pipelines!

Key Insight: Distill a Crisis Lexicon

- *damage*
- *affected people*
- *people displaced*
- *donate blood*
- *text redcross*
- *stay safe*
- *crisis deepens*
- *evacuated*
- *toll raises*
- … …

Gov. McDonnell: Virginia 'Spared' in Hurricane Sandy Damage patch.com/A-zgoL

Deeply touched by the blast at the Boston Marathon. Our thoughts with the affected people and their families. EH

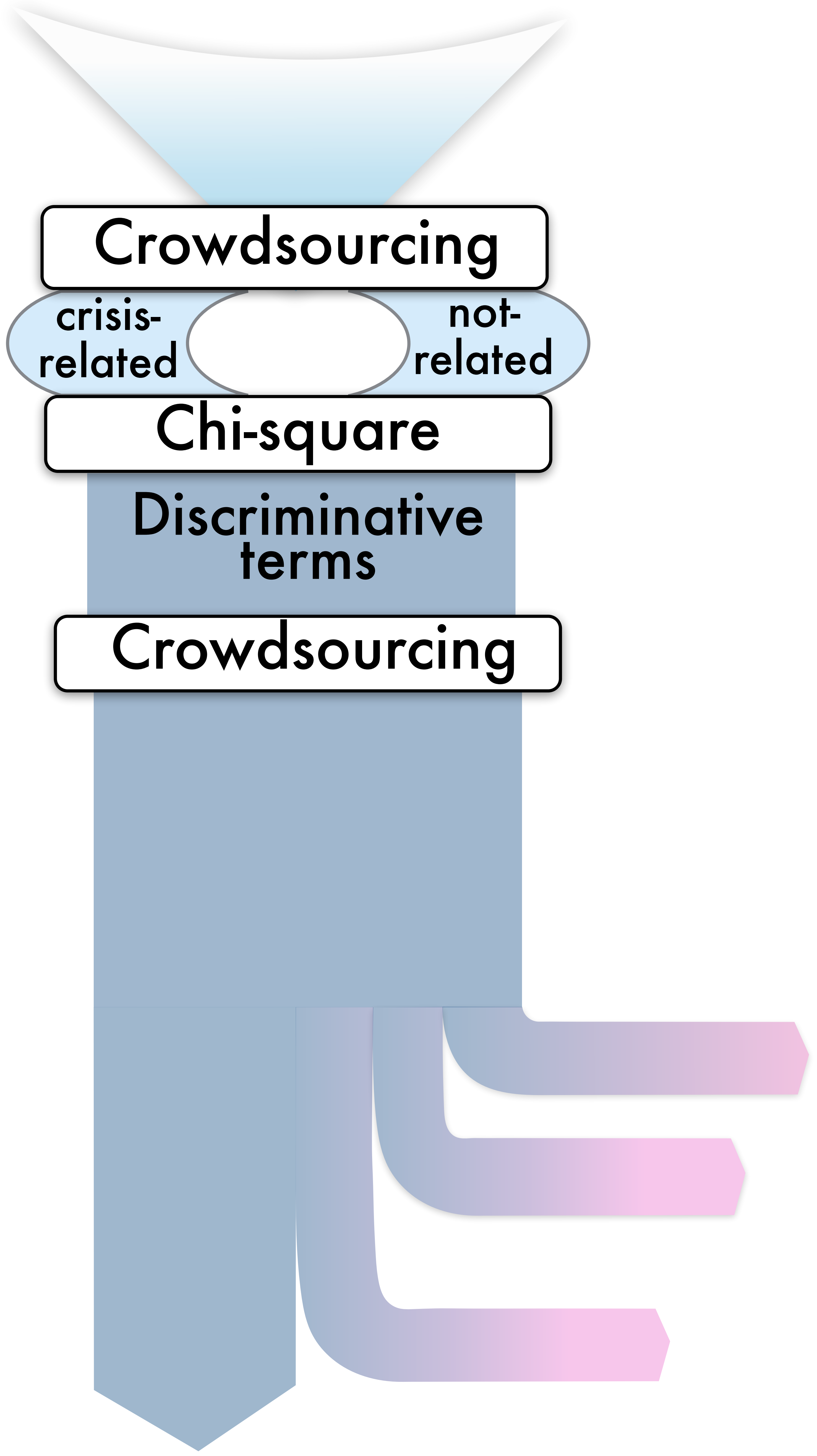
NOTE: Oklahoma University is providing shelter in their dorm facilities for people displaced by the tornado in OK today - @KFOR

If you are able to donate blood- Providence Hospital will have a blood drive in Waco from 11 a.m. to 5 p.m. #prayforwest

Best way to help tornado victims is to donate to the Red Cross at redcross.org or text REDCROSS to 90999. #okwx

This flooding is crazy! Hoping my fellow Albertans and Calgarians stay safe! #abflood #yycflood

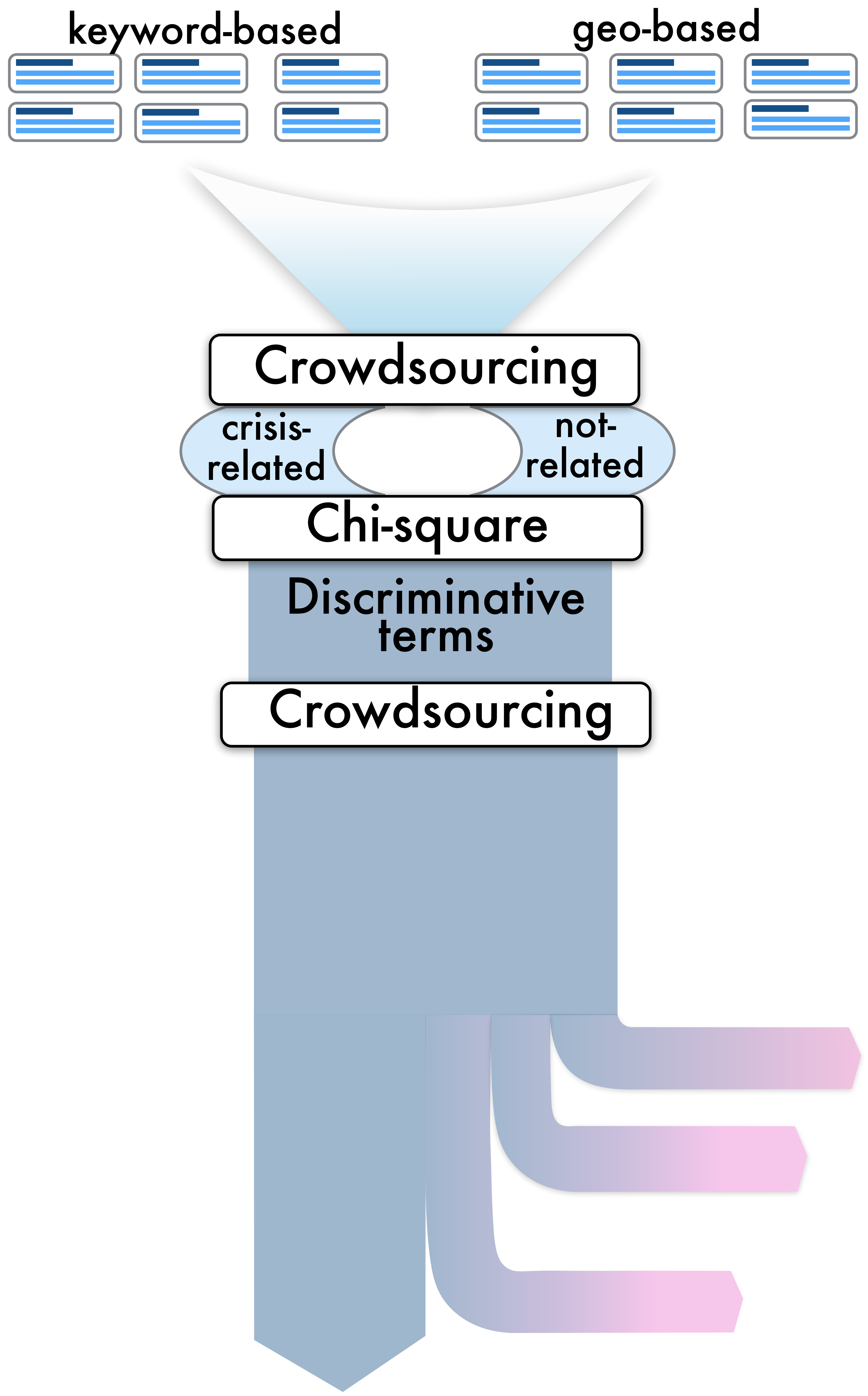
Lexicon Construction



Lexicon Construction

0. Offline collections

6x10,000 tweets per collection



Lexicon Construction

0. Offline collections

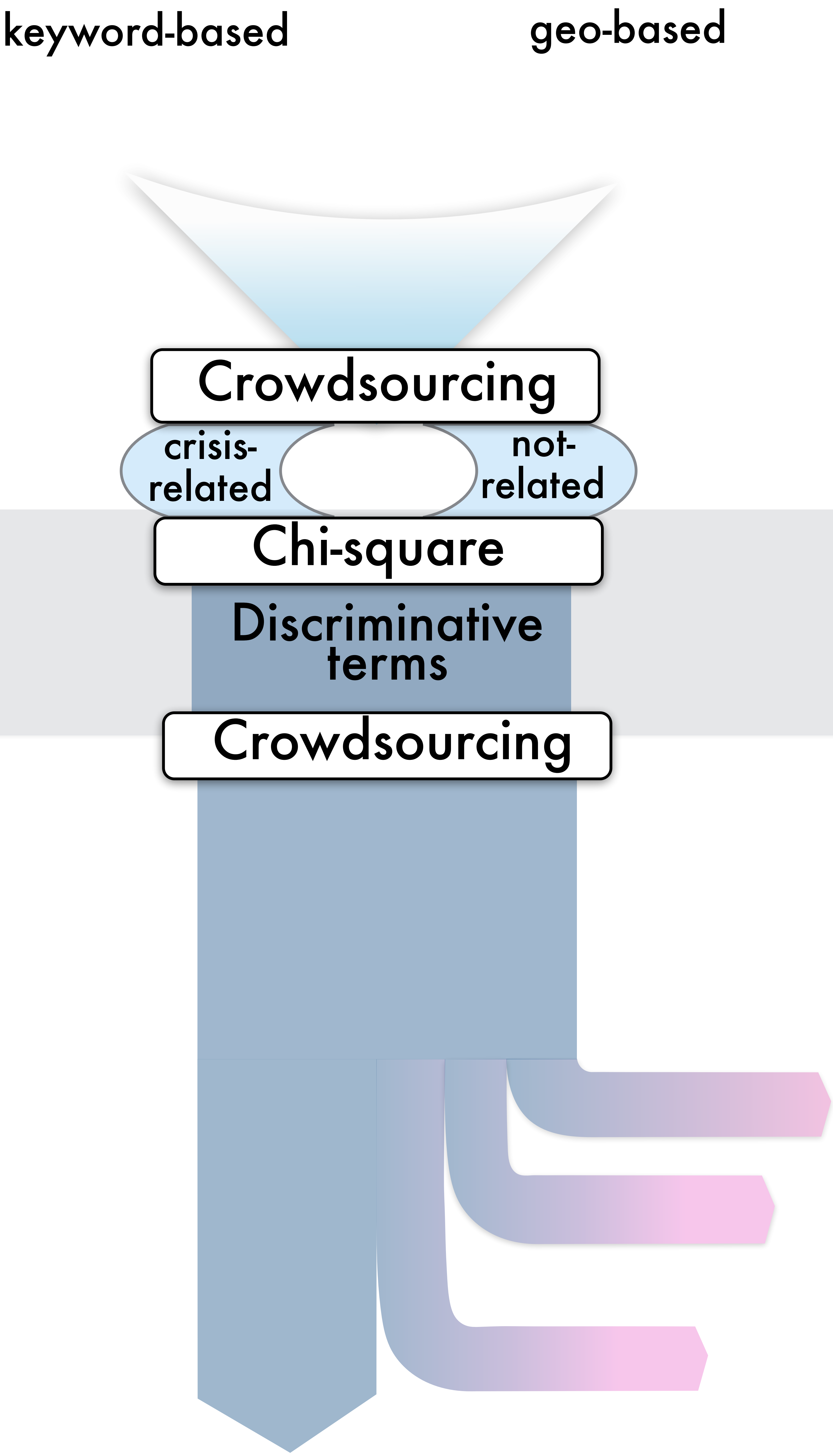
6x10,000 tweets per collection

1. Label tweets

Separate related from non-related tweets:
53.7% crisis-related

2. Extract & rank discriminative terms

Use statistical tests to extract terms more likely to appear in crisis tweets (e.g. Chi2, PMI)



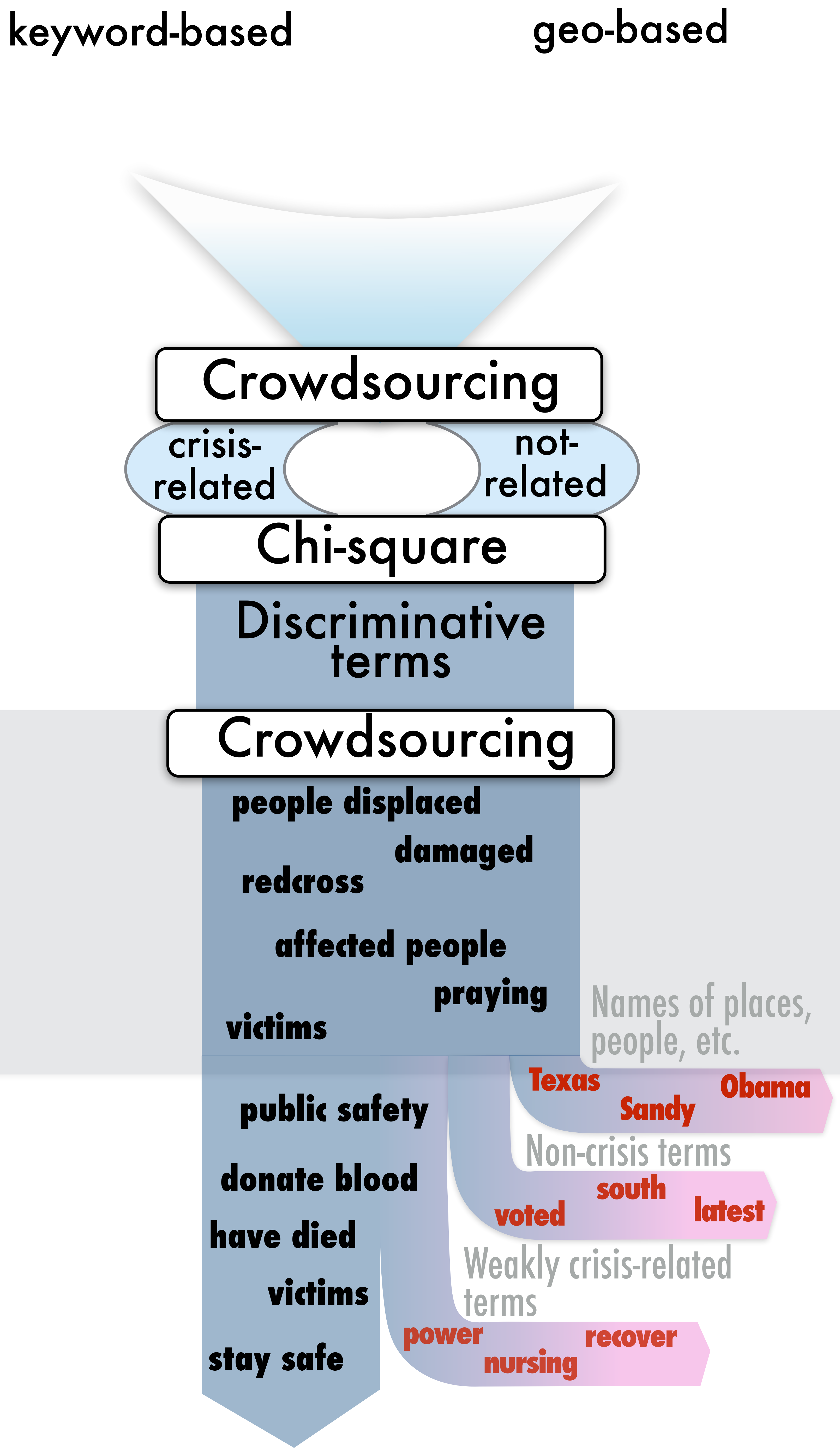
Lexicon Construction

0. Offline collections
6x10,000 tweets per collection

1. Label tweets
Separate related from non-related tweets:
53.7% crisis-related

2. Extract & rank discriminative terms
Use statistical tests to extract terms more likely to appear in crisis tweets (e.g. Chi2, PMI)

3. Revise terms
Strong crisis-terms.
Weakly crisis-terms.
Non-crisis terms.
Names of places, people, etc.



Lexicon Construction

0. Offline collections

6x10,000 tweets per collection

1. Label tweets

Separate related from non-related tweets:

53.7% crisis-related

2. Extract & rank discriminative terms

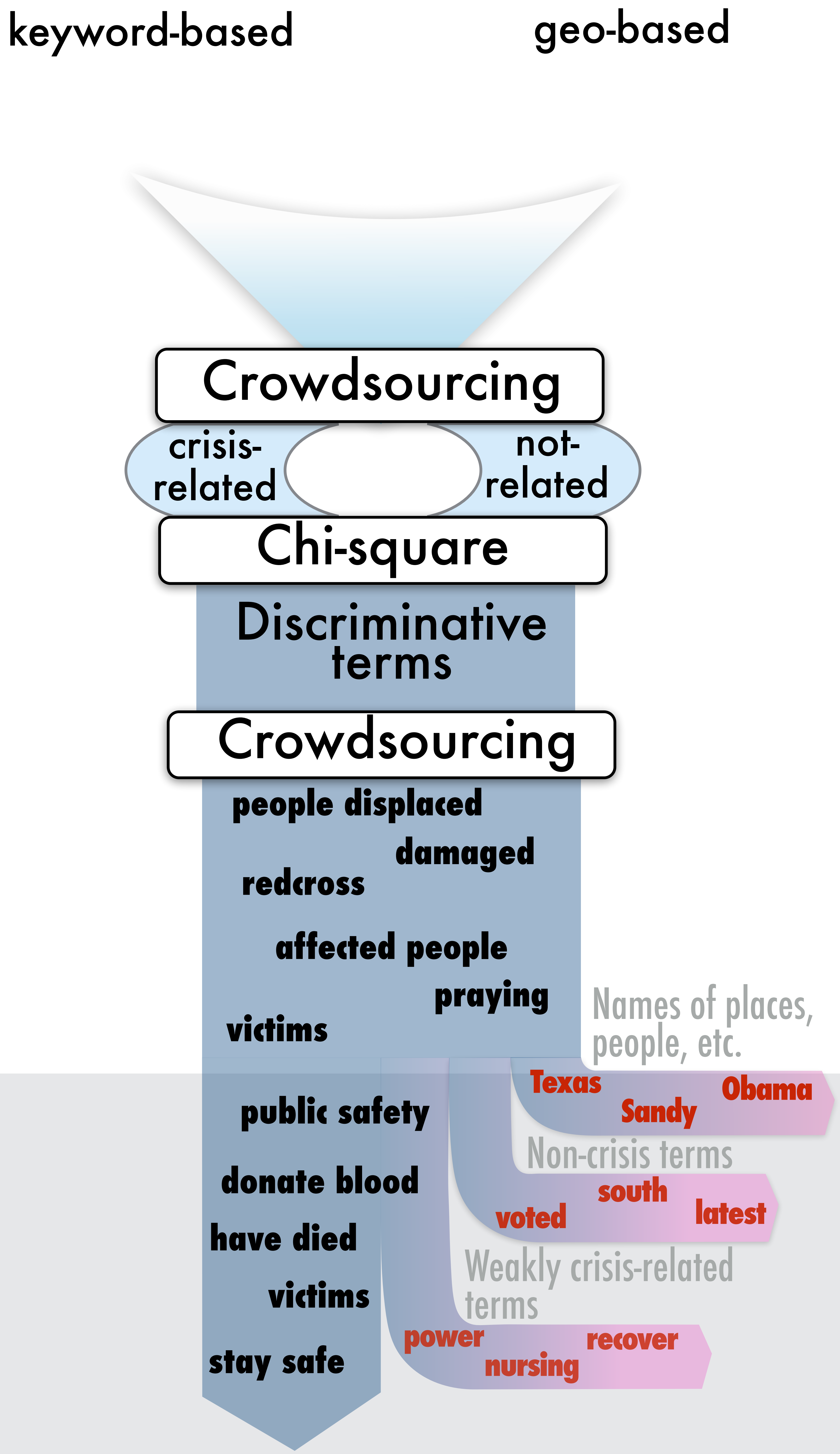
Use statistical tests to extract terms more likely to appear in crisis tweets (e.g. Chi2, PMI)

3. Revise terms

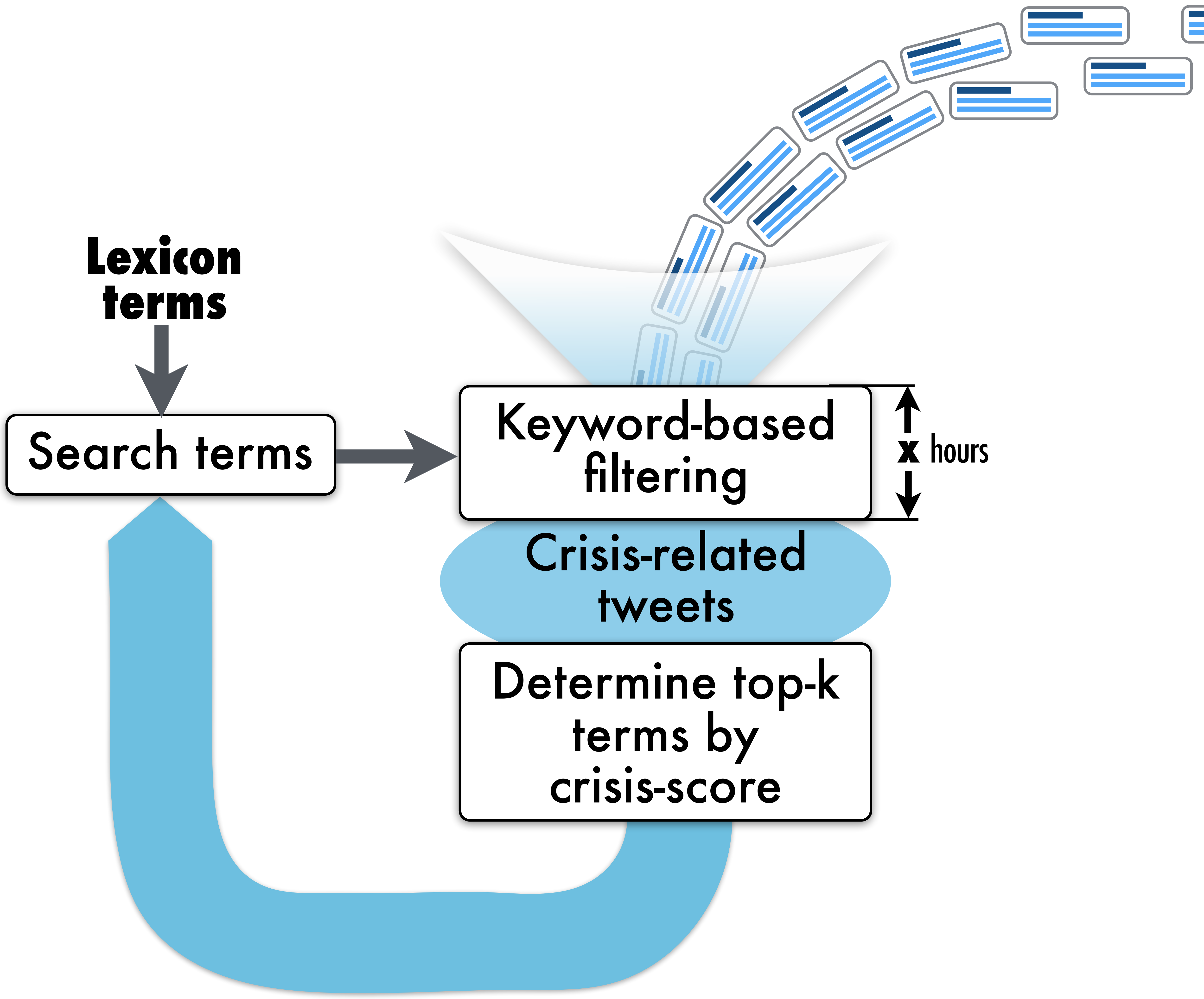
Strong crisis-terms.
Weakly crisis-terms.
Non-crisis terms.
Names of places, people, etc.

4. Remove co-occurring terms with lower scores

Maximum weighted coverage set on the term co-occurrence graph.



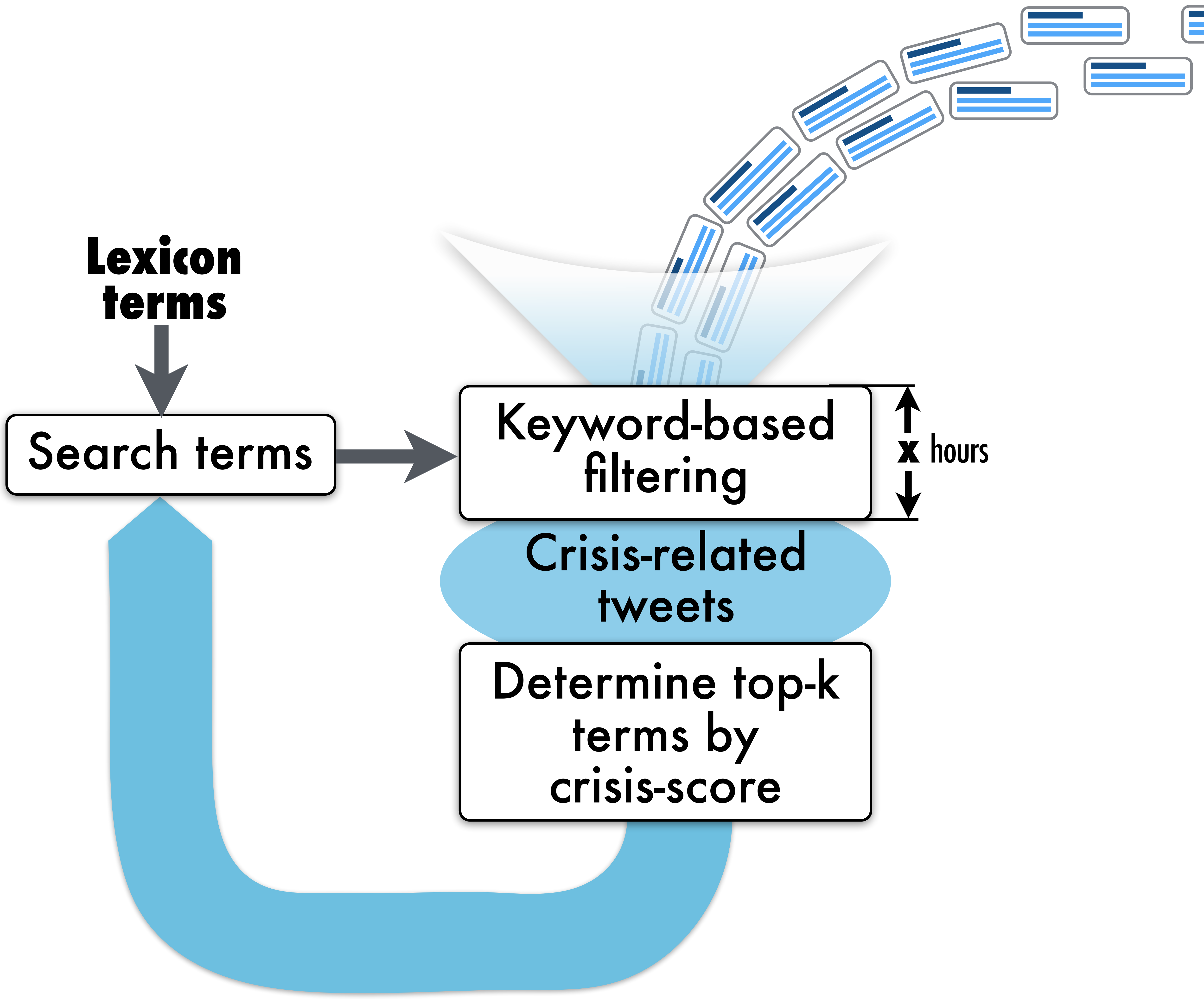
Apply and Expand the Lexicon



Apply and Expand the Lexicon

1. Collect tweets

x hours of pseudo-relevant tweets.



Apply and Expand the Lexicon

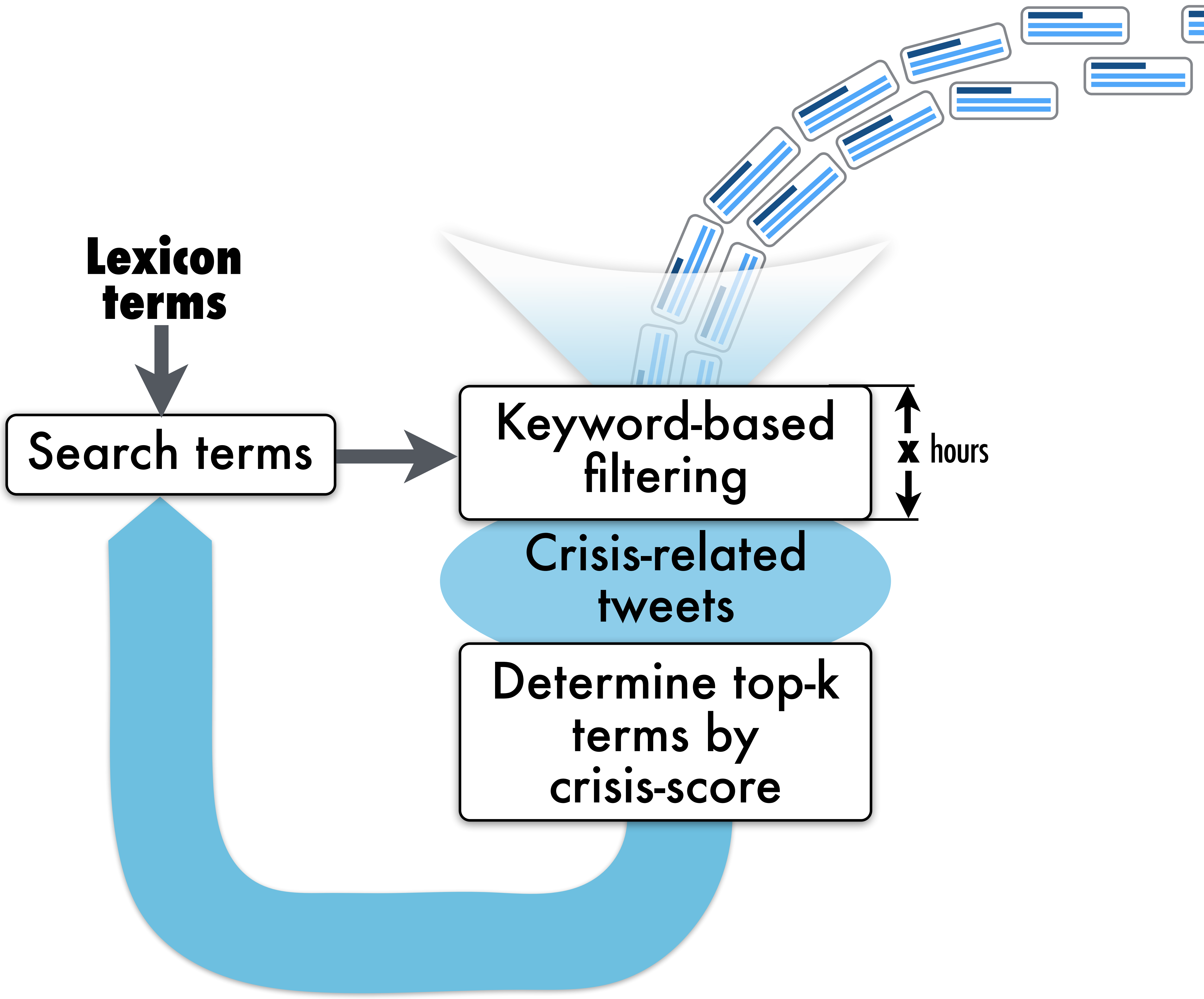
1. Collect tweets

   x hours of pseudo-relevant tweets.

2. Extract & rank terms

   Hashtags, unigrams & bigrams.

   Use label propagation or frequency to rank terms.



Apply and Expand the Lexicon

1. Collect tweets

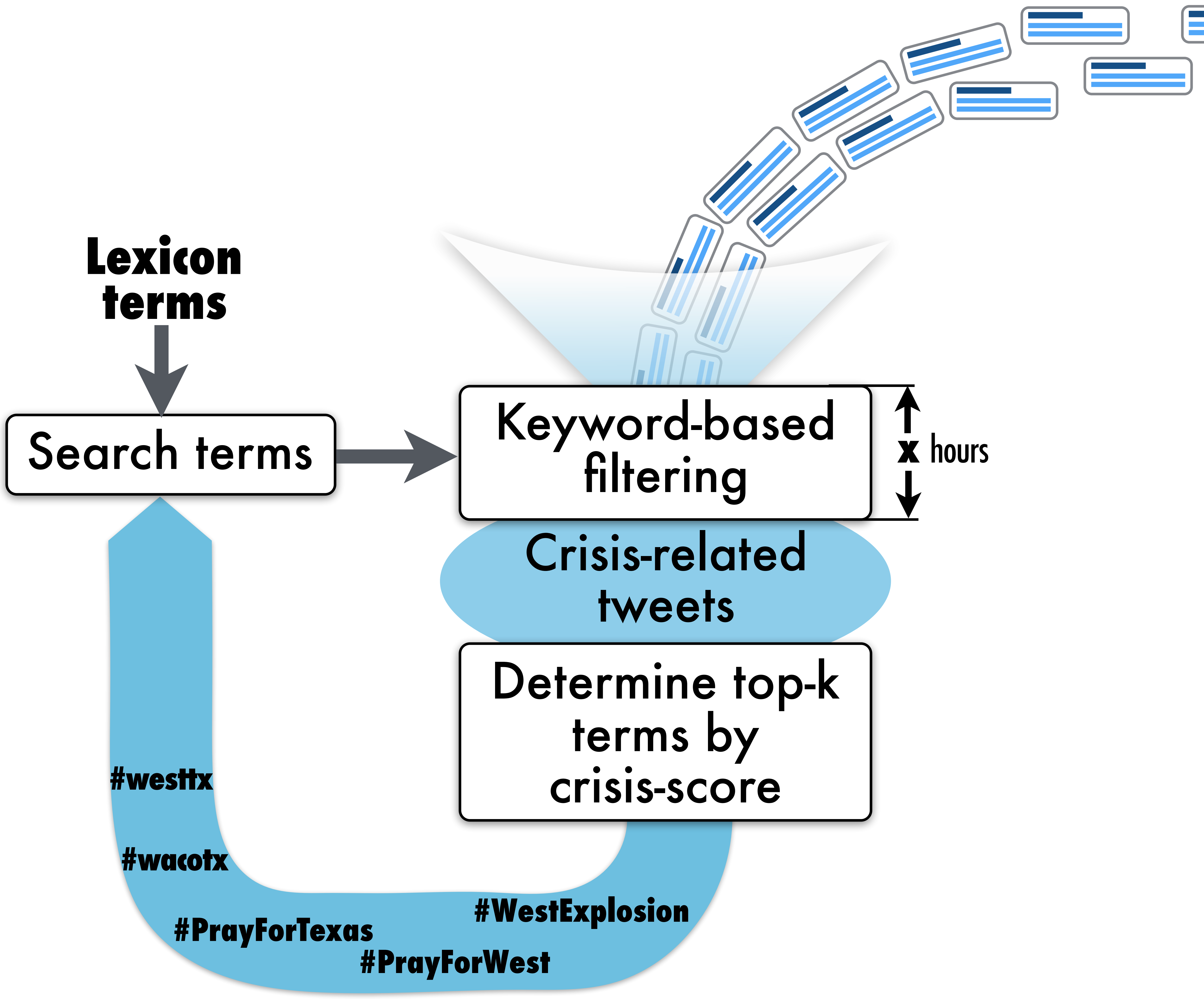
   x hours of pseudo-relevant tweets.

2. Extract & rank terms

   Hashtags, unigrams & bigrams.
   Use label propagation or frequency to rank terms.

3. Add **k** new terms to the lexicon

   Explore various sampling strategies.



Precision vs. Recall

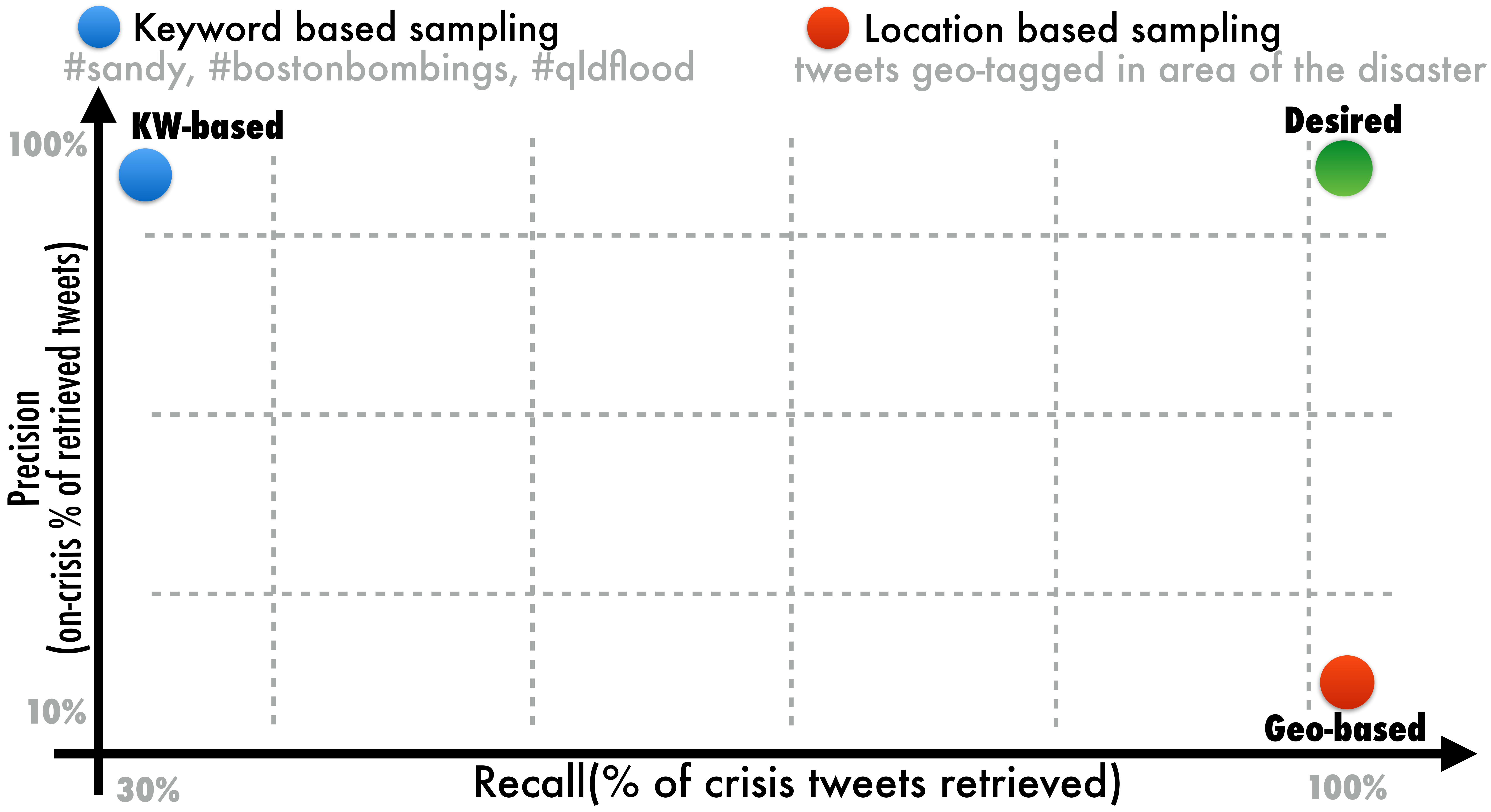
Precision is straightforward to measure.

Recall requires a complete data collection. We use geo data as proxy.

Precision vs. Recall

Precision is straightforward to measure.

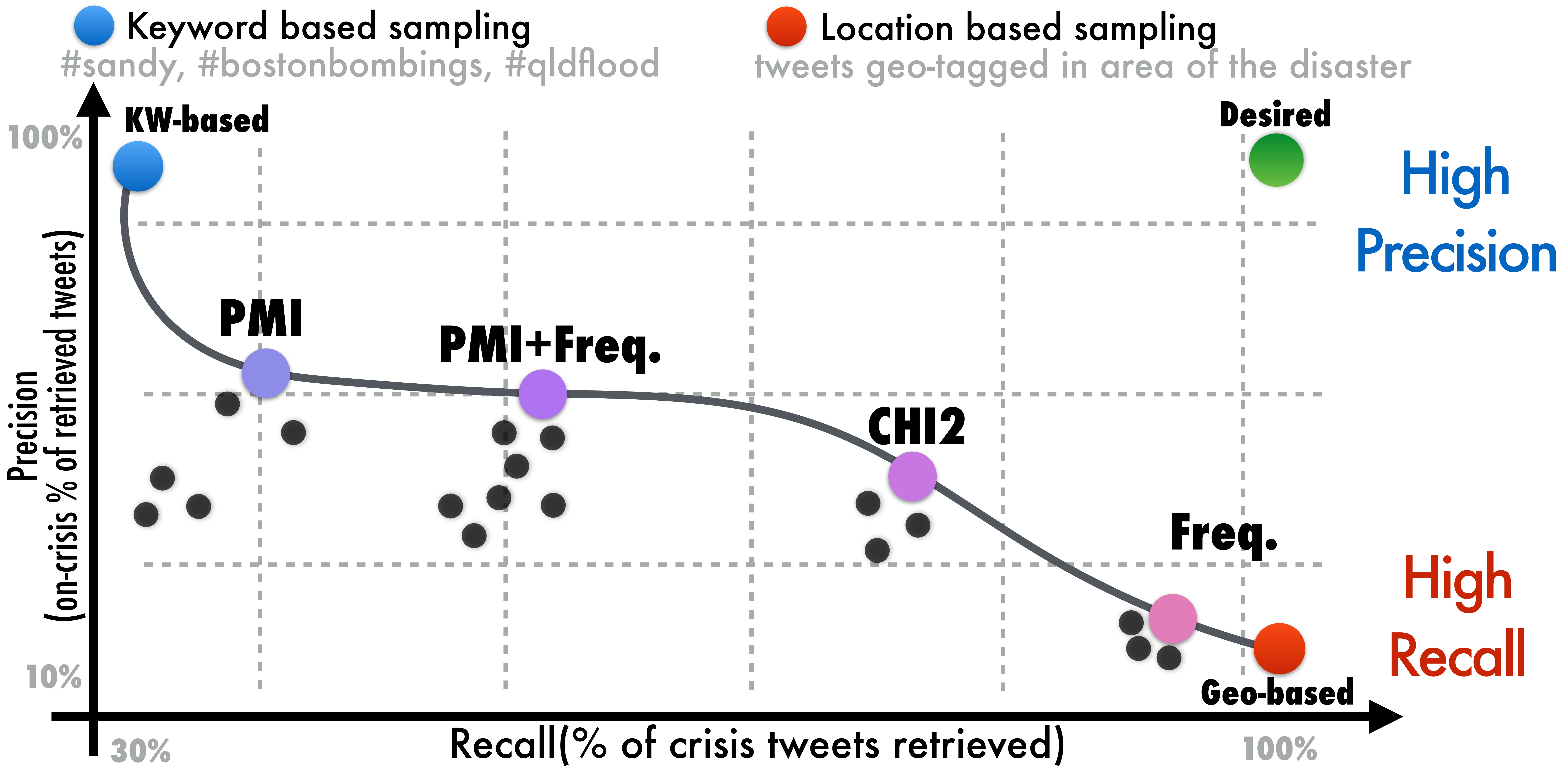
Recall requires a complete data collection. We use geo data as proxy.



Precision vs. Recall

Precision is straightforward to measure.

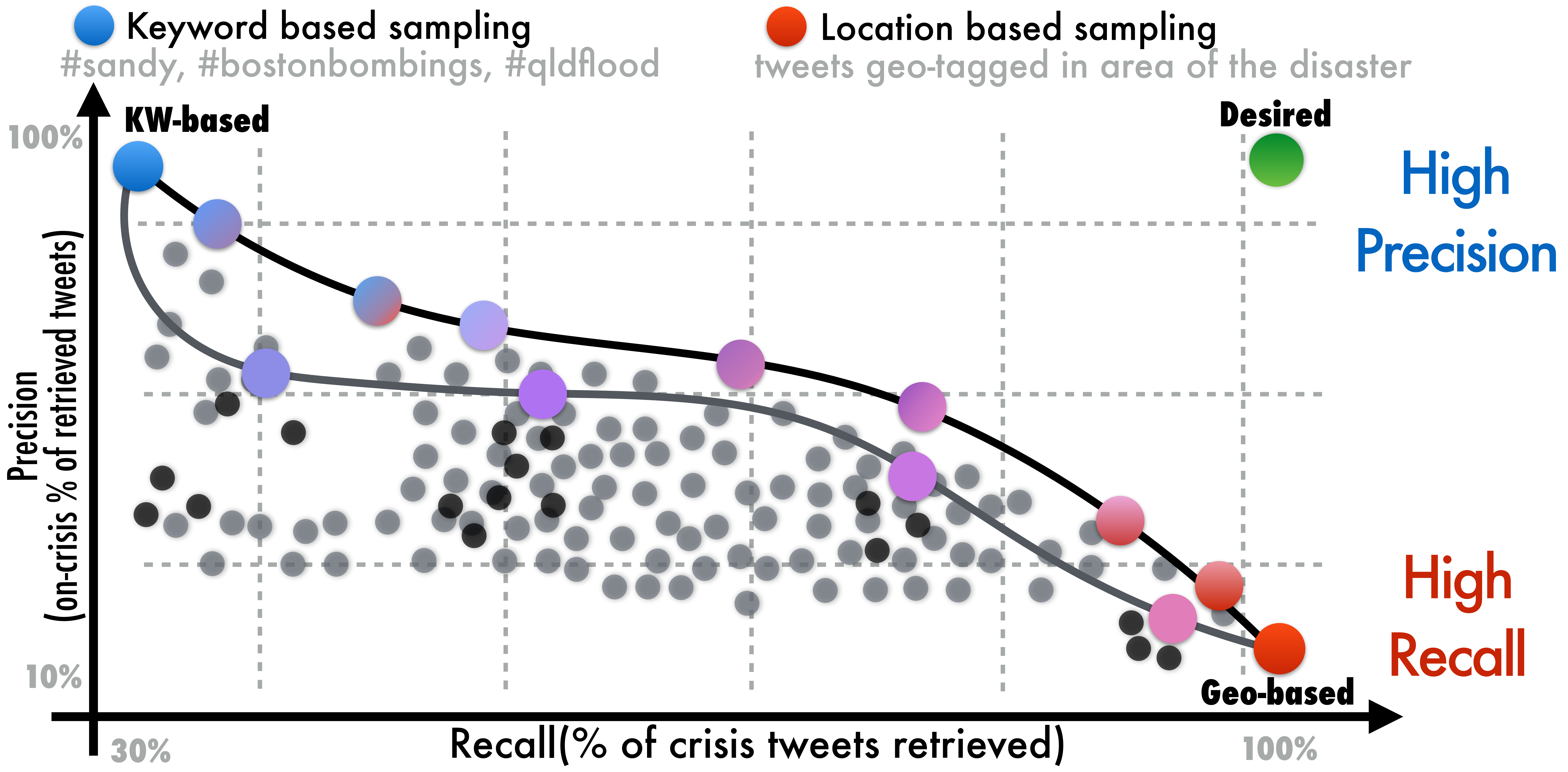
Recall requires a complete data collection. We use geo data as proxy.



Precision vs. Recall

Precision is straightforward to measure.

Recall requires a complete data collection. We use geo data as proxy.



Humanitarian crises/Social media use

How can we improve the data collection pipeline during crises/sudden onset events?
Do we collect the right data?

What are the differences in social media use during different type of crises?
Can we generalize findings from one dataset to other similar datasets?

with **Carlos Castillo** and **Sarah Vieweg** [CSCW'15]

Data Collection & Annotation

- ✓ Twitter sample API
- ✓ Keyword-based searches
- ✓ 26 crisis events
- ✓ 1000 annotated tweets per crisis

Data Collection & Annotation

- ✓ Twitter sample API
  - ✓ 2012 & 2013
  - ✓ ~1% random sample of Twitter public stream
  - ✓ ~130+ million tweets per month
- ✓ Keyword-based searches
- ✓ 26 crisis events
- ✓ 1000 annotated tweets per crisis

Data Collection & Annotation

- ✓ Twitter sample API
- ✓ Keyword-based searches
- ✓ 26 crisis events
- ✓ 1000 annotated tweets per crisis

Data Collection & Annotation

- ✓ Twitter sample API
- ✓ Keyword-based searches
  - ✓ proper names of affected location
    - ✓ manila floods, boston bombings, #newyork derailment
  - ✓ proper names of meteorological phenomena
    - ✓ sandy hurricane, typhoon yolanda
  - ✓ promoted hashtags
    - ✓ #SafeNow, #RescuePH, #ReliefPH
- ✓ 26 crisis events
- ✓ 1000 annotated tweets per crisis

Data Collection & Annotation

- ✓ Twitter sample API
- ✓ Keyword-based searches
- ✓ 26 crisis events
- ✓ 1000 annotated tweets per crisis

Data Collection & Annotation

- ✓ Twitter sample API
- ✓ Keyword-based searches
- ✓ 26 crisis events
  - ✓ 14 countries and 8 languages
  - ✓ 12 different hazard types
    - ✓ *earthquakes, wildfires, floods, bombings, shootings, etc.*
  - ✓ 15 instantaneous crises
- ✓ 1000 annotated tweets per crisis

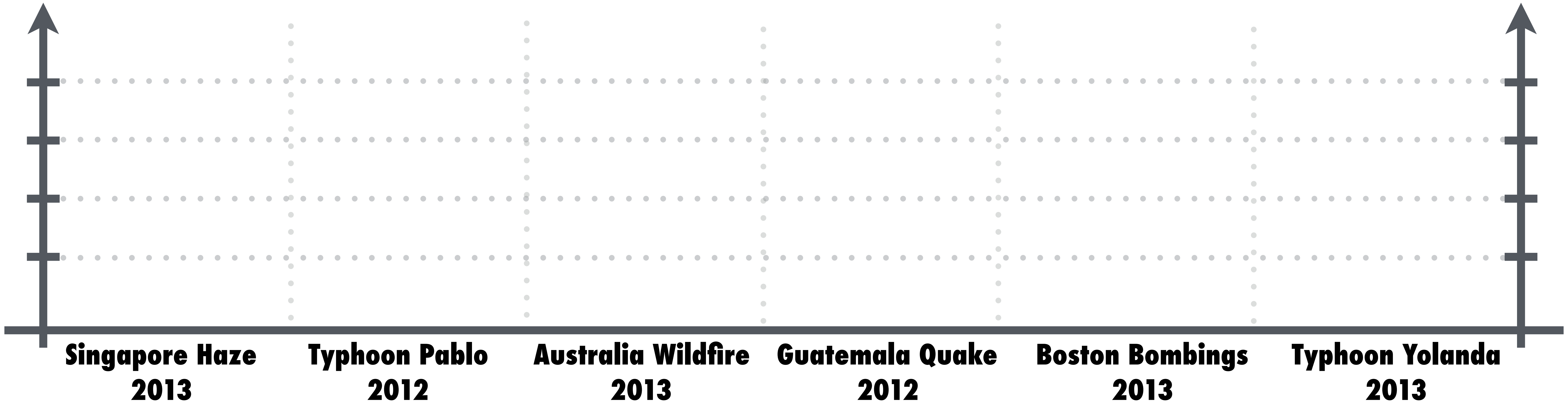
Data Collection & Annotation

- ✓ Twitter sample API
- ✓ Keyword-based searches
- ✓ 26 crisis events
- ✓ 1000 annotated tweets per crisis

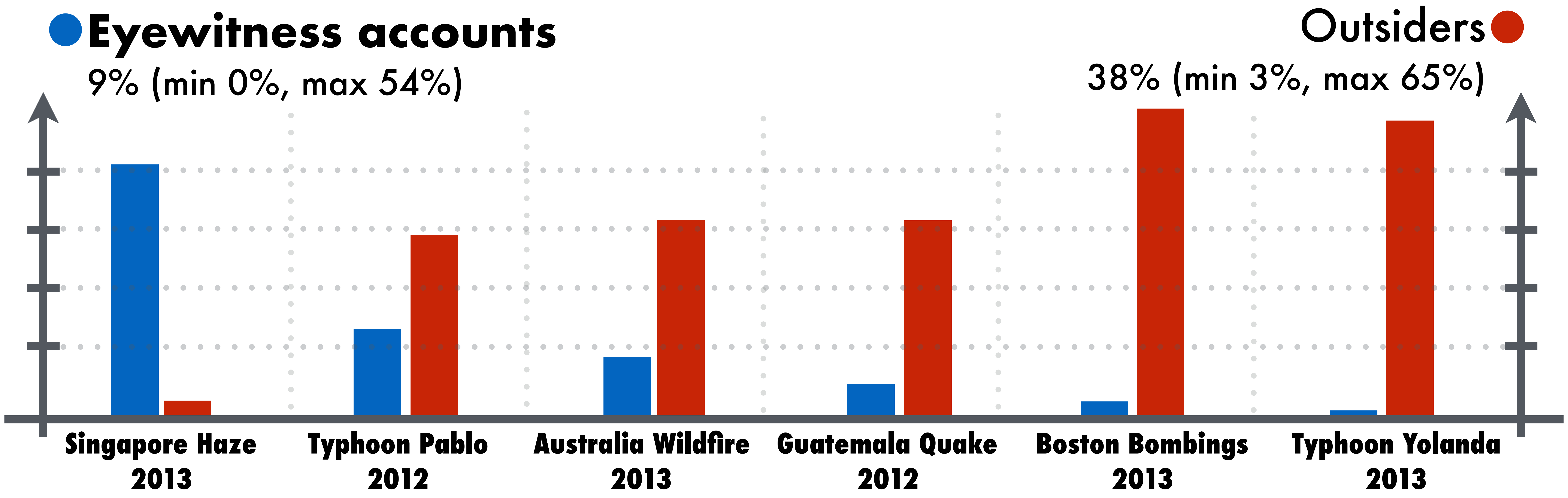
Data Collection & Annotation

- ✓ Twitter sample API
- ✓ Keyword-based searches
- ✓ 26 crisis events
- ✓ 1000 annotated tweets per crisis
  - ✓ Content dimensions
    - ✓ Informativeness
    - ✓ Source of information
    - ✓ Type of information
  - ✓ Crowdsource workers from the affected countries

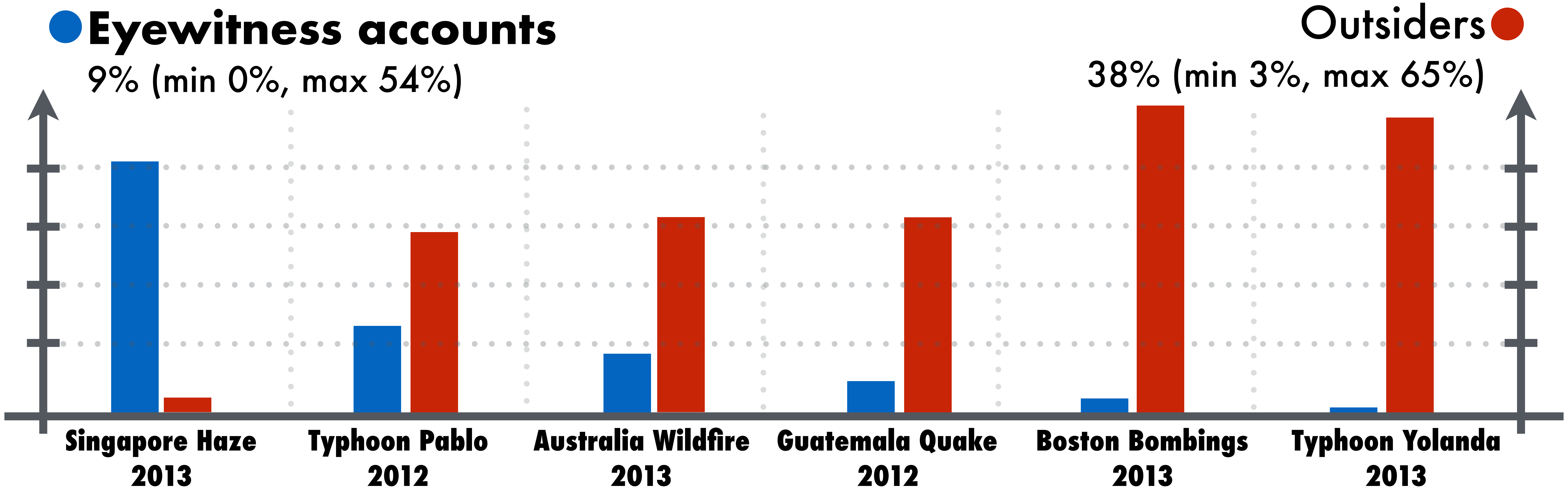
Content & Source Variations

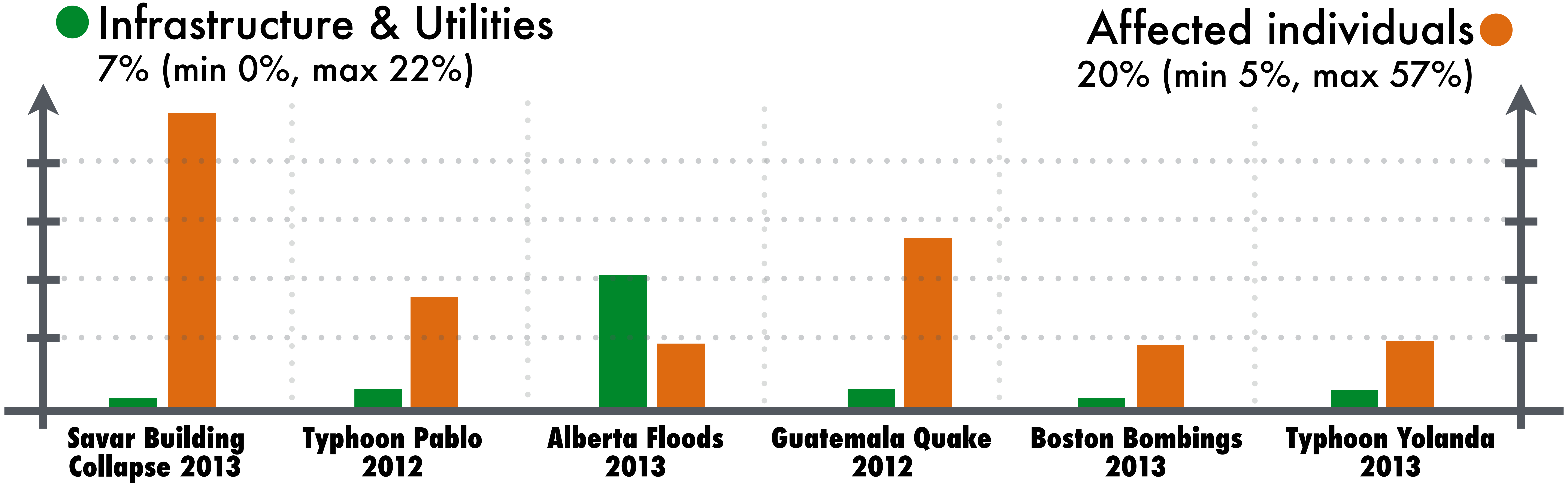


Content & Source Variations

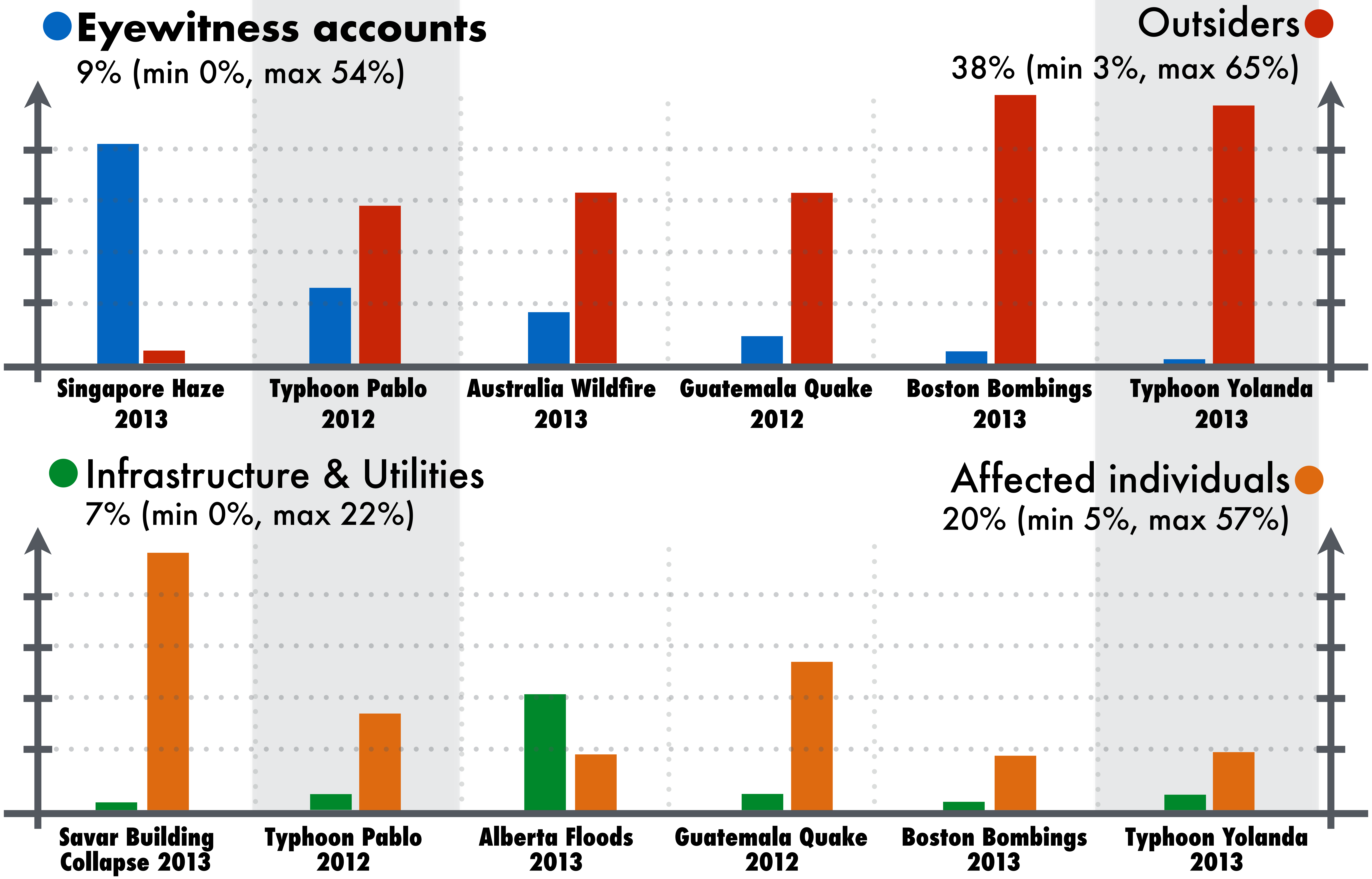


Content & Source Variations

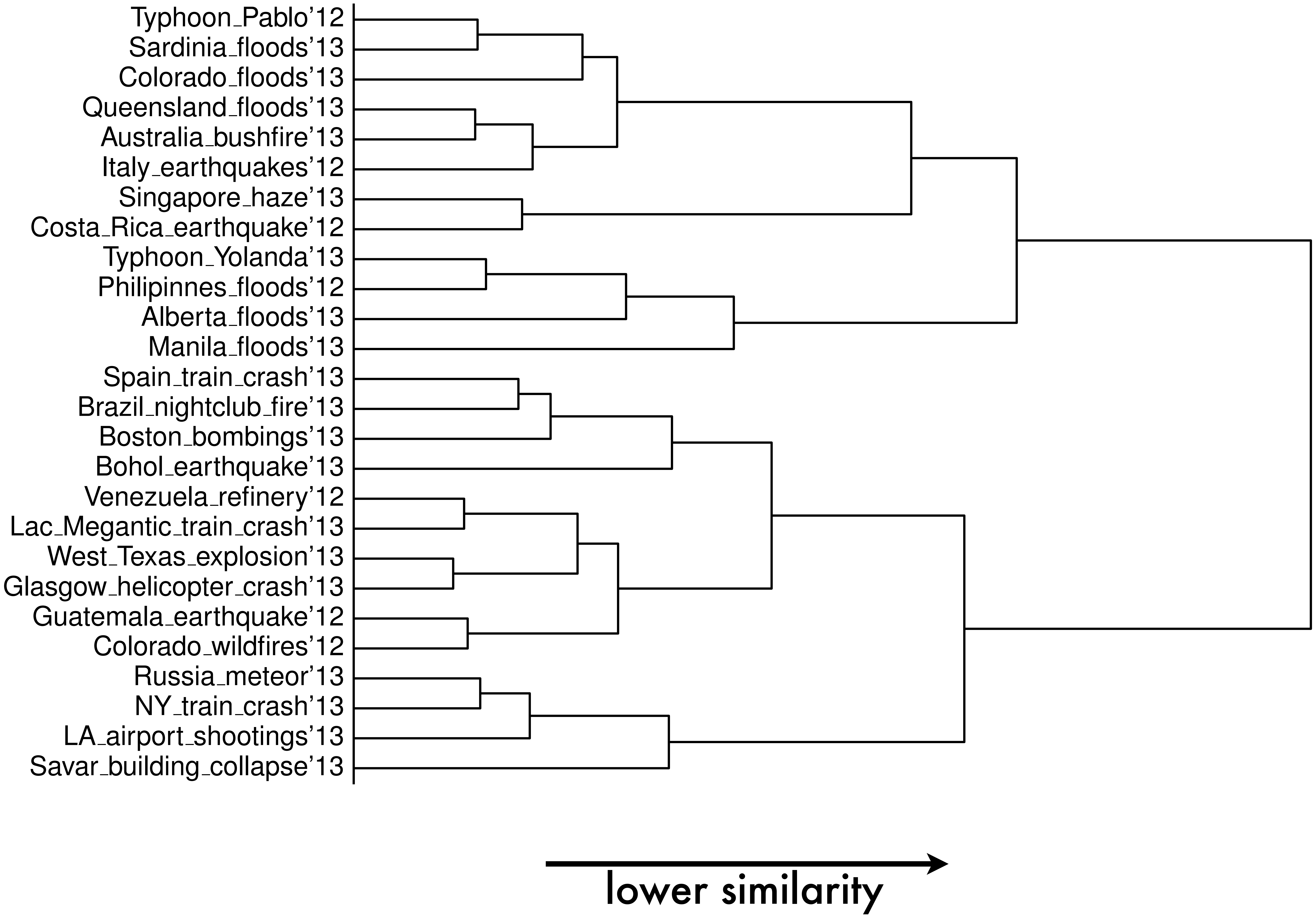




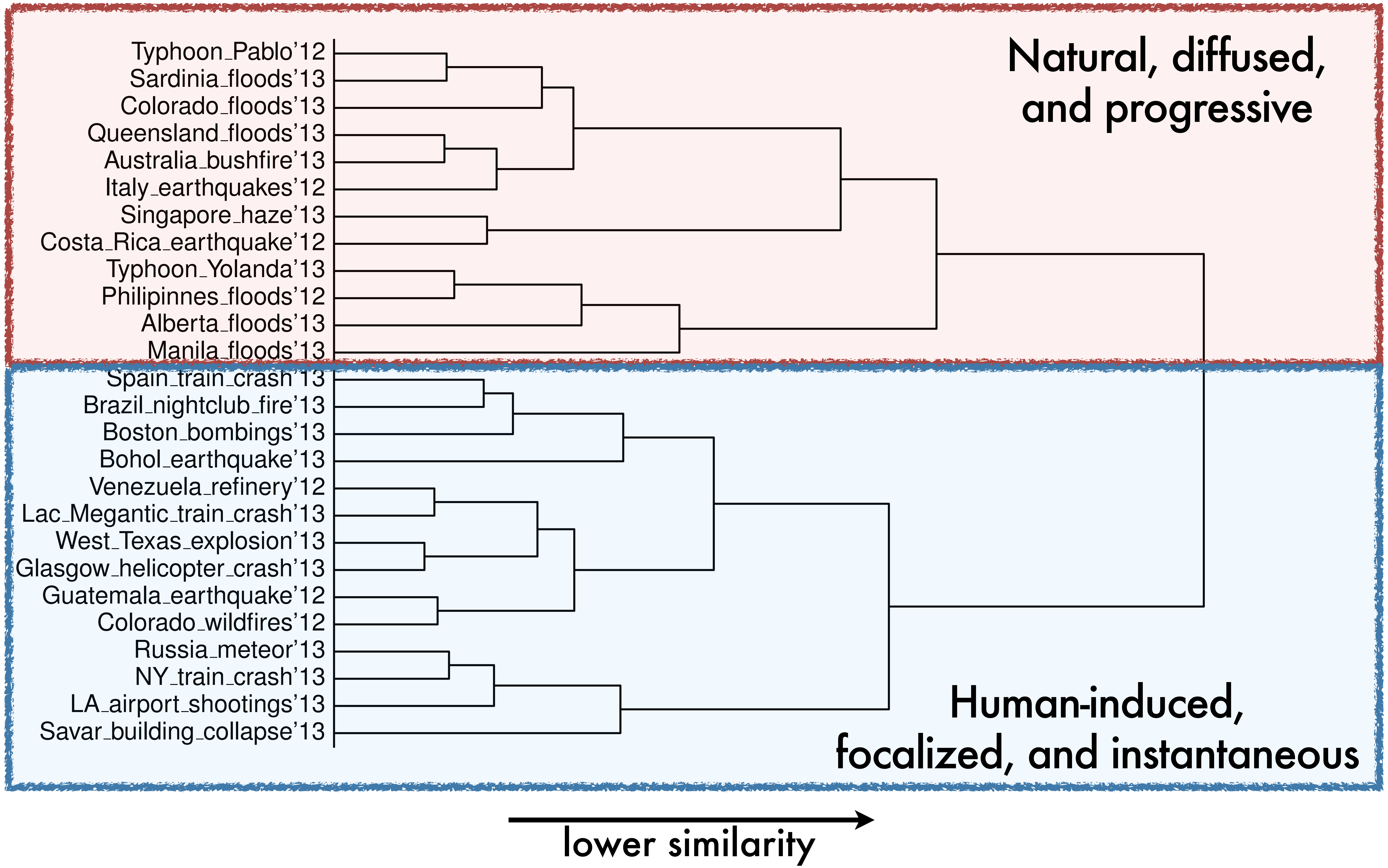
Content & Source Variations



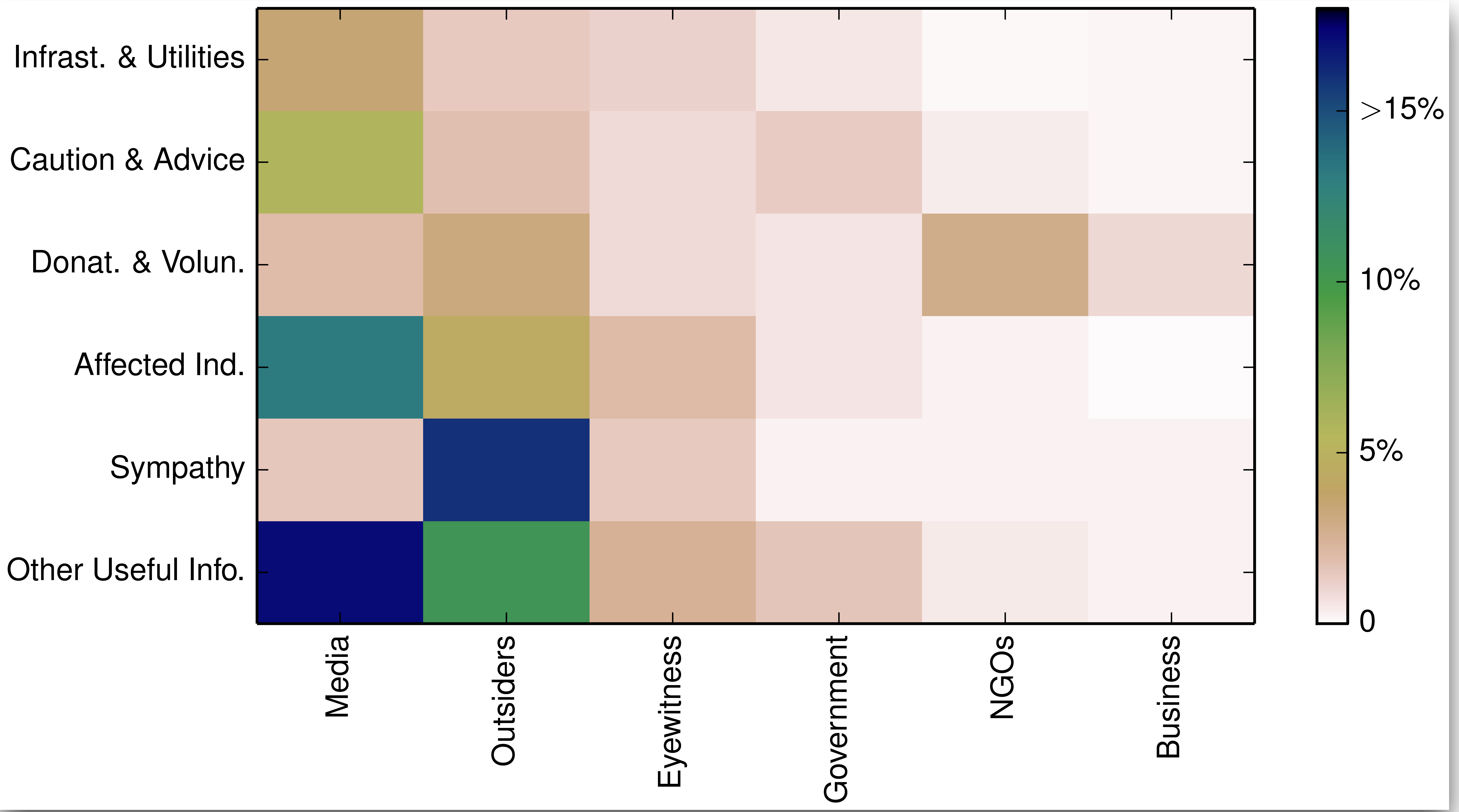
Data Patterns: Types



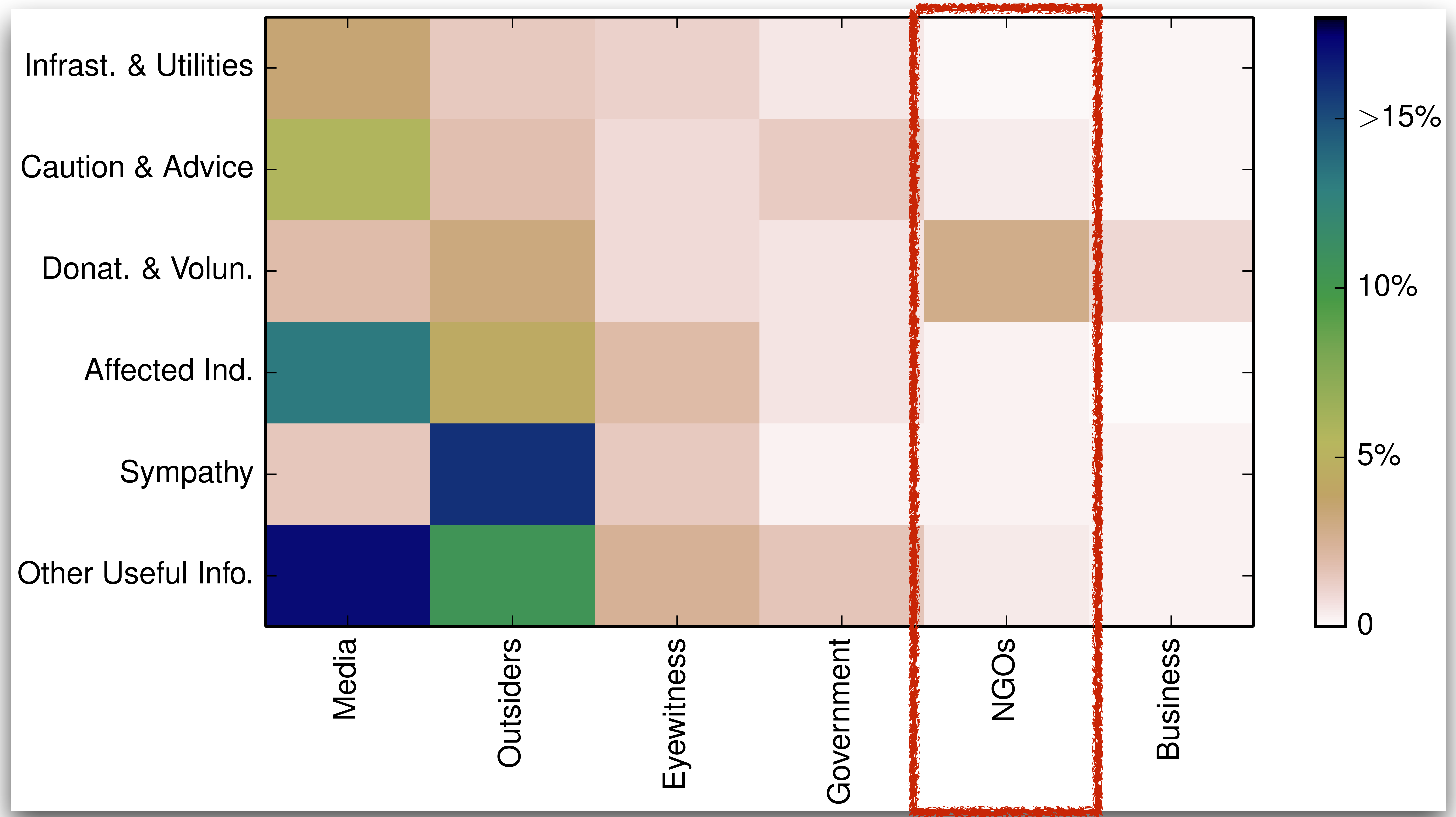
Data Patterns: Types



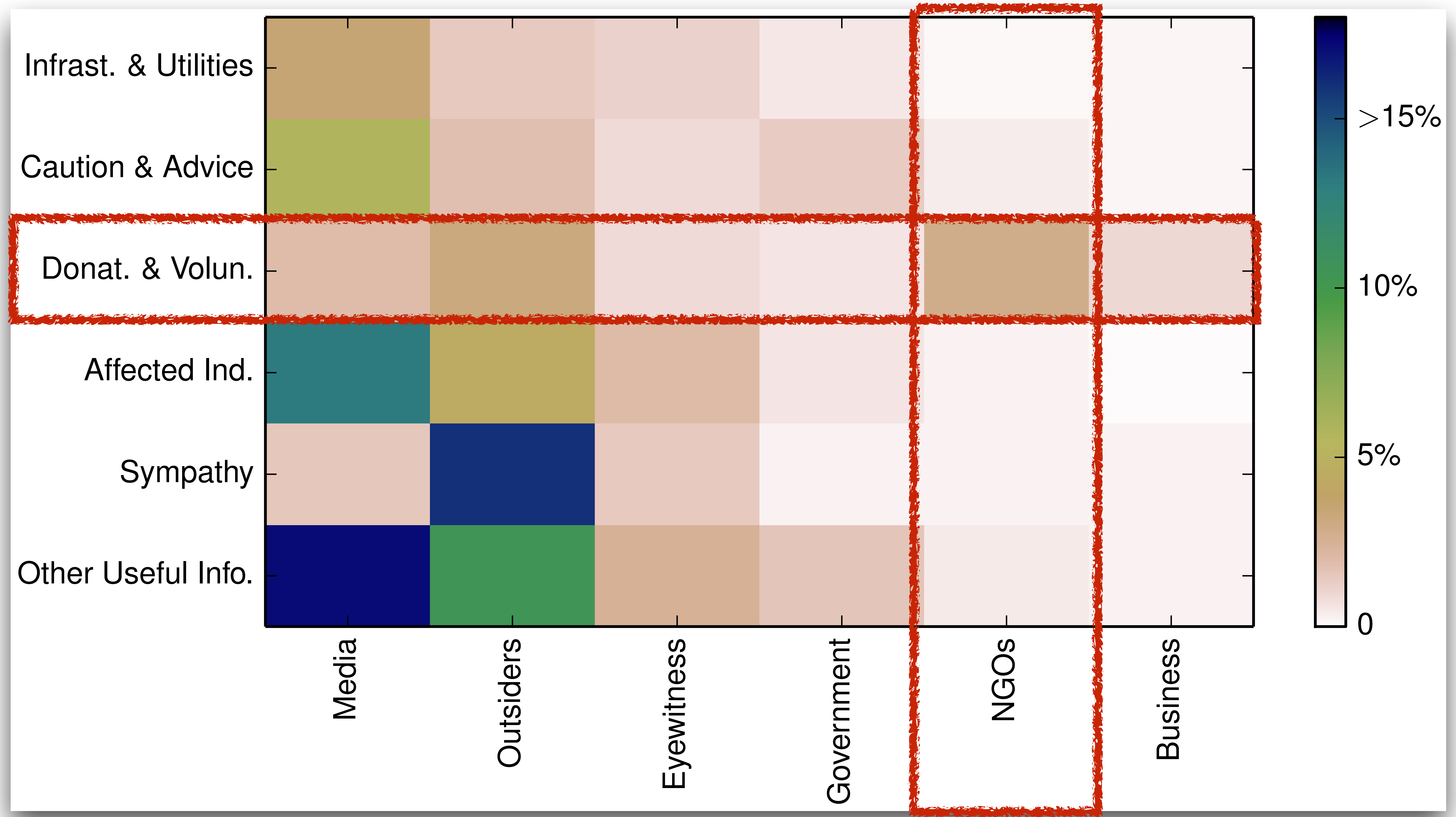
Sources & Message Types



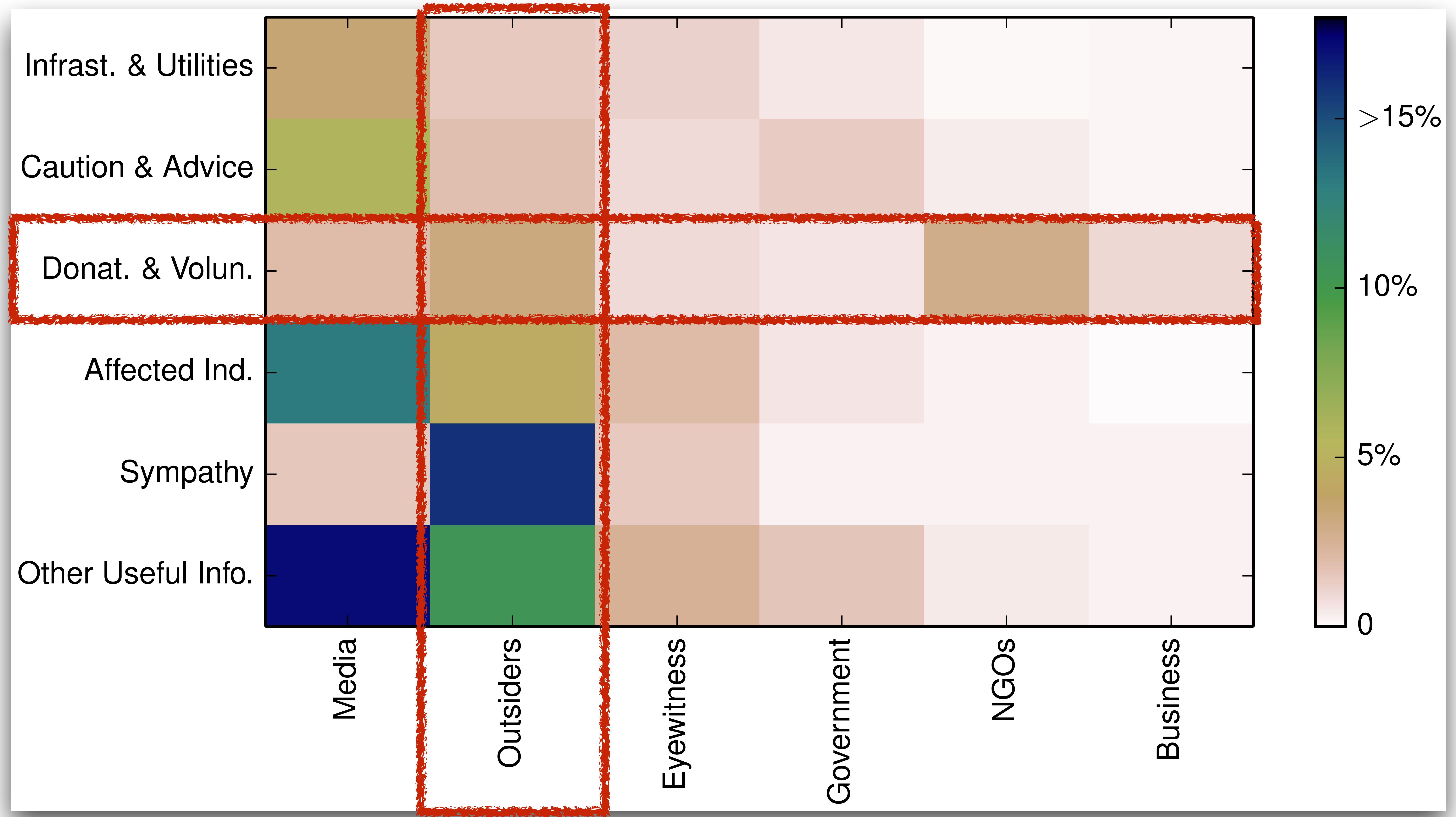
Sources & Message Types



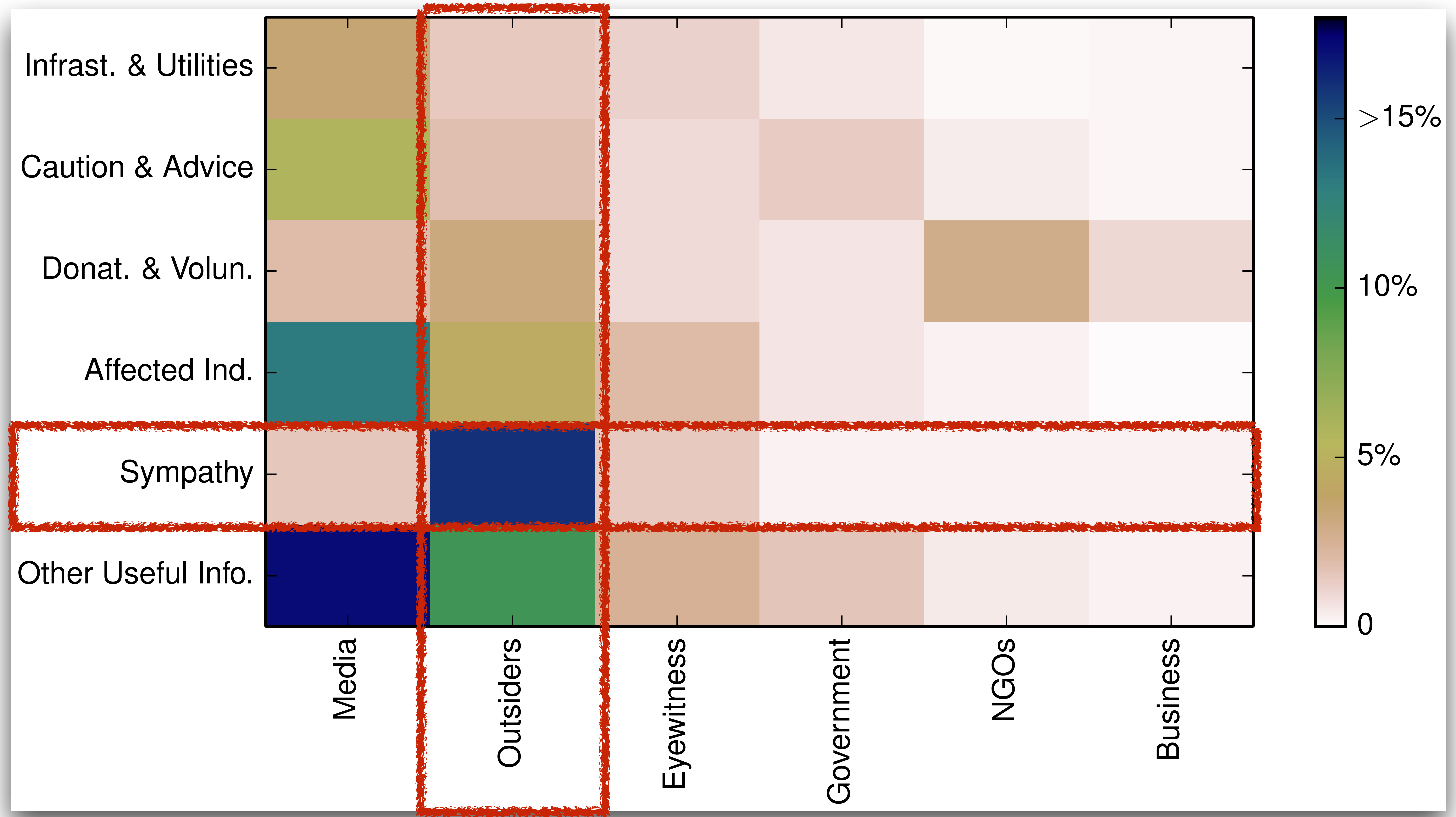
Sources & Message Types



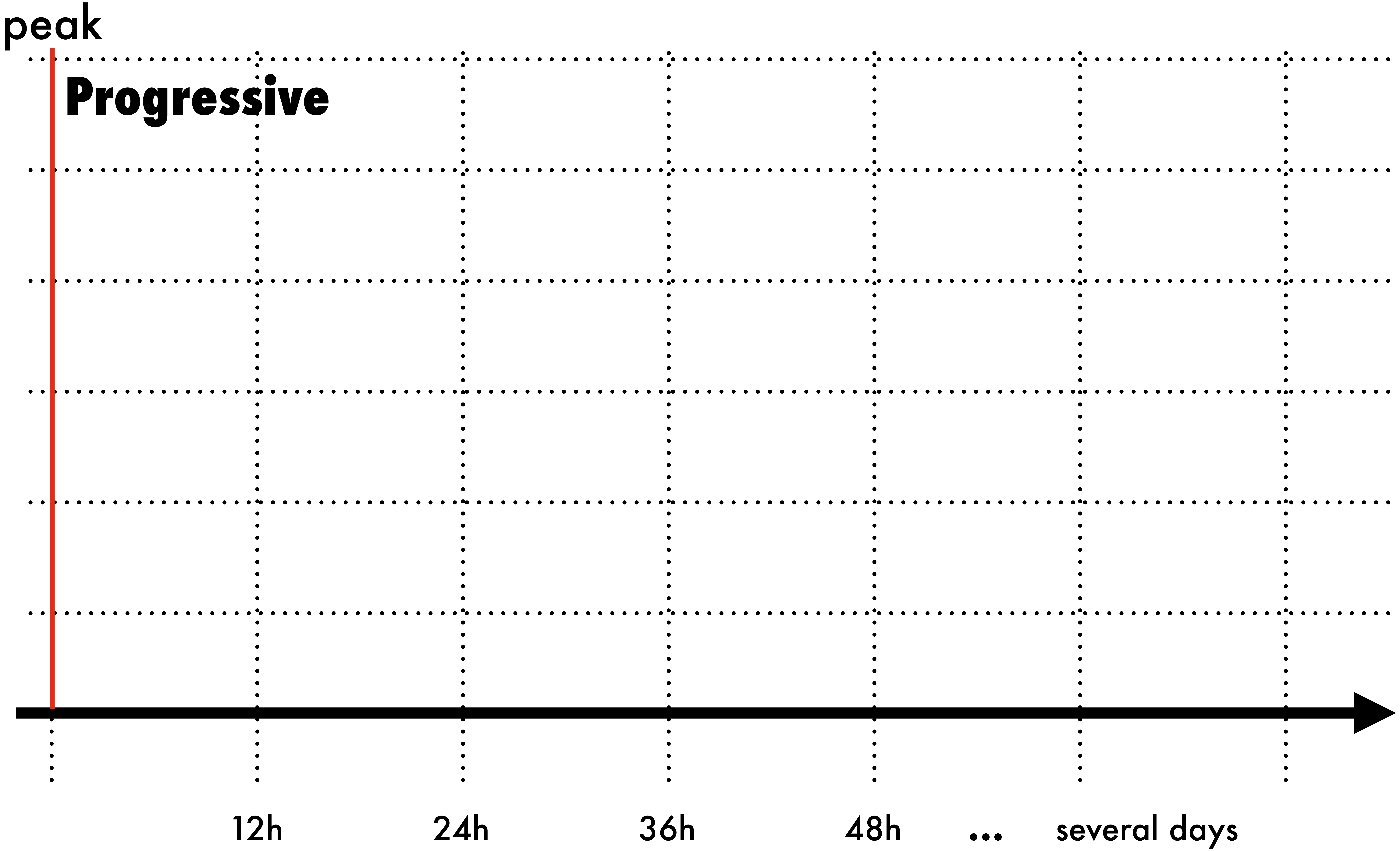
Sources & Message Types



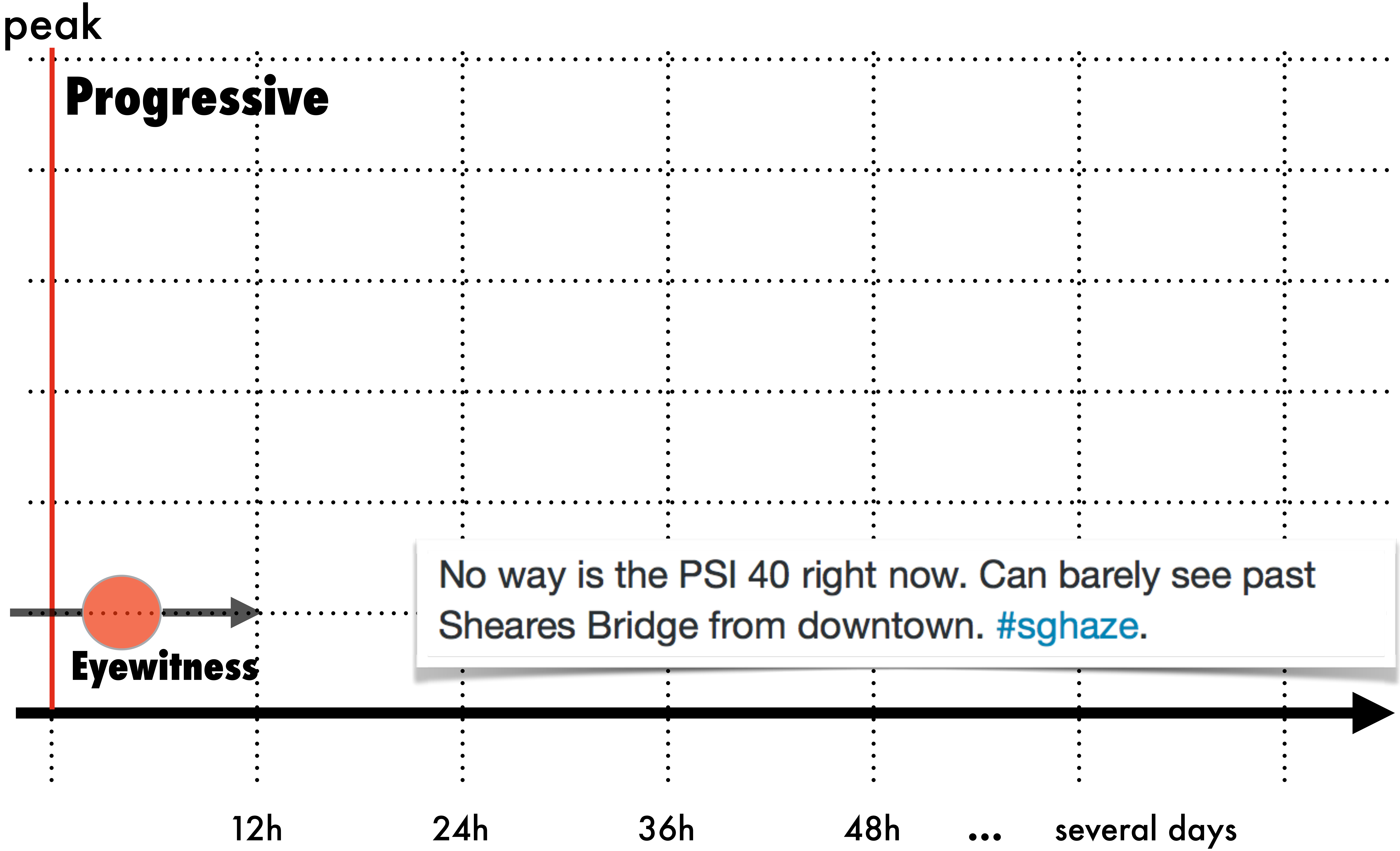
Sources & Message Types



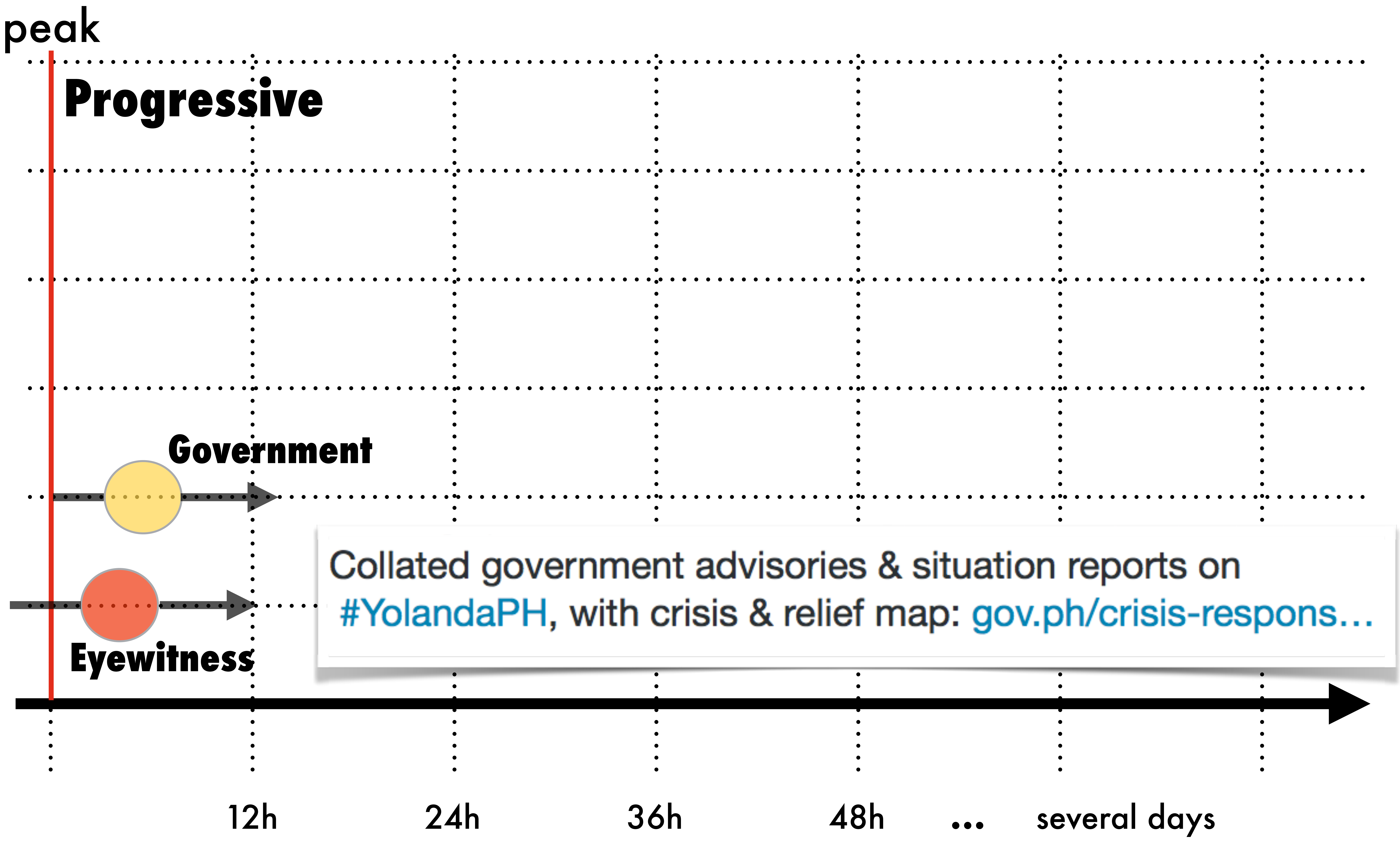
Temporal Distribution: Sources



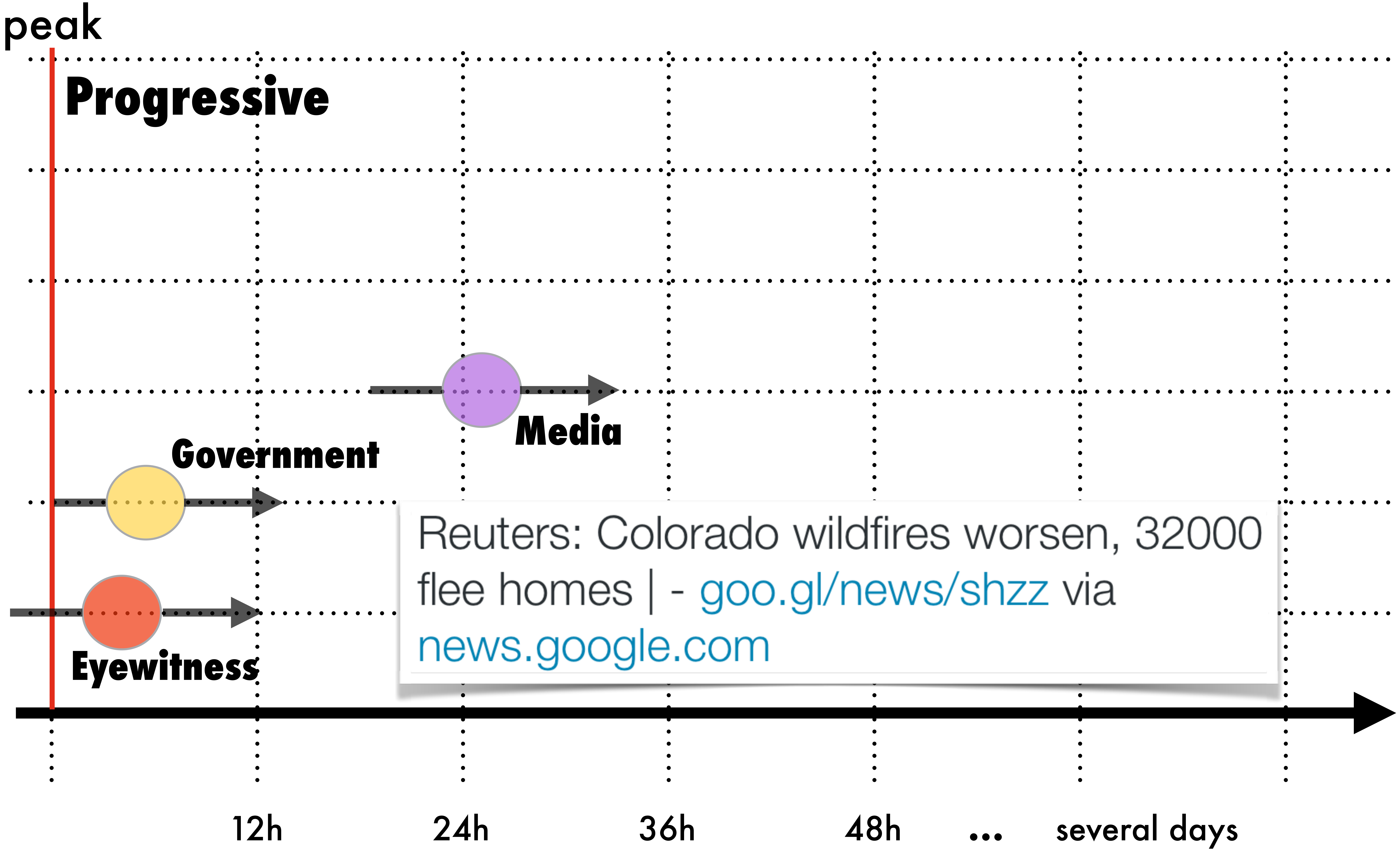
Temporal Distribution: Sources



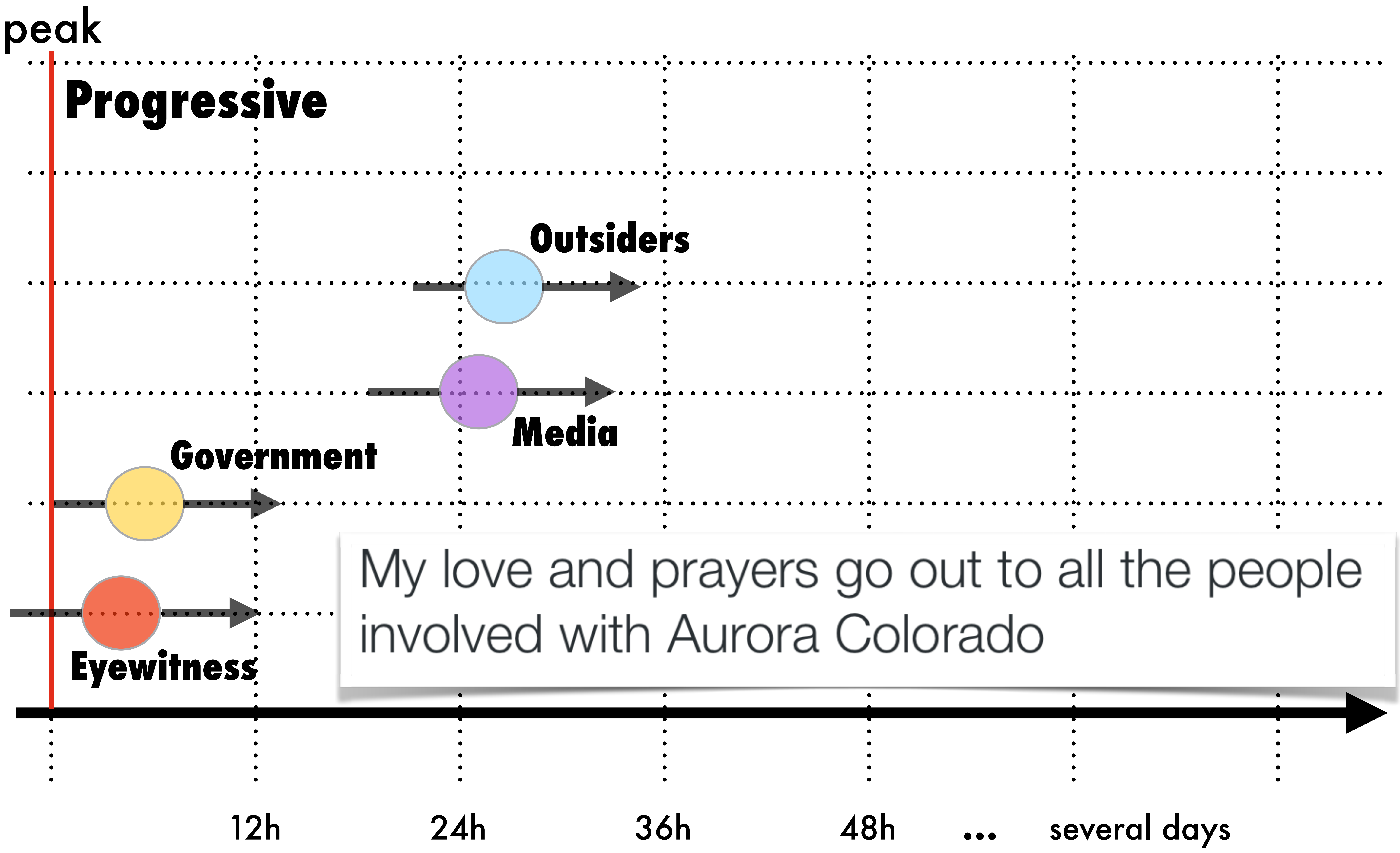
Temporal Distribution: Sources



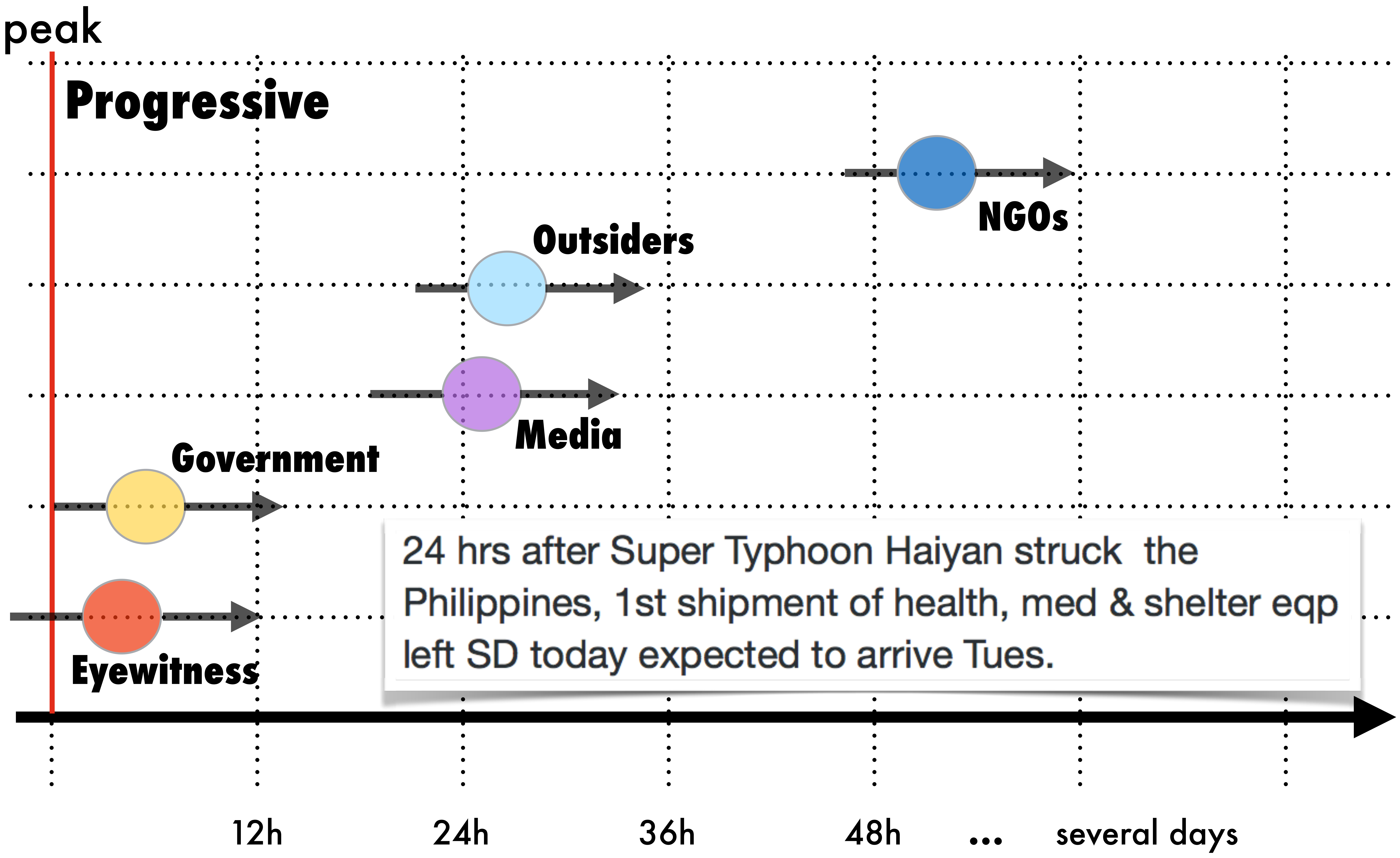
Temporal Distribution: Sources



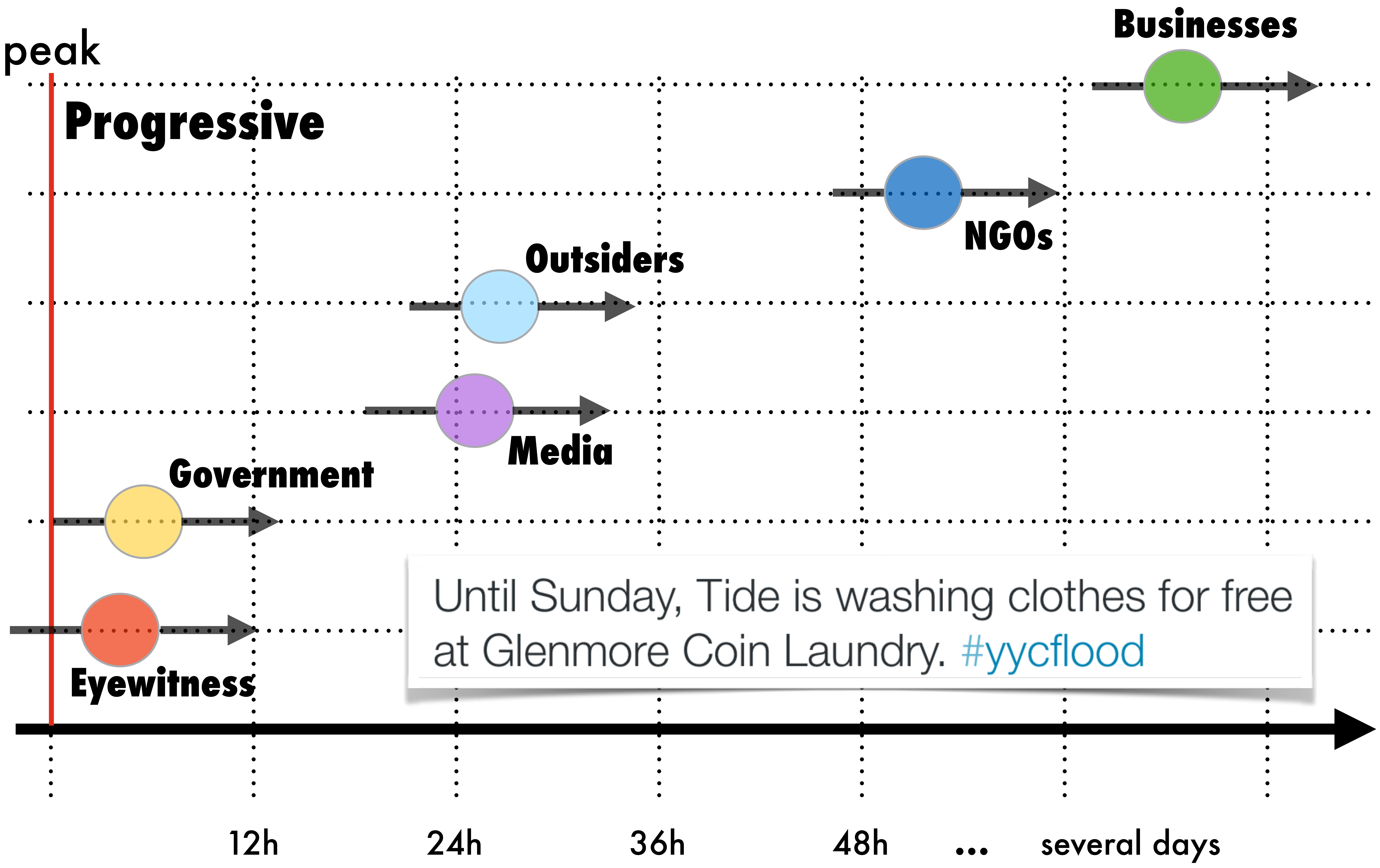
Temporal Distribution: Sources



Temporal Distribution: Sources



Temporal Distribution: Sources



Why It Matters?

UNOCHA World Humanitarian Data and Trends 2014

Annotated datasets used by 100+ new studies (see crisislex.org)

Lexicons used by e.g., GDELT to annotate news

We need better data collection pipelines

We need to better understand what factors shape the datasets at origin

<u>**Climate change**/Media coverage bias</u>

Do social and mainstream media differ in their coverage of climate change?

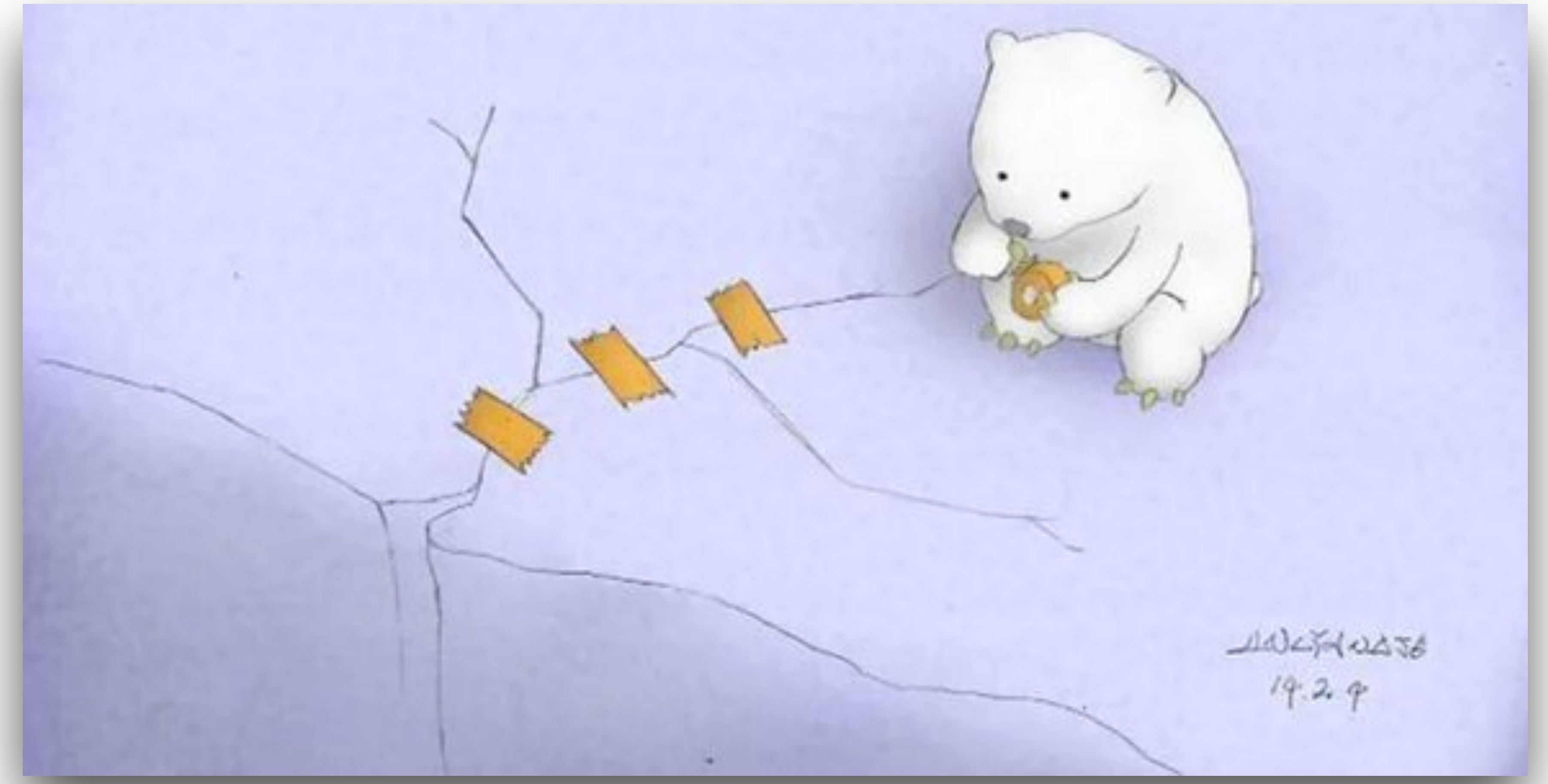
Is social media a good proxy for some phenomena of interest?

with **Carlos Castillo**, **Nick Diakopoulos**, and **Karl Aberer** [ICWSM'15]

Operational Definition

"A change of climate which is attributed directly or indirectly to **human activity** that alters the composition of the **global atmosphere** and which is in addition to natural **climate variability** observed over comparable time periods."

United Nations Framework for Climate Change

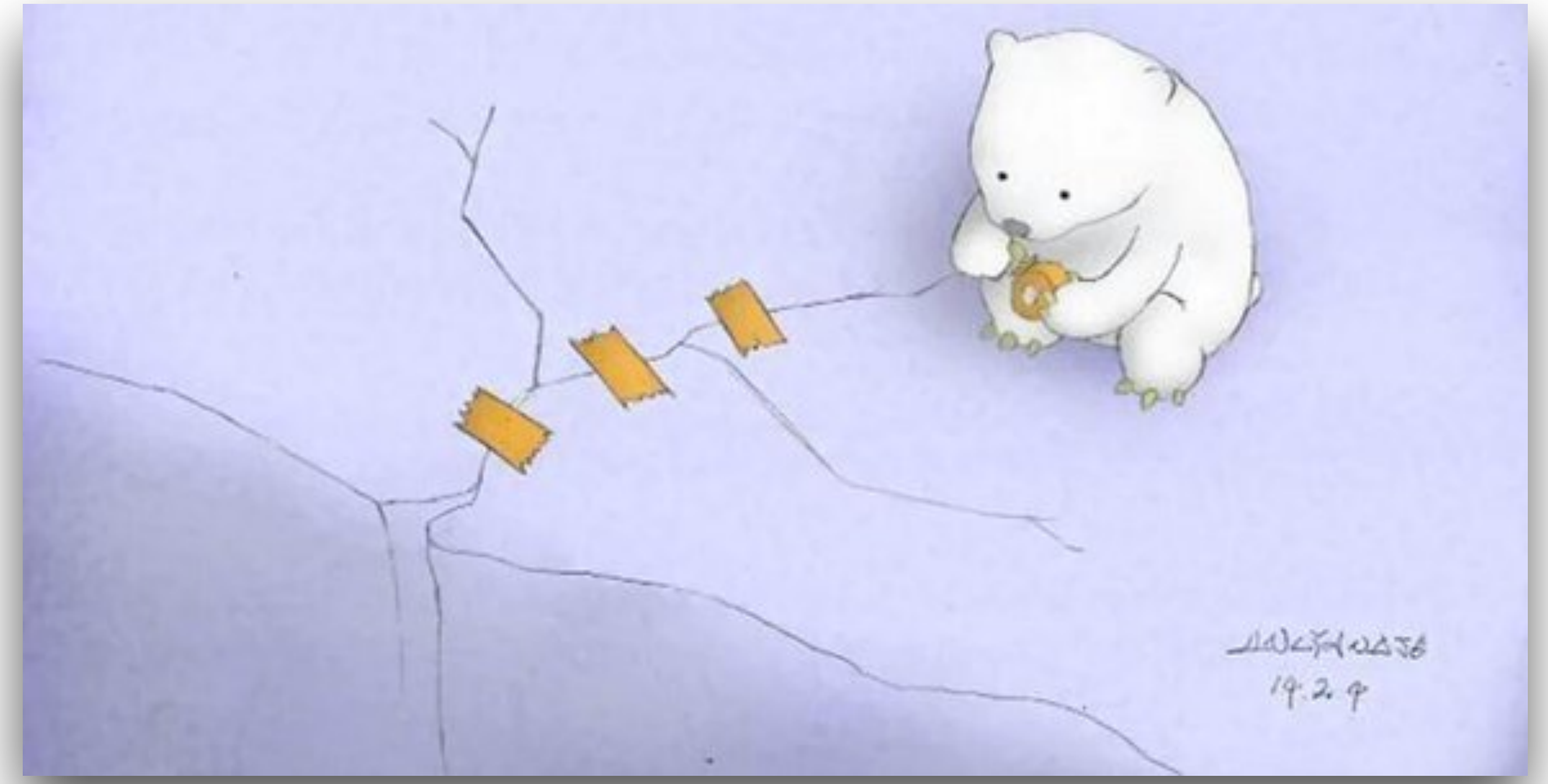


Operational Definition

"A change of climate which is attributed directly or indirectly to **human activity** that alters the composition of the **global atmosphere** and which is in addition to natural **climate variability** observed over comparable time periods."

United Nations Framework for Climate Change

- ✓ defines a problem
- ✓ identifies its causes
- ✓ makes a moral judgement
- ✓ suggests a remedy



Analysis Pipeline

Highly specific query (keywords, themes)

Data Sources (e.g. URLs, Tweets)

Peak detection over coverage time series

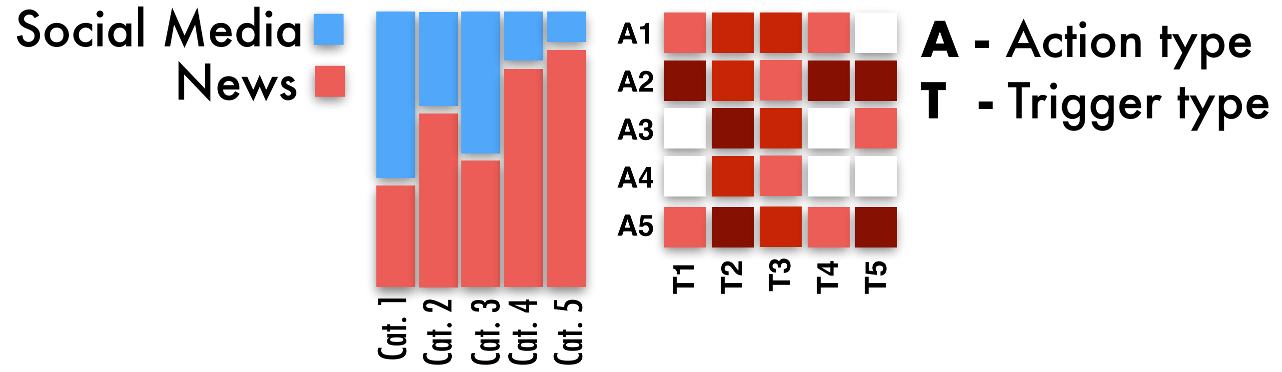
Candidate Events

Manual annotations/ Crowdsourcing

Climate-change Events

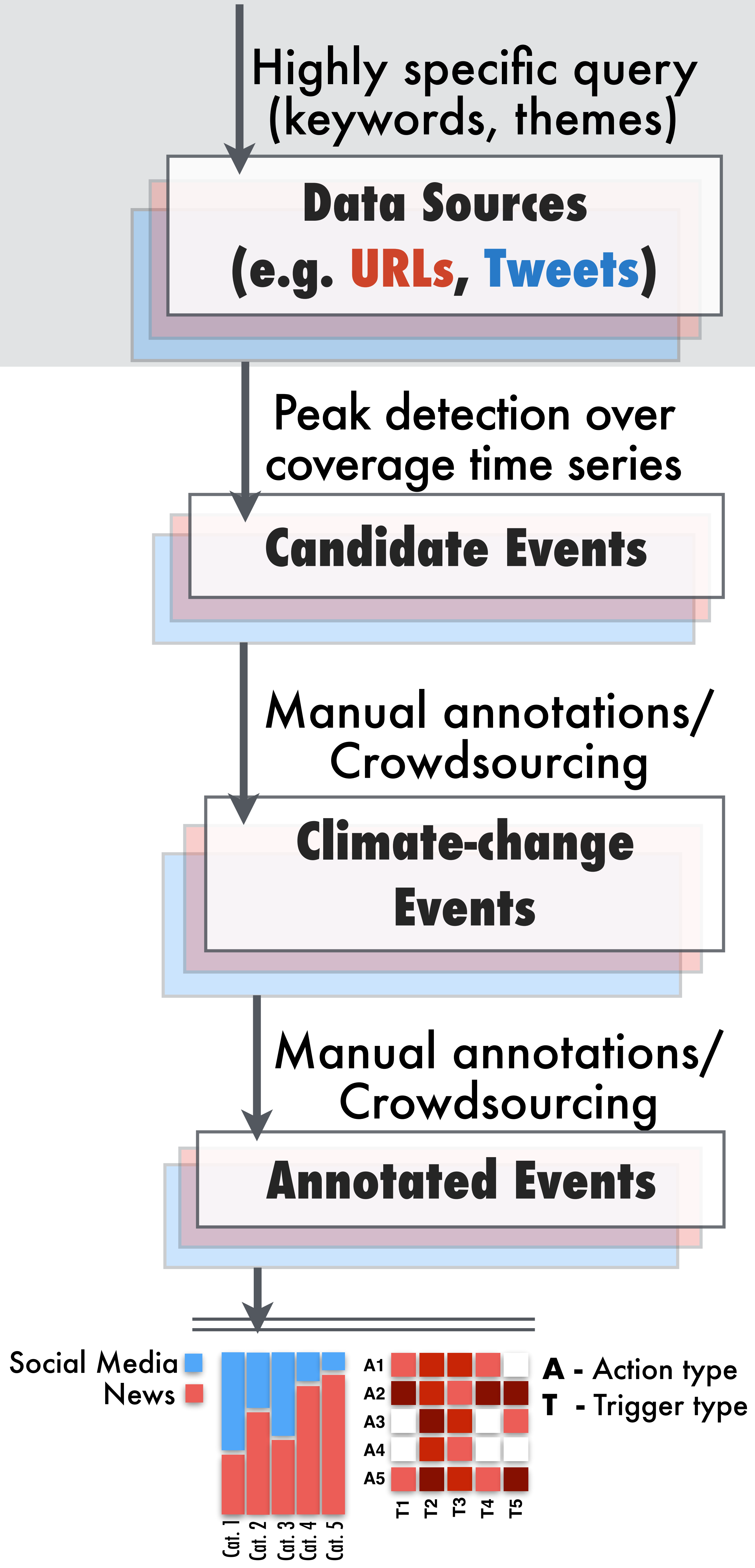
Manual annotations/ Crowdsourcing

Annotated Events



Analysis Pipeline

1. Domain Data

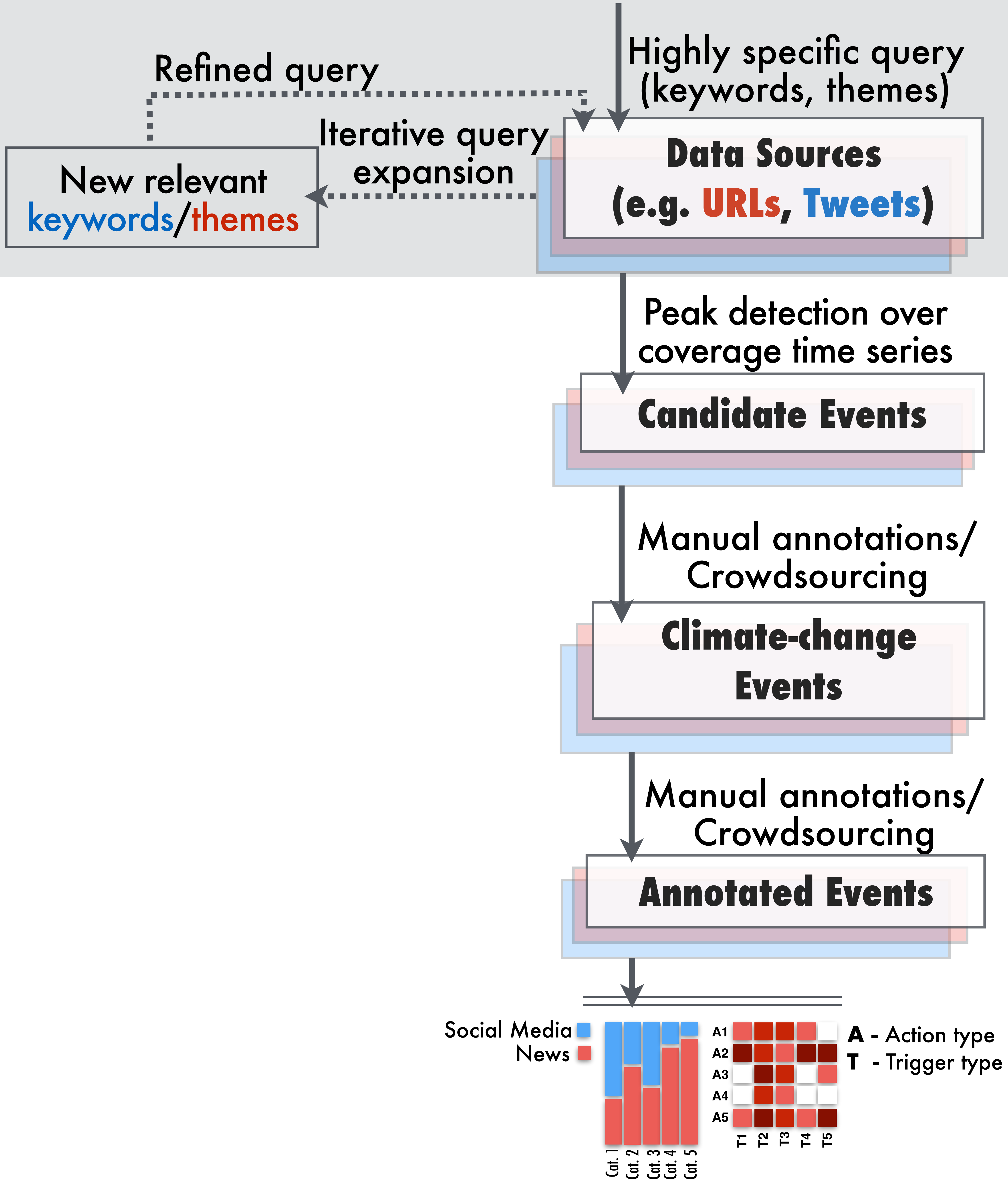


Analysis Pipeline

1. Domain Data

#ClimateChange
#cop21, rising seas, ...

Env_ClimateChange
Env_CarbonCapture, ...



Analysis Pipeline

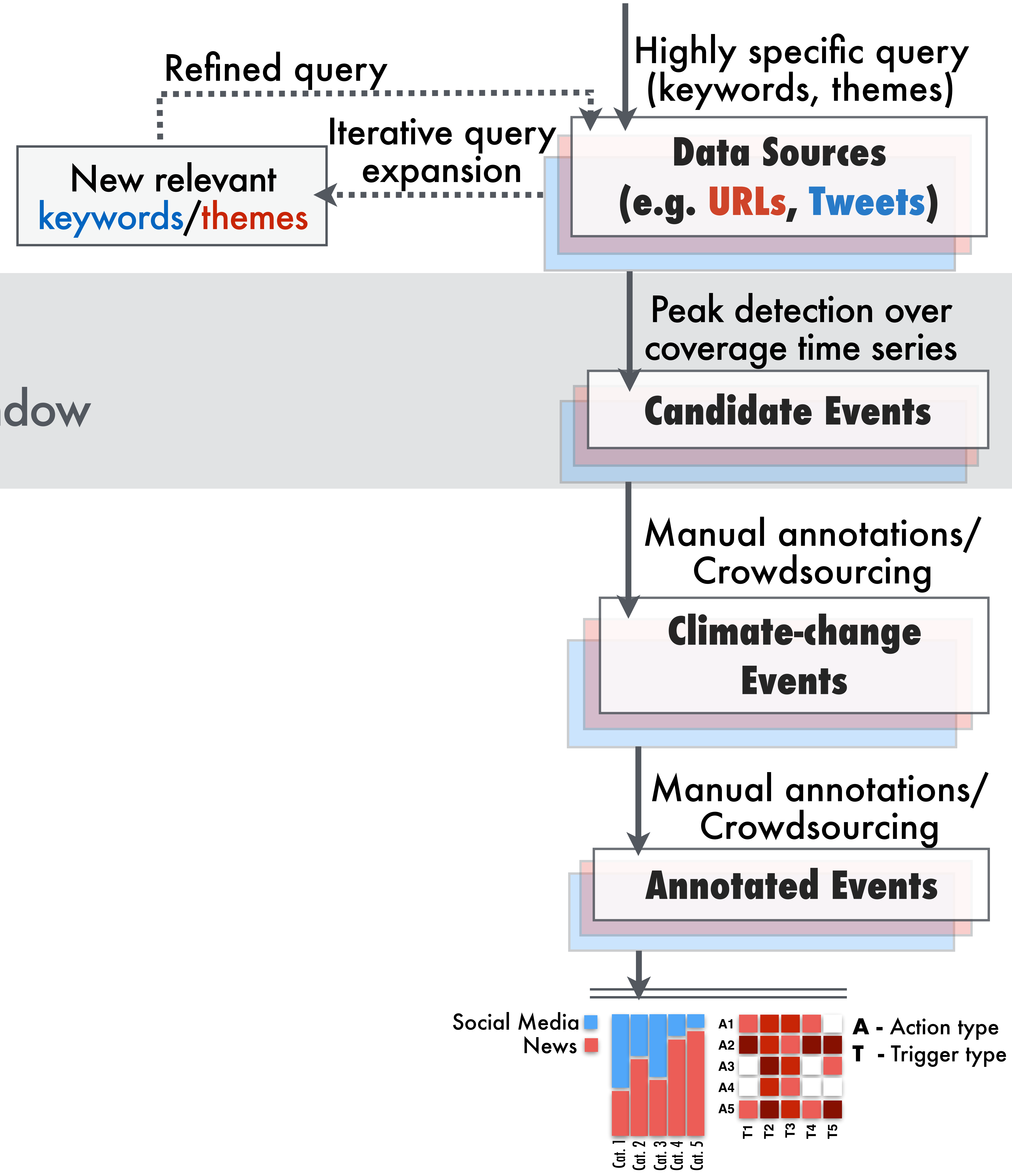
1. Domain Data

#ClimateChange
#cop21, rising seas, …

Env_ClimateChange
Env_CarbonCapture, …

2. Automated Event

Detects spikes within a month-long time window
[Lehmann et al., WWW'12].



Analysis Pipeline

1. Domain Data

#ClimateChange
#cop21, rising seas, ...

Env_ClimateChange
Env_CarbonCapture, ...

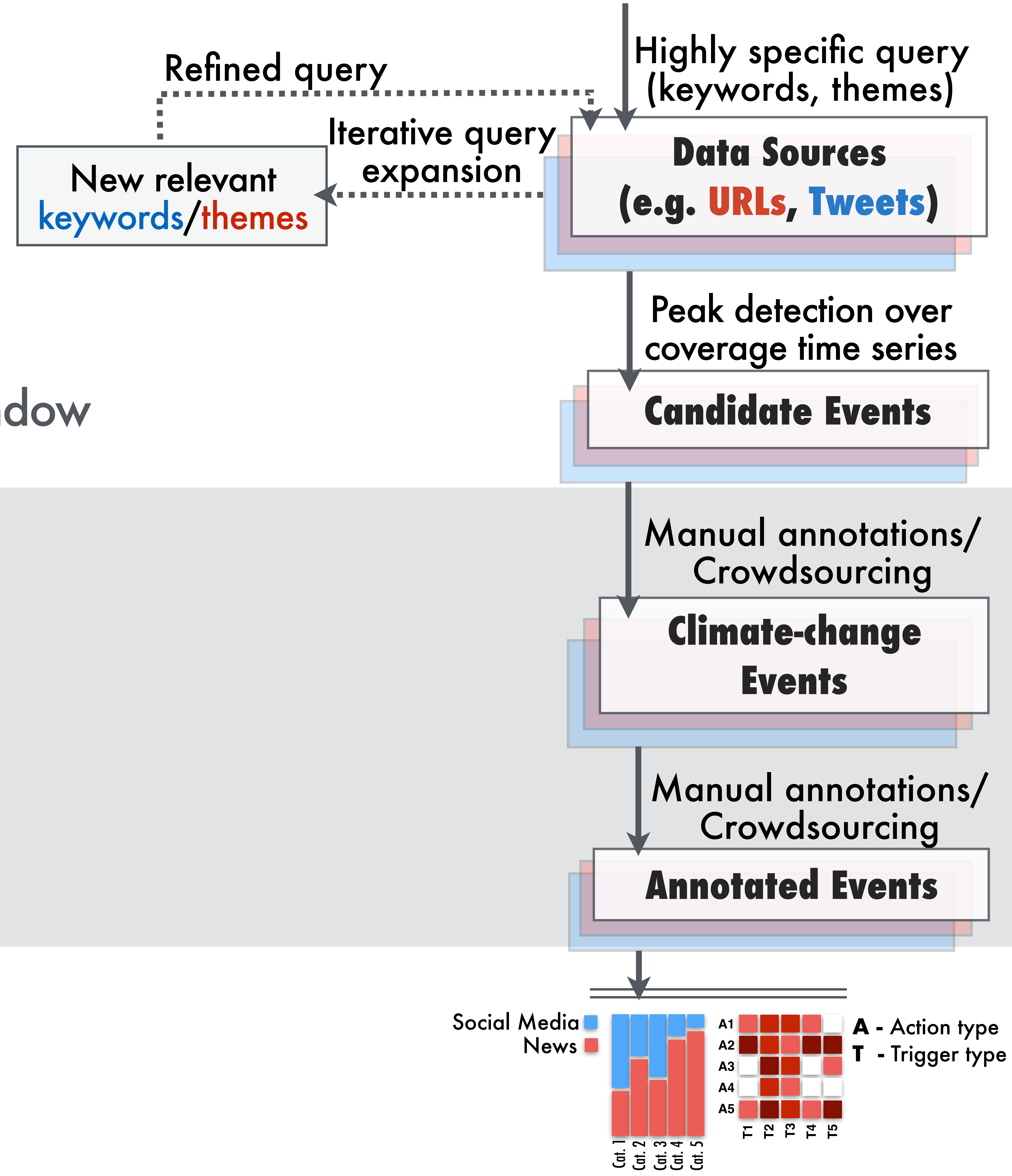
2. Automated Event

Detects spikes within a month-long time window
[Lehmann et al., WWW'12].

3. Events Curation &

Merge duplicates & remove
ambiguous or not-relate events.

Categorize events.



Analysis Pipeline

1. Domain Data

#ClimateChange
#cop21, rising seas, …

Env_ClimateChange
Env_CarbonCapture, …

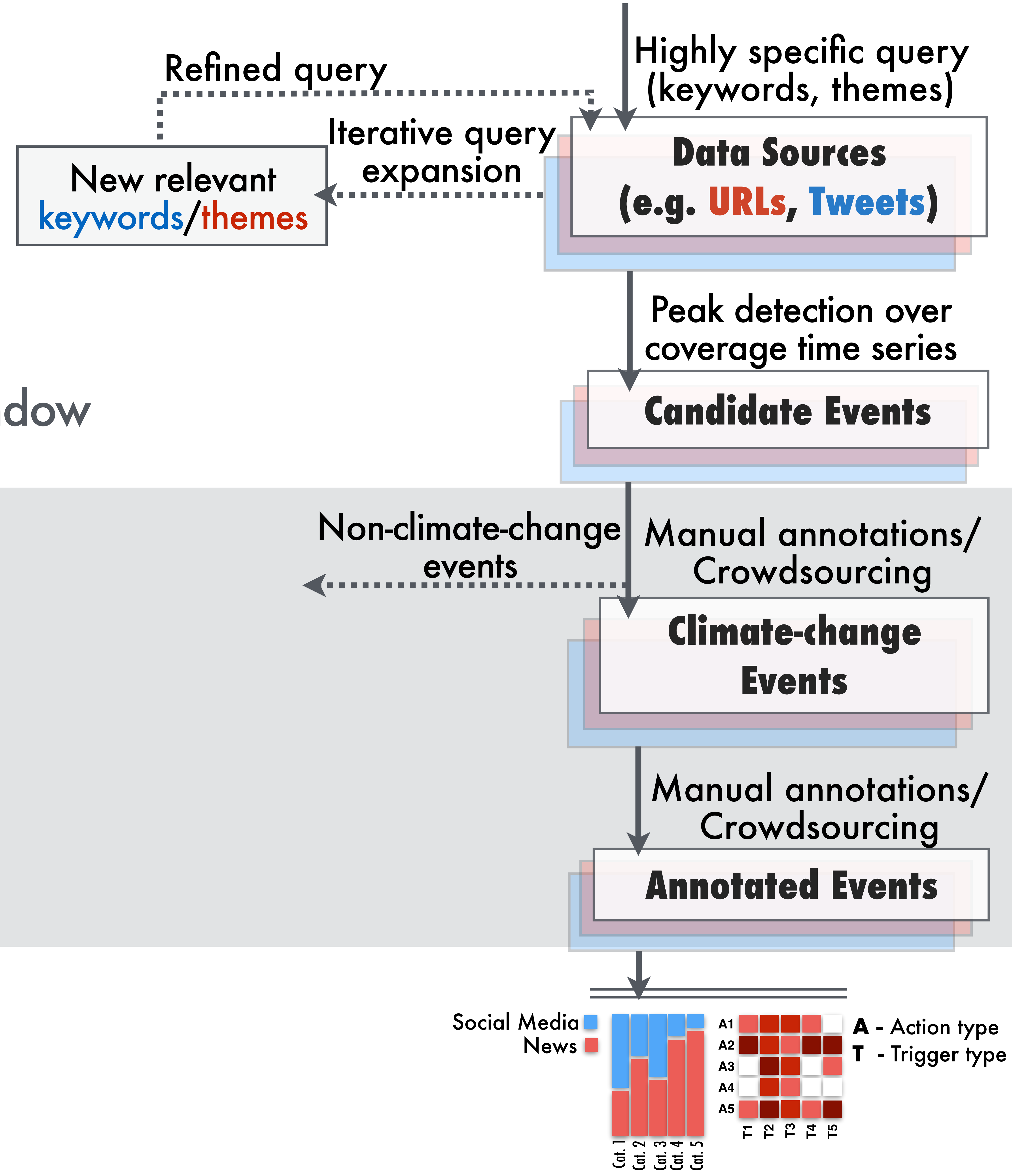
2. Automated Event

Detects spikes within a month-long time window [Lehmann et al., WWW'12].

3. Events Curation &

Merge duplicates & remove ambiguous or not-relate events.

Categorize events.



Analysis Pipeline

1. Domain Data

#ClimateChange
#cop21, rising seas, …

Env_ClimateChange
Env_CarbonCapture, …

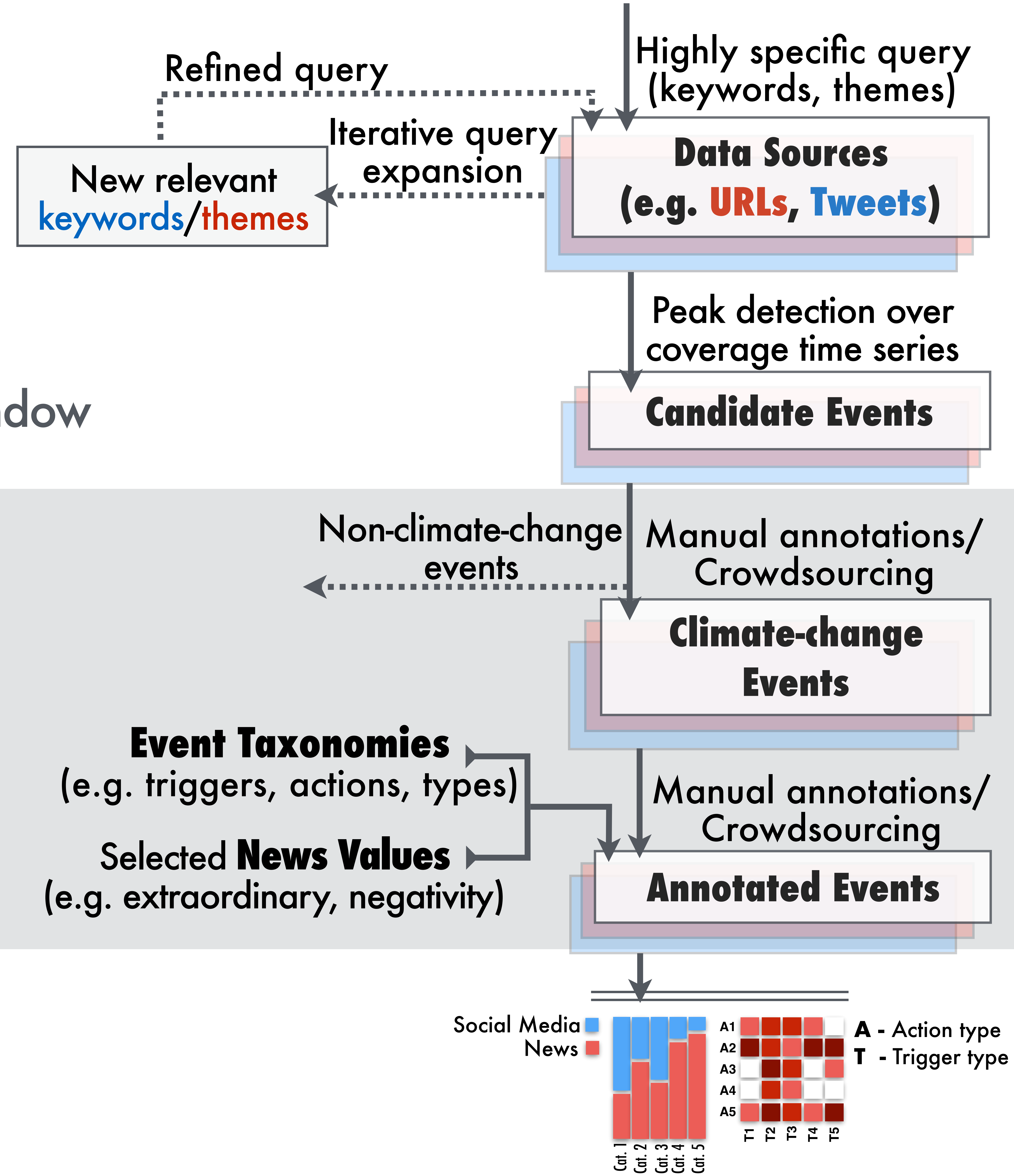
2. Automated Event

Detects spikes within a month-long time window
[Lehmann et al., WWW'12].

3. Events Curation &

Merge duplicates & remove ambiguous or not-relate events.

Categorize events.



Analysis Pipeline

1. Domain Data

#ClimateChange
#cop21, rising seas, …

Env_ClimateChange
Env_CarbonCapture, …

2. Automated Event

Detects spikes within a month-long time window
[Lehmann et al., WWW'12].

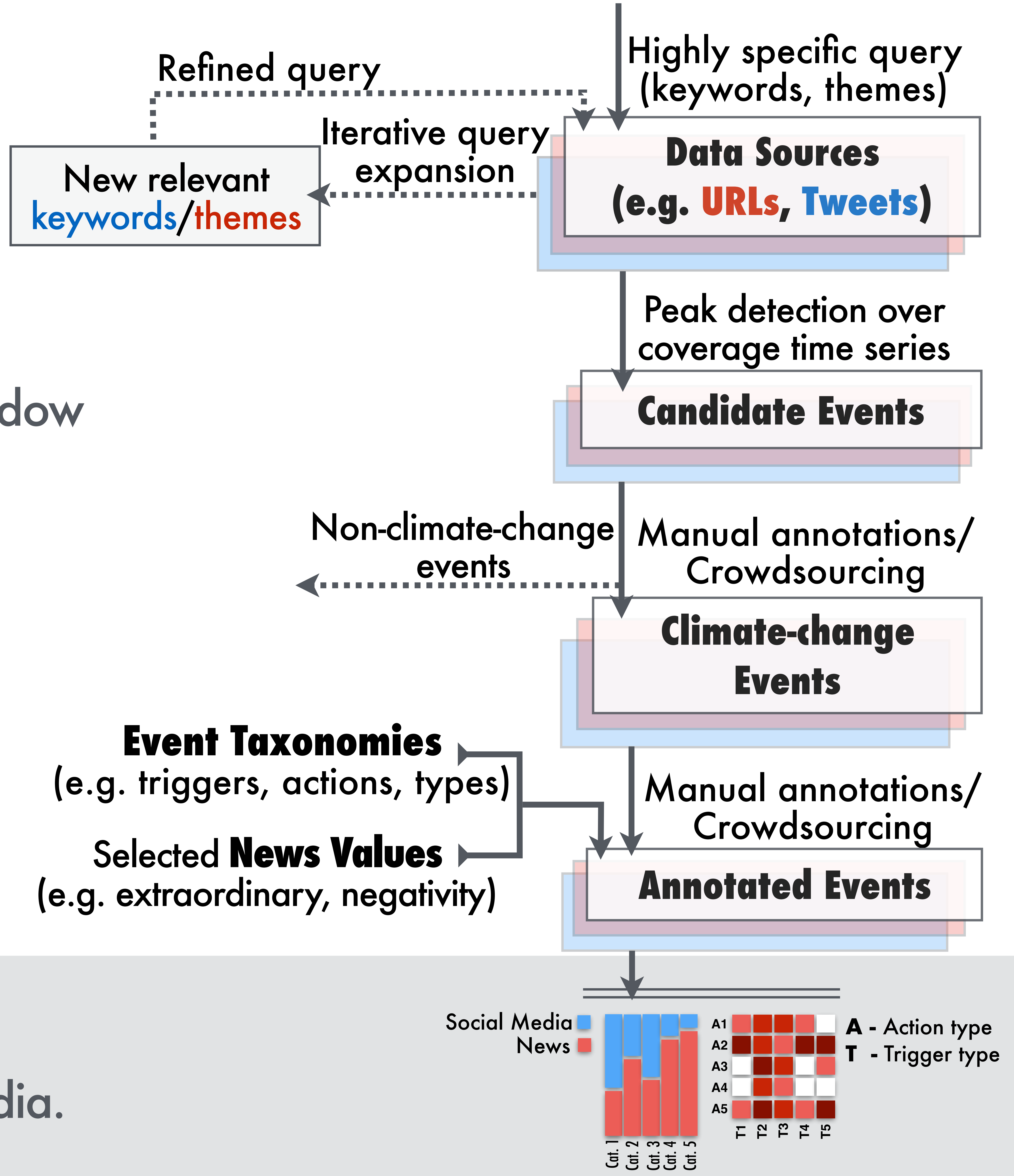
3. Events Curation &

Merge duplicates & remove ambiguous or not-relate events.

Categorize events.

4. Data

Event types prevalence across the two media.



Data Collections

17 months in 2013/2014

Social Media (Twitter)

1% stream
(via Internet Archive)

~**2 billion** tweets

+

240 terms
(e.g., rise sea, #acidification)

↓

~480,000

↓

428 peaks/**111 events**

News Media (GDELT)

major international, national,
regional, and local news sources

~**30 million** news articles

+

40 themes & taxonomies
(e.g., movement_environmental)

↓

~560,000

↓

218 peaks/**100 events**

Data Collections

17 months in 2013/2014

Social Media (Twitter)

1% stream
(via Internet Archive)

~**2 billion** tweets

+

240 terms
(e.g., rise sea, #acidification)

428 peaks/**111 events**

News Media (GDELT)

major international, national,
regional, and local news sources

~**30 million** news articles

+

40 themes & taxonomies
(e.g., movement_environmental)

218 peaks/**100 events**

Only 25 events occur in both!

Coverage Across Media

Prevalence of Actor/Action combination for extensively covered events in the News and Twitter

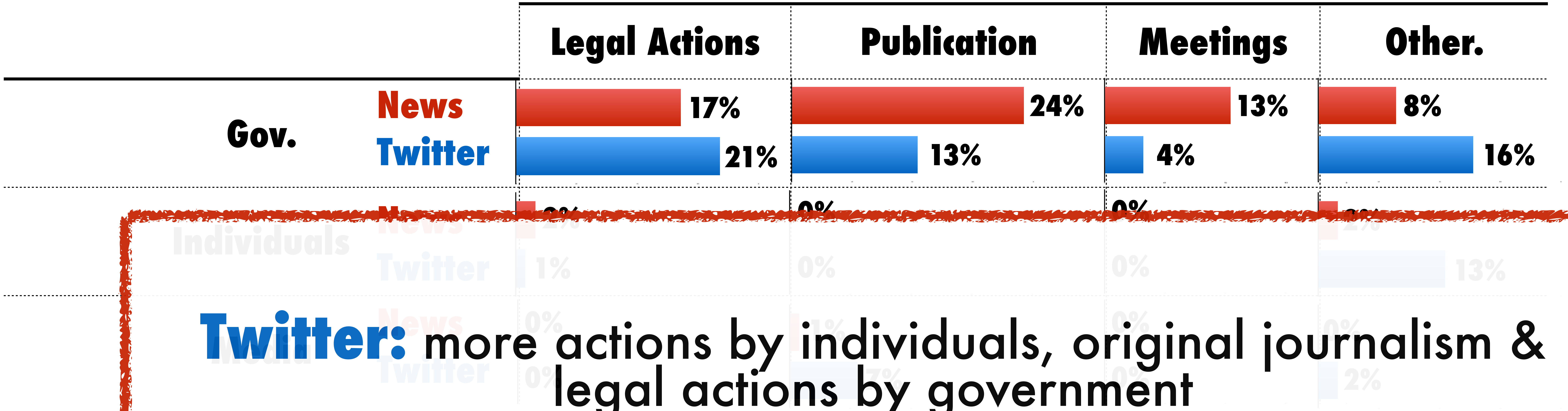
		Legal Actions	Publication	Meetings	Other.
Gov.	News	17%	24%	13%	8%
	Twitter	21%	13%	4%	16%
Individuals	News	2%	0%	0%	2%
	Twitter	1%	0%	0%	13%
Media	News	0%	1%	0%	0%
	Twitter	0%	7%	0%	2%

Prevalence of extensively covered disaster events in the News and Twitter

		Human-Induced Hazards	Natural Hazards
Disaster	News	3%	17%
	Twitter	1%	6%

Coverage Across Media

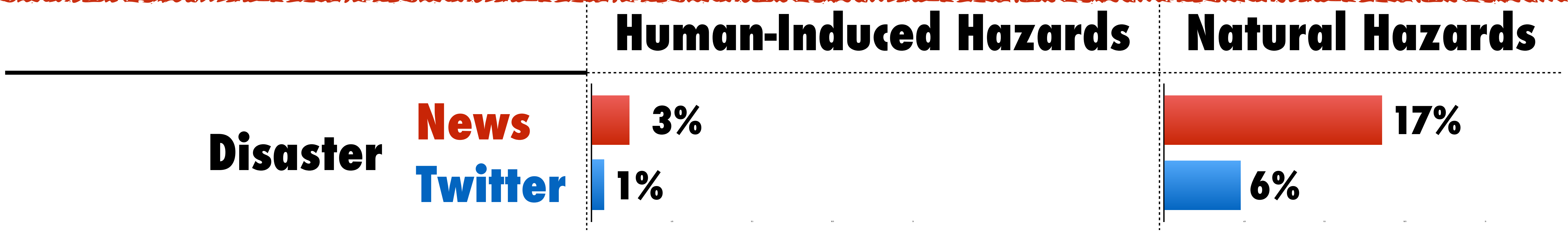
Prevalence of Actor/Action combination for extensively covered events in the News and Twitter



Twitter: more actions by individuals, original journalism & legal actions by government

News: more disasters, publications and governmental meetings

Prevalence of extensively covered disaster events in the News and Twitter



Events Newsworthiness

News Values (how much prominence is given to a news story)

Extraordinary

Unpredictable

High magnitude

Negative

Conflictive

Related to elite persons

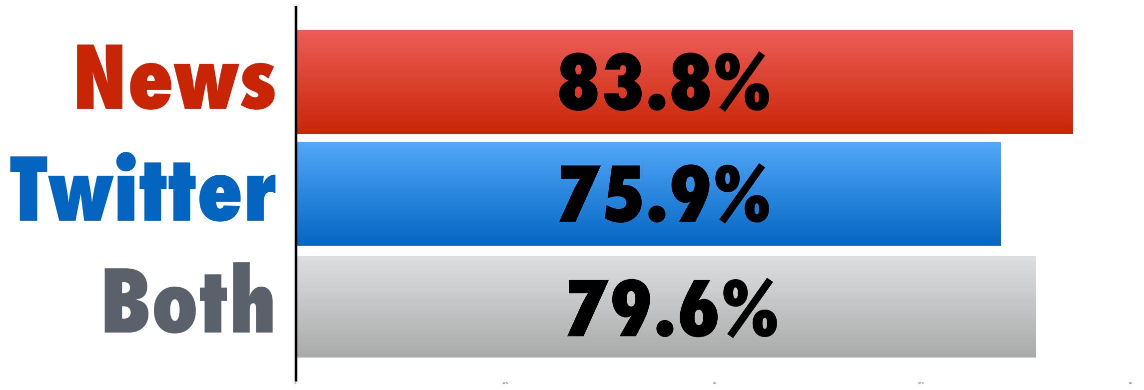
Differences significant at $p<0.05$

Events Newsworthiness

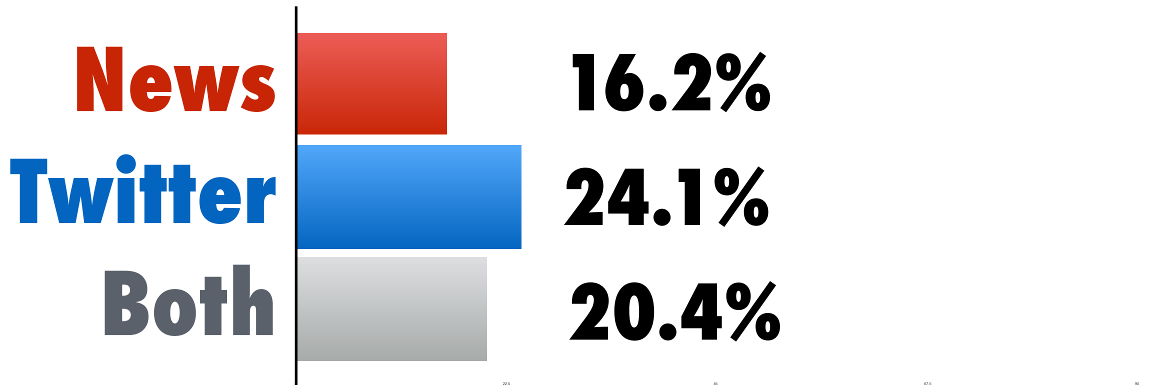
News Values (how much prominence is given to a news story)

Extraordinary

Negative

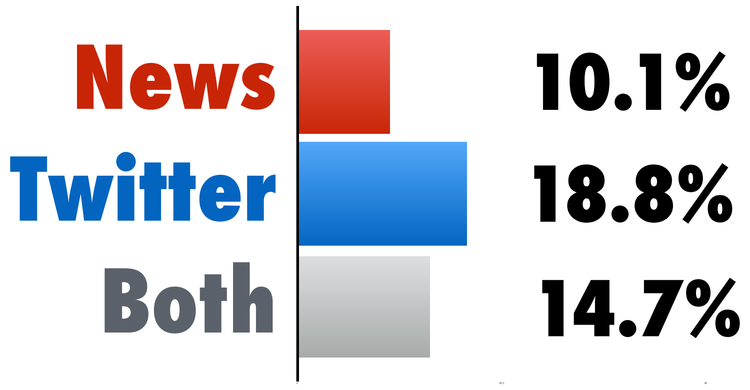


Ordinary

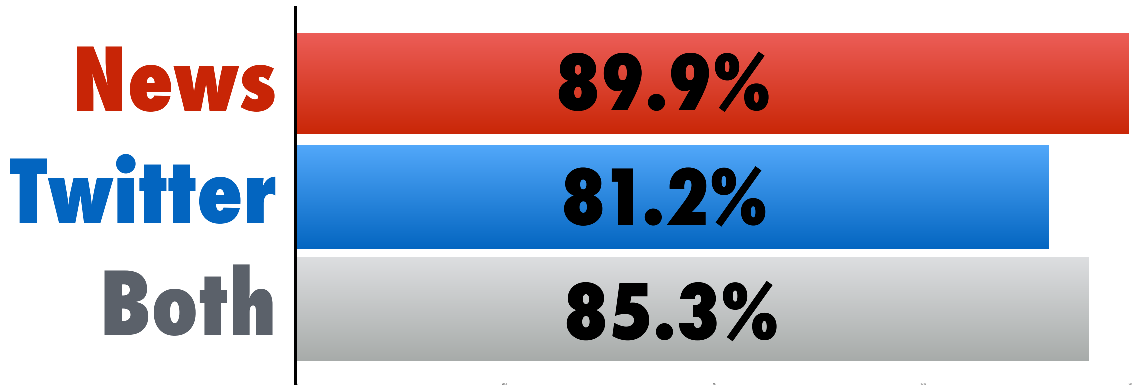


Unpredictable

Conflictive

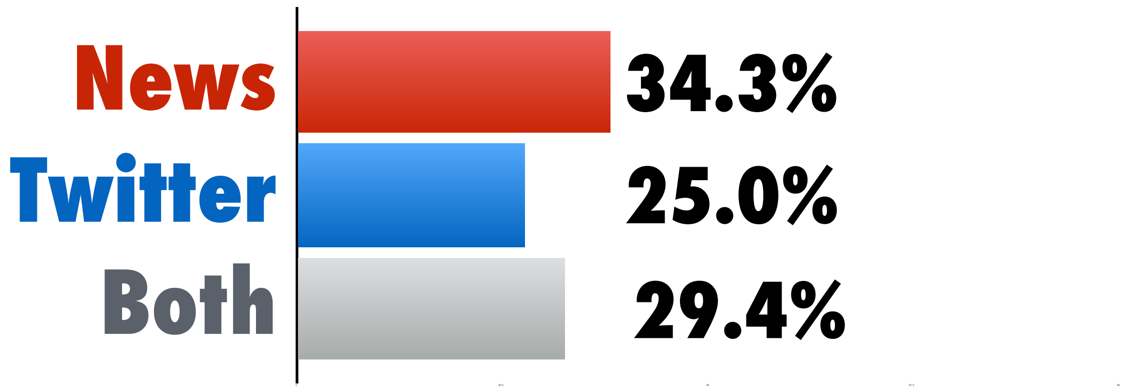


No

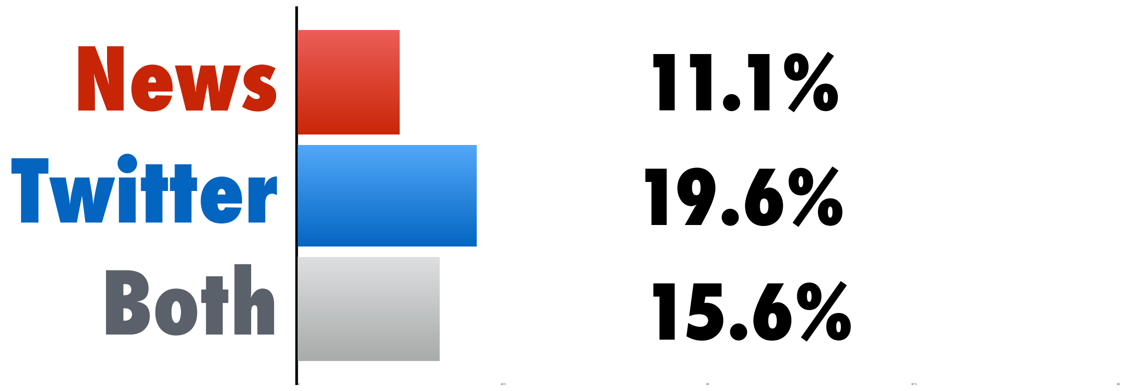


High magnitude

Related to elite persons



Low magnitude



Differences significant at $p<0.05$

Minority issues/BlackLivesMatter movement demographics

What demographics use more the #BlackLivesMatter hashtag on social media?

Is the user sample representative?

with **Ingmar Weber** and **Daniel Gatica-Perez** [AAAI SSS'15]

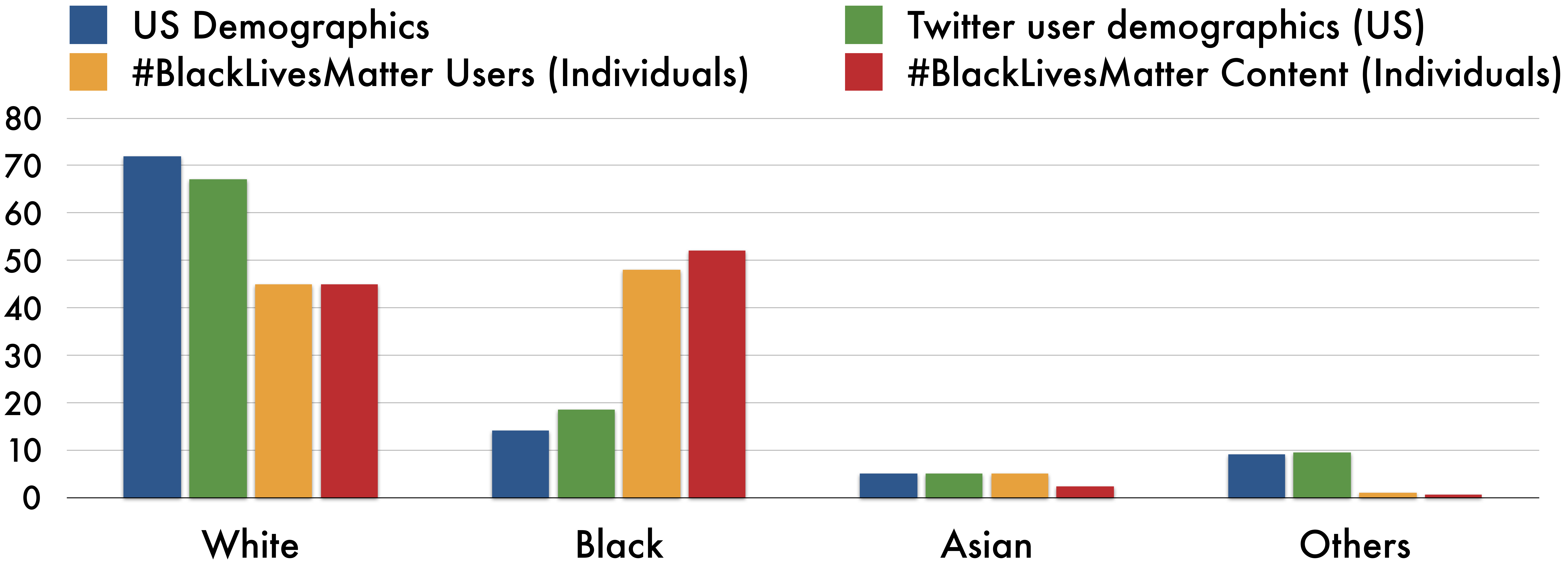
Observing the #BlackLivesMatter Movement

Observing the #BlackLivesMatter Movement

There is a growing number of discussions on minority issues.

Observing the #BlackLivesMatter Movement

There is a growing number of discussions on minority issues.



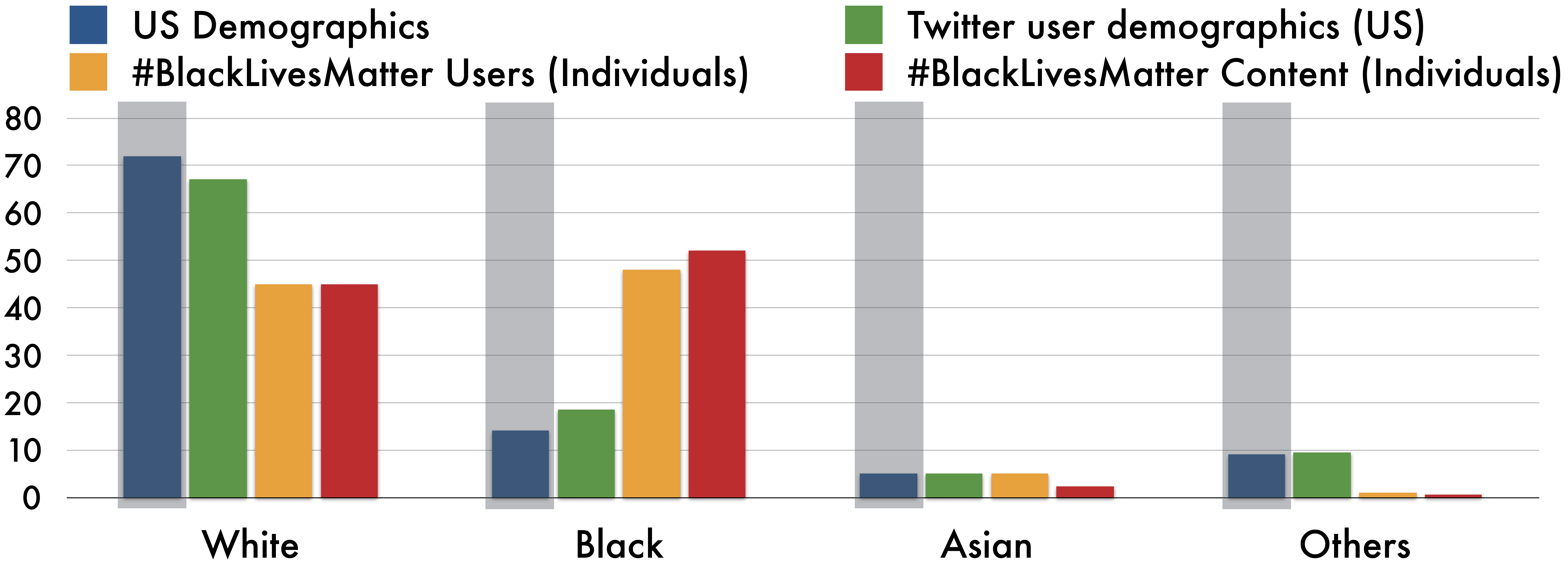
Very rough estimates based on:

- Demographics of social media, Pew Research 2015
- Demographics of the United States in 2010, US census

#BlackLivesMatter: accounts of organizations represent ~5% of all accounts and have a 3 times higher tweeting rate than individuals.

Observing the #BlackLivesMatter Movement

There is a growing number of discussions on minority issues.



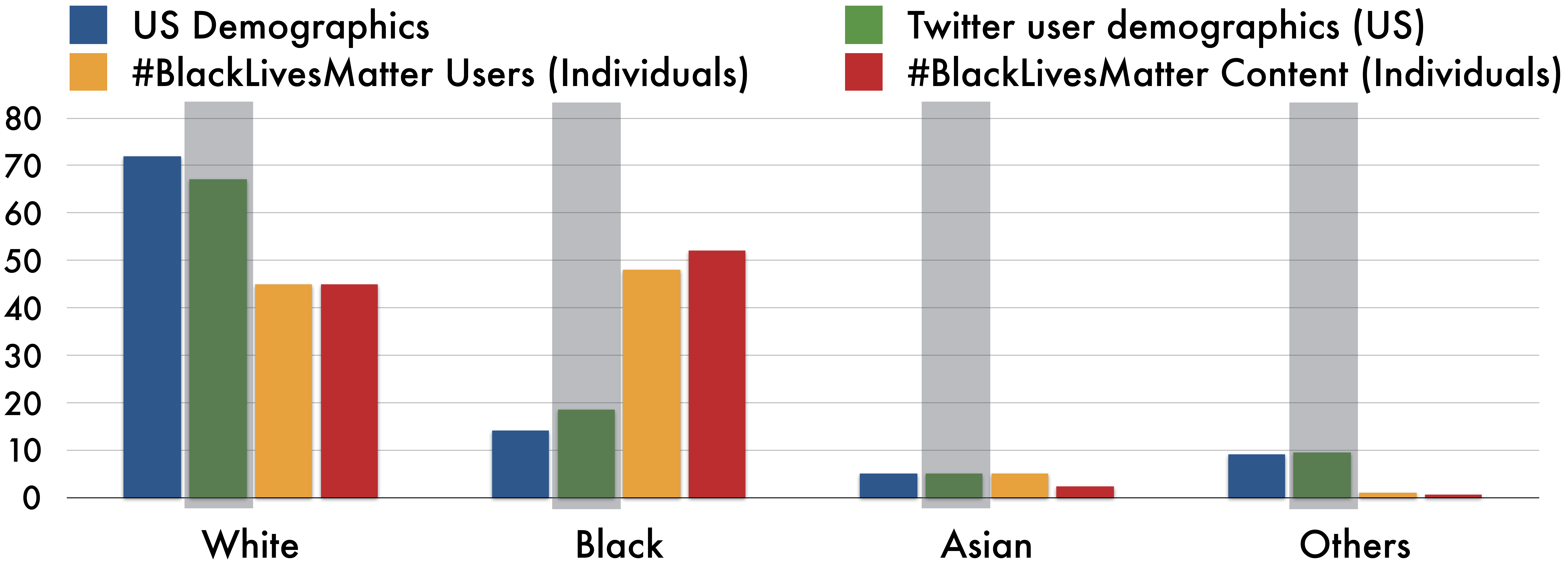
Very rough estimates based on:

- Demographics of social media, Pew Research 2015
- Demographics of the United States in 2010, US census

#BlackLivesMatter: accounts of organizations represent ~5% of all accounts and have a 3 times higher tweeting rate than individuals.

Observing the #BlackLivesMatter Movement

There is a growing number of discussions on minority issues.



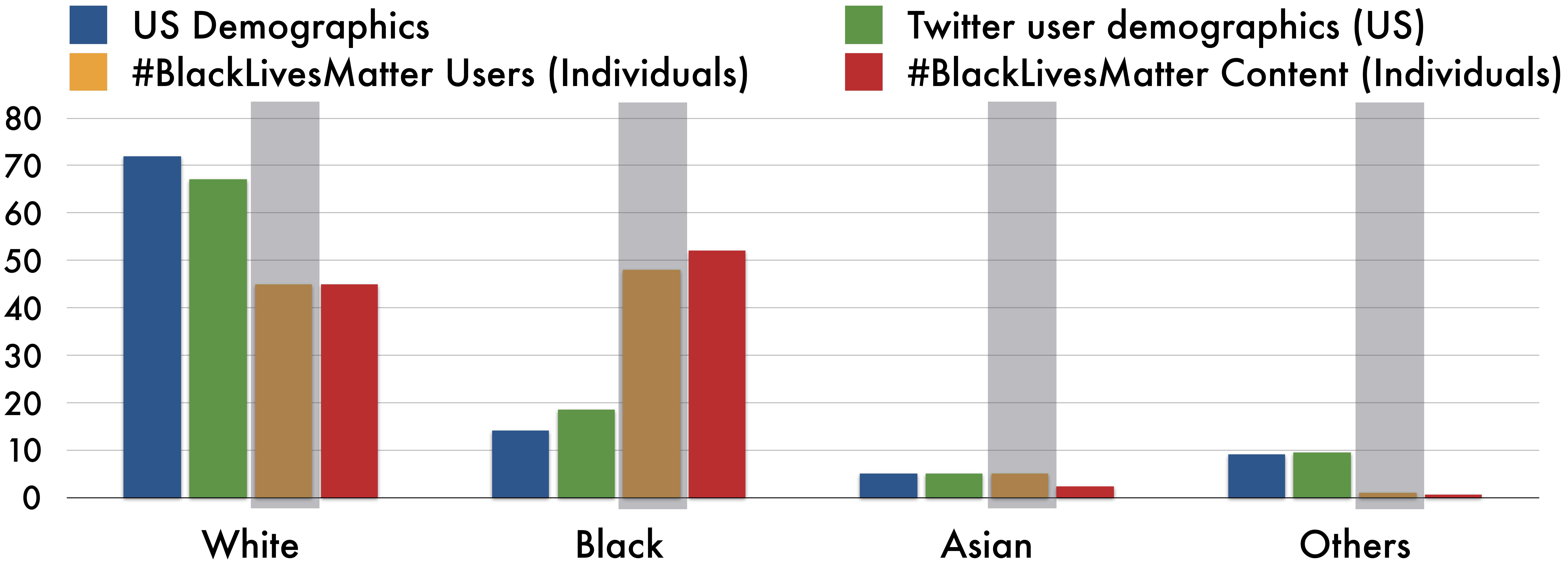
Very rough estimates based on:

- Demographics of social media, Pew Research 2015
- Demographics of the United States in 2010, US census

#BlackLivesMatter: accounts of organizations represent ~5% of all accounts and have a 3 times higher tweeting rate than individuals.

Observing the #BlackLivesMatter Movement

There is a growing number of discussions on minority issues.



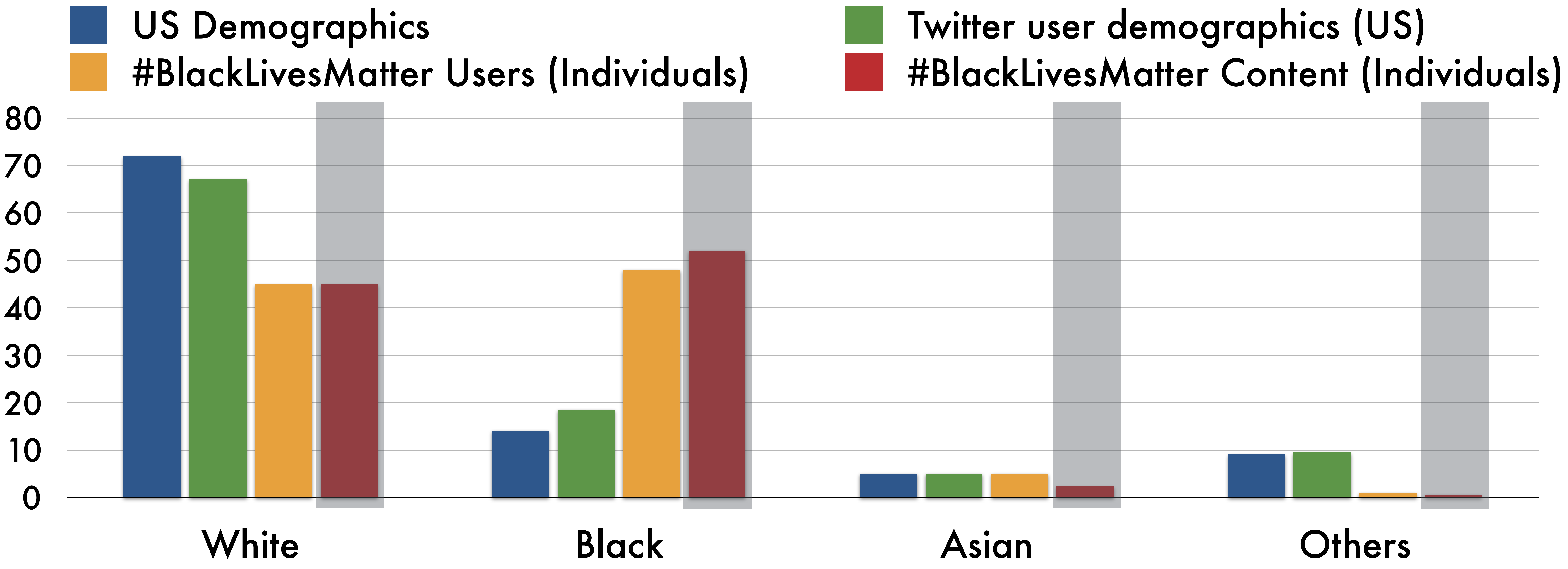
Very rough estimates based on:

- Demographics of social media, Pew Research 2015
- Demographics of the United States in 2010, US census

#BlackLivesMatter: accounts of organizations represent ~5% of all accounts and have a 3 times higher tweeting rate than individuals.

Observing the #BlackLivesMatter Movement

There is a growing number of discussions on minority issues.

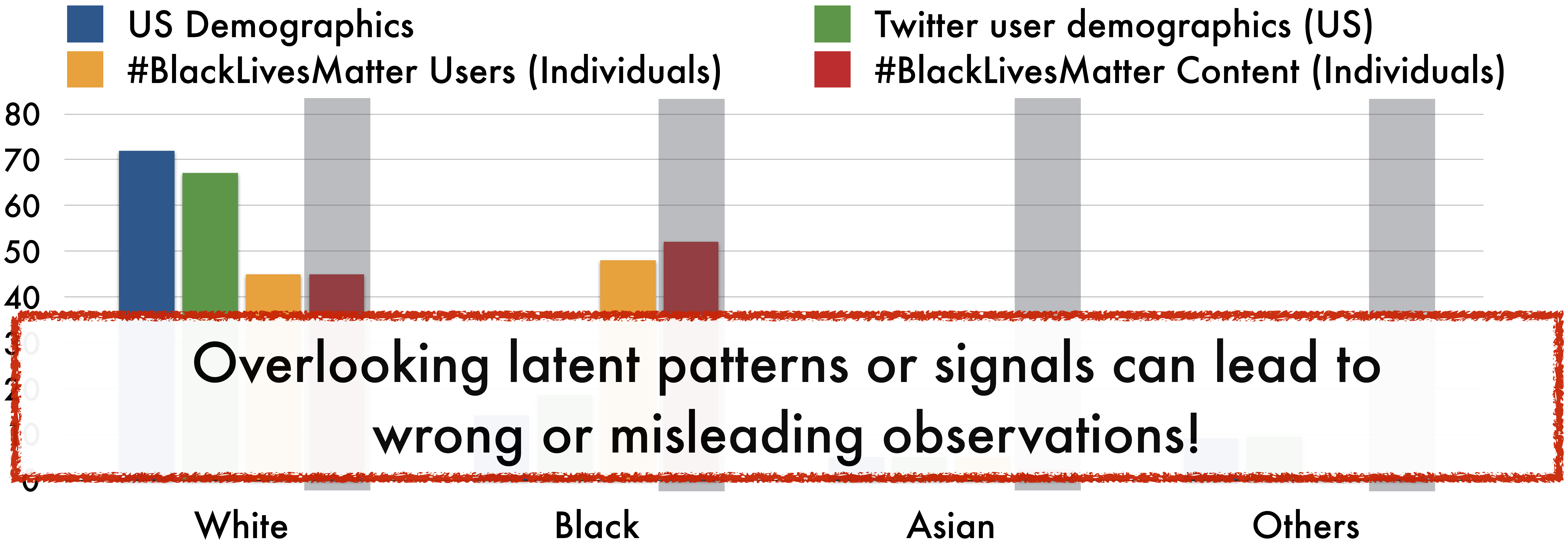


Very rough estimates based on:
- Demographics of social media, Pew Research 2015
- Demographics of the United States in 2010, US census

#BlackLivesMatter: accounts of organizations represent ~5% of all accounts and have a 3 times higher tweeting rate than individuals.

Observing the #BlackLivesMatter Movement

There is a growing number of discussions on minority issues.



Very rough estimates based on:

- Demographics of social media, Pew Research 2015
- Demographics of the United States in 2010, US census

#BlackLivesMatter: accounts of organizations represent ~5% of all accounts and have a 3 times higher tweeting rate than individuals.

Why It Matters?

Inform the public through citations, e.g., by mainstream media or in Amici Curiae Briefs

We need to understand the validity of our assumptions about the working datasets

Health/Distilling outcomes from self-reports

What can we learn about the outcomes of people experiences from social media?

Can we extract causal relations among personal event from social media data?

Health/Distilling outcomes from self-reports

What can we learn about the outcomes of people experiences from social media?

Can we extract causal relations among personal event from social media data?

with **Emre Kiciman** and **Onur Varol** [ICWSM'16, CSCW'17]

Goal: Build an **open** and **domain agnostic** system for querying about the outcomes of **any** experience people may have.

Goal: Build an **open** and **domain agnostic** system for querying about the outcomes of **any** experience people may have.

“Long-tail” of situations and experiences

- **Explore a situation: What happens …?**
  - when depressed, after disease diagnosis, after being fired, …
- **Understand the effects of a potential action: Should I …?**
  - get pregnant, ask for divorce, lose belly fat, change last name, …
- **Plan for outcomes/goals: How to …?**
  - lose weight, get admitted to MIT, increase income, find true love

Applications for individuals, policy makers, and others

Social Media Posts: A Proxy to User Experiences

Social Media Posts: A Proxy to User Experiences

Experiences & situations

I ate lots of fried things **today** and thoroughly enjoyed it. 🍔🍟🍗

I'm glad **I went** to the **show**. It was an experience **I** had to have, and **I'm** thankful.

I had my first **car accident** this morning

Social Media Posts: A Proxy to User Experiences

Experiences & situations

Post-hoc events (potential outcomes)

I ate lots of fried things **today** and thoroughly enjoyed it. 🍔🍟🍗

Everyone got problems losing weight and **I** got problems **gaining weight** 😒

I'm glad **I went** to the **show**. It was an experience **I** had to have, and **I**'m thankful.

i was just woken up to a strawberry milkshake and a **relaxed** household. this is nice.

I had my first **car accident** this morning

Not having a car this week and maybe next week will be the longest, most hardest thing ever 😒 What a great Valentines Day 😟

Social Media Timelines: Experiencing Depression

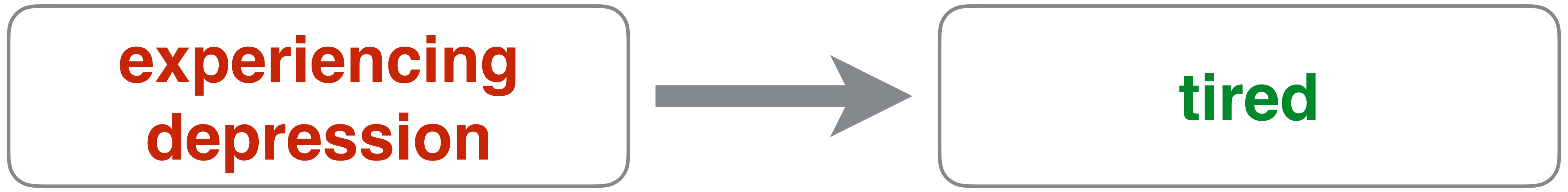
Social Media Timelines: Experiencing Depression

Social Media Timelines: Experiencing Depression

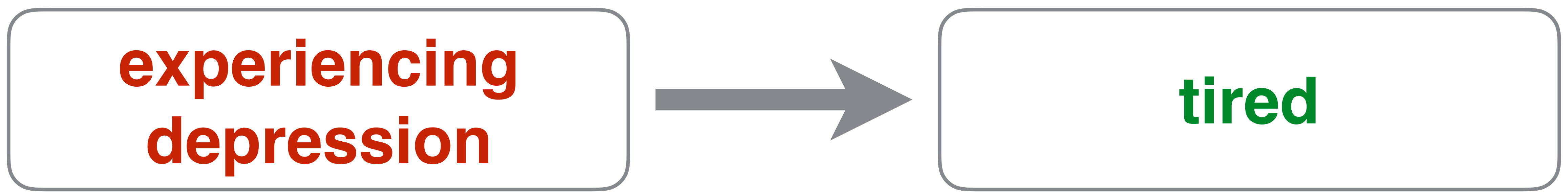
Social Media Timelines: Experiencing Depression

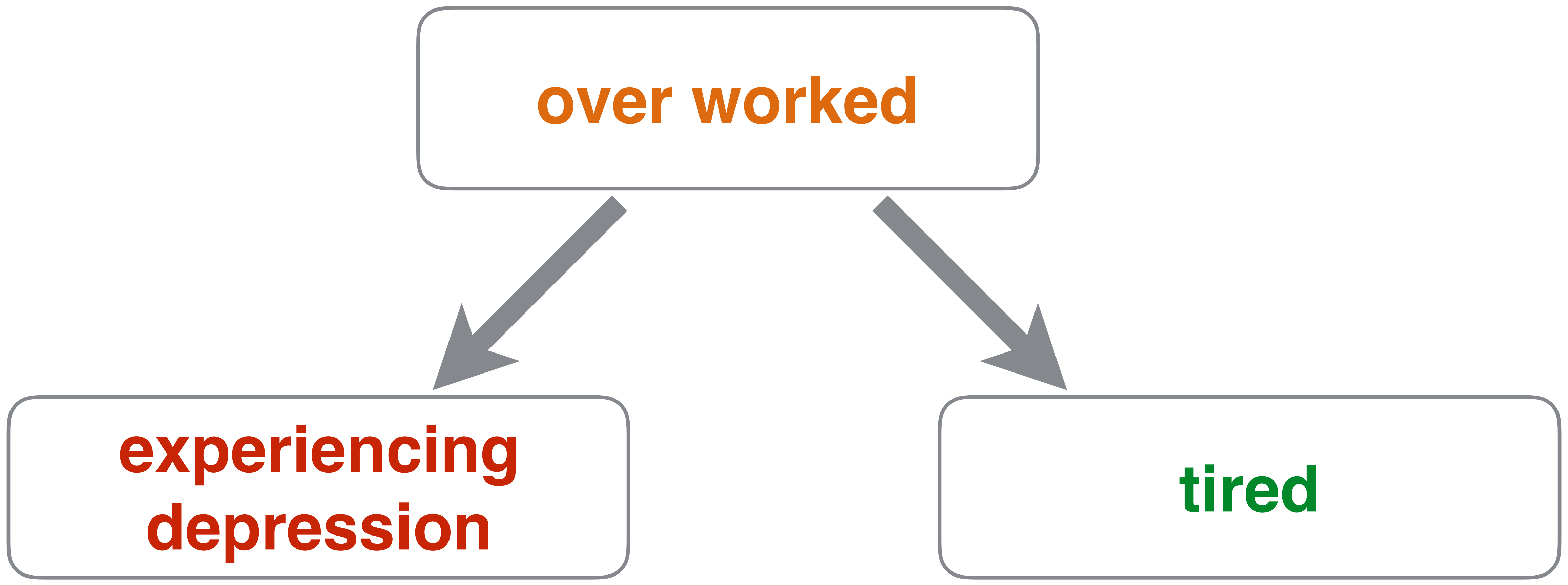
Social Media Timelines: Experiencing Depression

Confounding Bias & Matching

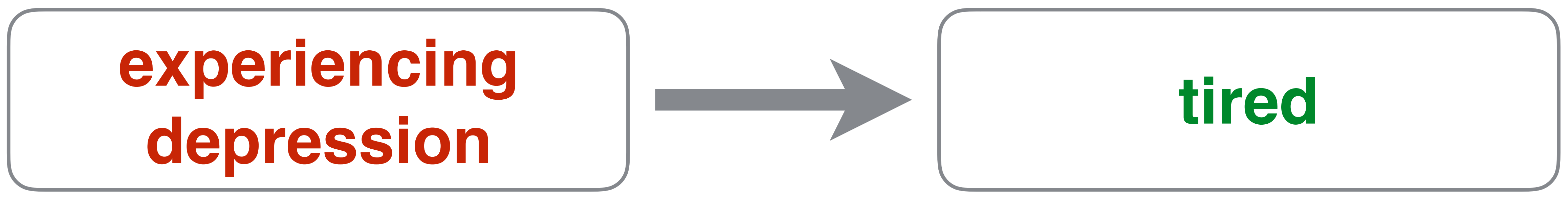


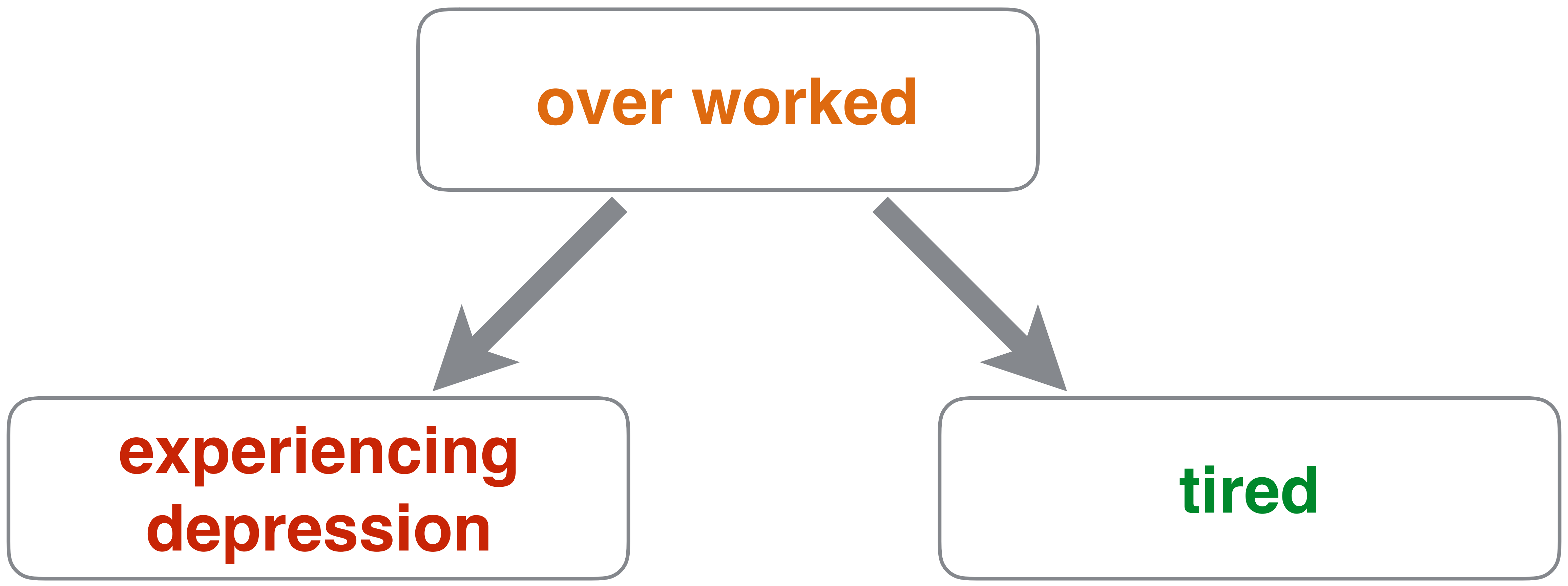
Confounding Bias & Matching





Confounding Bias & Matching





Matching: For every user with a particular experience, find another user with identical characteristics (prior to the experience) who didn't have the experience.

What Kind of Outcomes We Distill?

Causal-like relations discovered at higher rates (from **ConceptNet5**)

- **implementation steps** e.g., HasSubEvent, HasFirstSubEvent
- **motivations and prerequisites** e.g., MotivatedByGoal, HasPrerequisite
- **implications** e.g., Desires, NotDesires, CapableOf, UsedFor, Causes

We miss more conceptual or descriptive relationships

- **definitions, alternate names or similar actions** e.g., DefinedAs, RelatedTo, IsA, SimilarTo

	HasFirstSubevent	HasPrerequisite	MotivatedByGoal	HasLastSubevent	Desires	CapableOf	HasSubevent	UsedFor	NotDesires	Causes	ReceivesAction	CausesDesire	HasProperty	HasA	DefinedAs	NotCapableOf	RelatedTo	IsA	SimilarTo	DerivedFrom
Prozac	66%	60%	56%	51%	54%	52%	50%	44%	51%	46%	42%	41%	40%	38%	31%	38%	30%	25%	9%	13%
Xanax	68%	57%	57%	50%	54%	51%	52%	42%	49%	47%	43%	45%	41%	40%	39%	38%	26%	26%	17%	12%
Lorazepam	65%	59%	59%	67%	56%	53%	52%	54%	52%	45%	44%	53%	41%	42%	42%	38%	28%	27%	20%	14%
Promethazine	60%	65%	61%	68%	56%	54%	54%	56%	52%	50%	49%	41%	45%	43%	38%	39%	42%	32%	20%	17%
Tramadol	68%	66%	64%	61%	56%	55%	56%	60%	52%	55%	46%	33%	44%	44%	44%	38%	43%	32%	17%	22%

Hate speech/Impact of external events & user aspects

How do extremist events impact the prevalence of hate speech online?

How do external factors impact online phenomena?

Do user aspects impact how they perceive online hate speech?

How do we evaluate systems that work with "subjective" concepts?

Hate speech/Impact of external events & user aspects

How do extremist events impact the prevalence of hate speech online?

How do external factors impact online phenomena?

Do user aspects impact how they perceive online hate speech?

How do we evaluate systems that work with "subjective" concepts?

with **Carlos Castillo**, **Jeremy Boy**, and **Kush Varshney** [ICWSM'18]

Operationalizing Hate Speech

#agendaofevil, #attackamosque, #banislam, #bansharia, #cantcoexistwithislam, #deathcult, #deleteislam, #deportallmuslims, **#extremistsarenotmuslim**, #fuckallah, #fuckislam, **#illridewithyou**, #islamicinvasion, #islamistheproblem, #killallmuslims, #marchagainstsharia, #norapeugees, **#notinmyname**, #religionofhate, **#takeonhate**, #stopimportingislam, **#weareallmuslim**, #stopmoslemsinvasion, #islamisevil, **#terrorismhasnoreligion**

Speech that could be perceived as offensive, derogatory, or in any way harmful, and that is motivated, in whole or in a part, by someone's bias against an aspect of a group of people, or related to commentary about such speech by others, or related to speech that aims to counter any type of speech that this definition covers.

Study Setup

Two different social platforms: **Twitter** (107 M) and **Reddit** (45 M)

Various types of hate speech, based on **stance, intensity, target, frame**

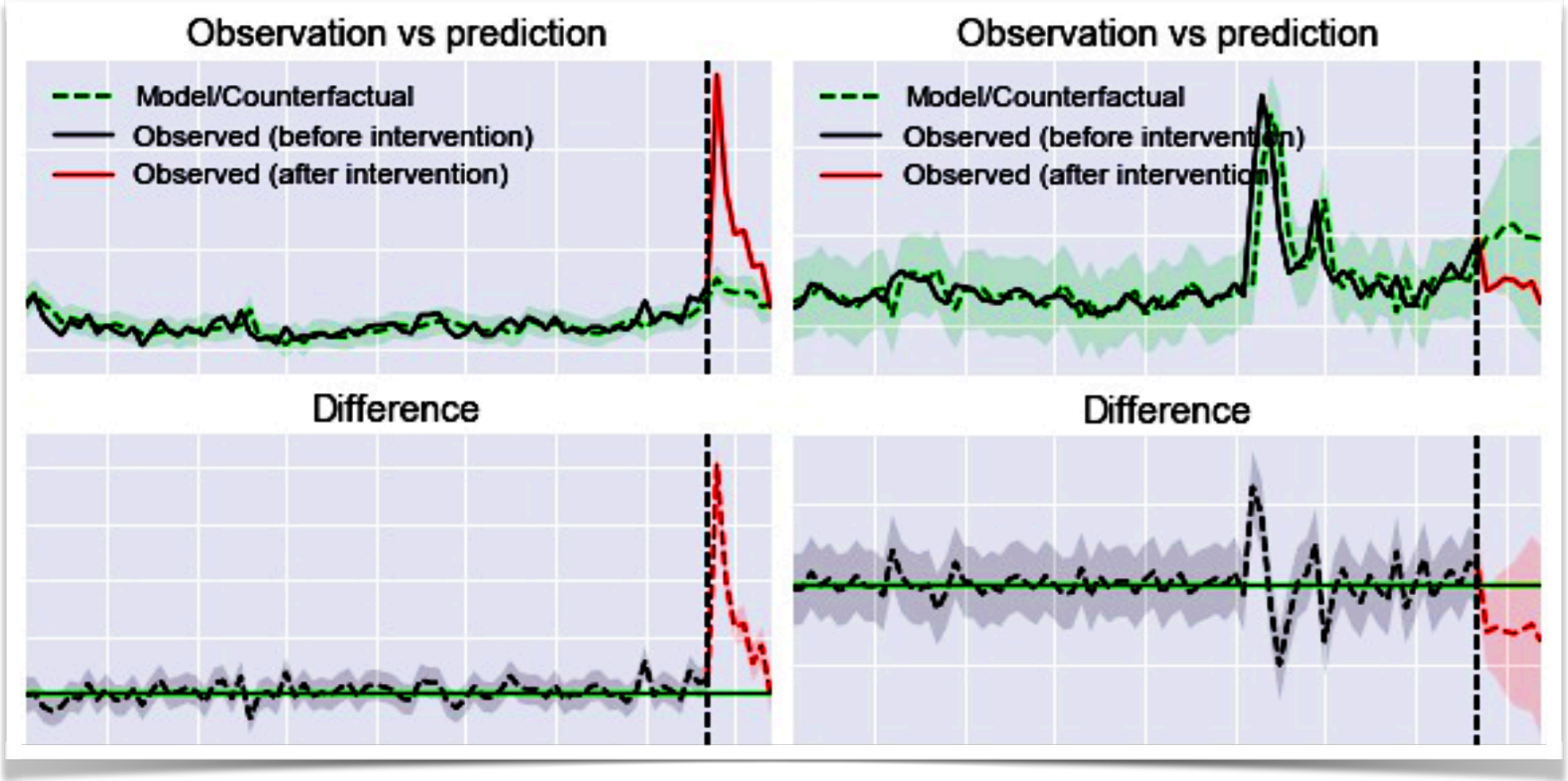
Extremist attacks involving Arabs and Muslims as **perpetrators** or **victims**

The Events Impact on Hateful Speech

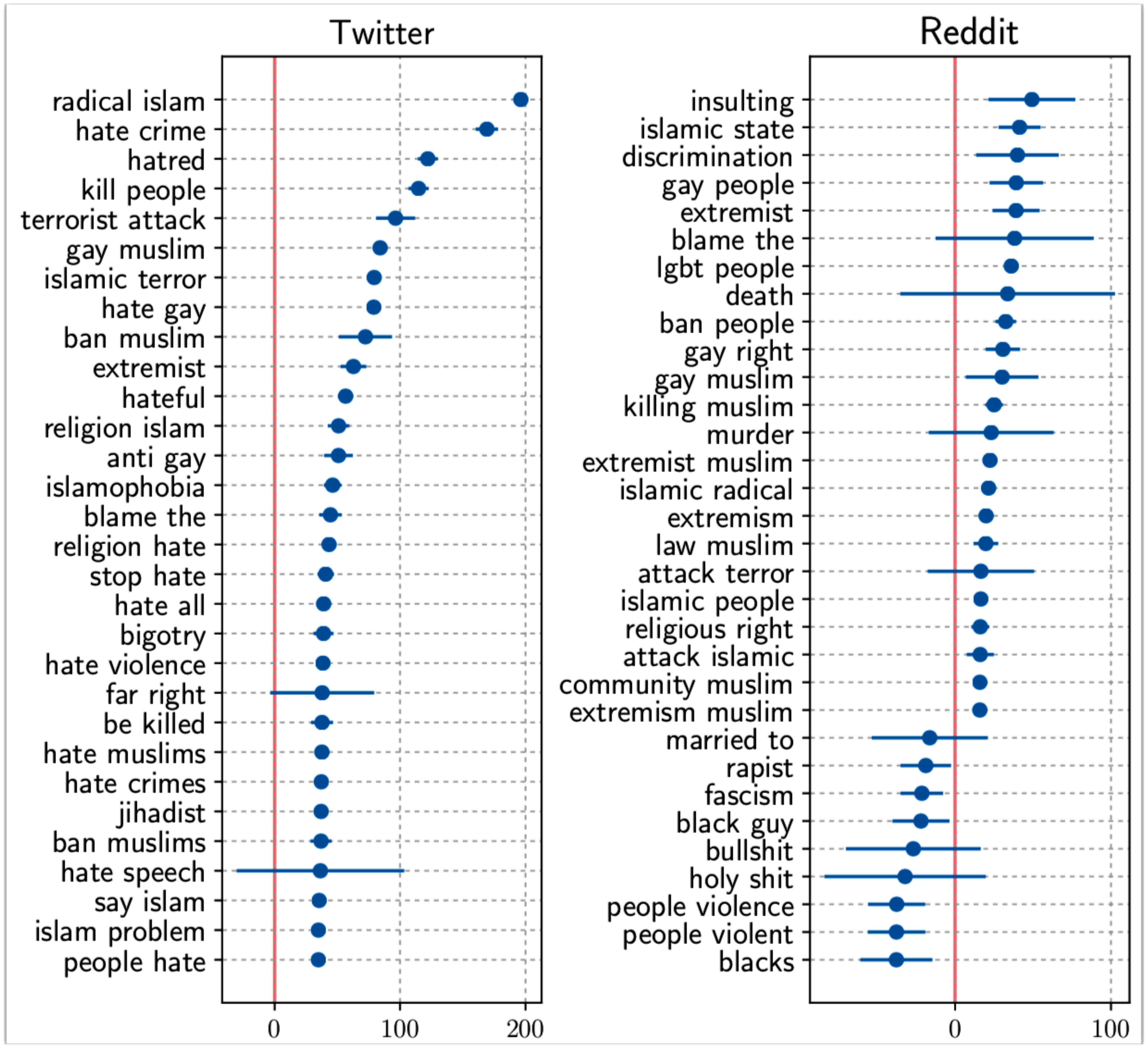
1. **Predict the counterfactual**: what would had happen had no event taken place
2. **Estimate the effect**: the difference among observed series and the predicted ones
3. **Aggregate results**: distribution of effects across platforms and types of speech

Query: evil muslim

Events:
(Right) Olathe Kansas shooting
(**Left**) Orlando nightclub shooting



Estimated Relative Effects



Orlando nightclub shooting

We observed:

- an increase in hate speech targeting Muslims after Islamic terrorist attacks
- an increase in counter speech terms after Islamic terrorist attacks
- an increase in counter speech terms related to religion
- ...

Hate speech/Impact of external events & user aspects

How do external events impact the prevalence of hateful chatter online?

How do external factors impact online phenomena?

Do user traits impact how they perceive online hateful chatter?

How do we evaluate systems that deal with "subjective" concepts?

with **Kartik Talamadupula** and **Kush Varshney** [WebSci'17]

Hate Speech Classification: Experimental Setup

Hypothetical classification task:

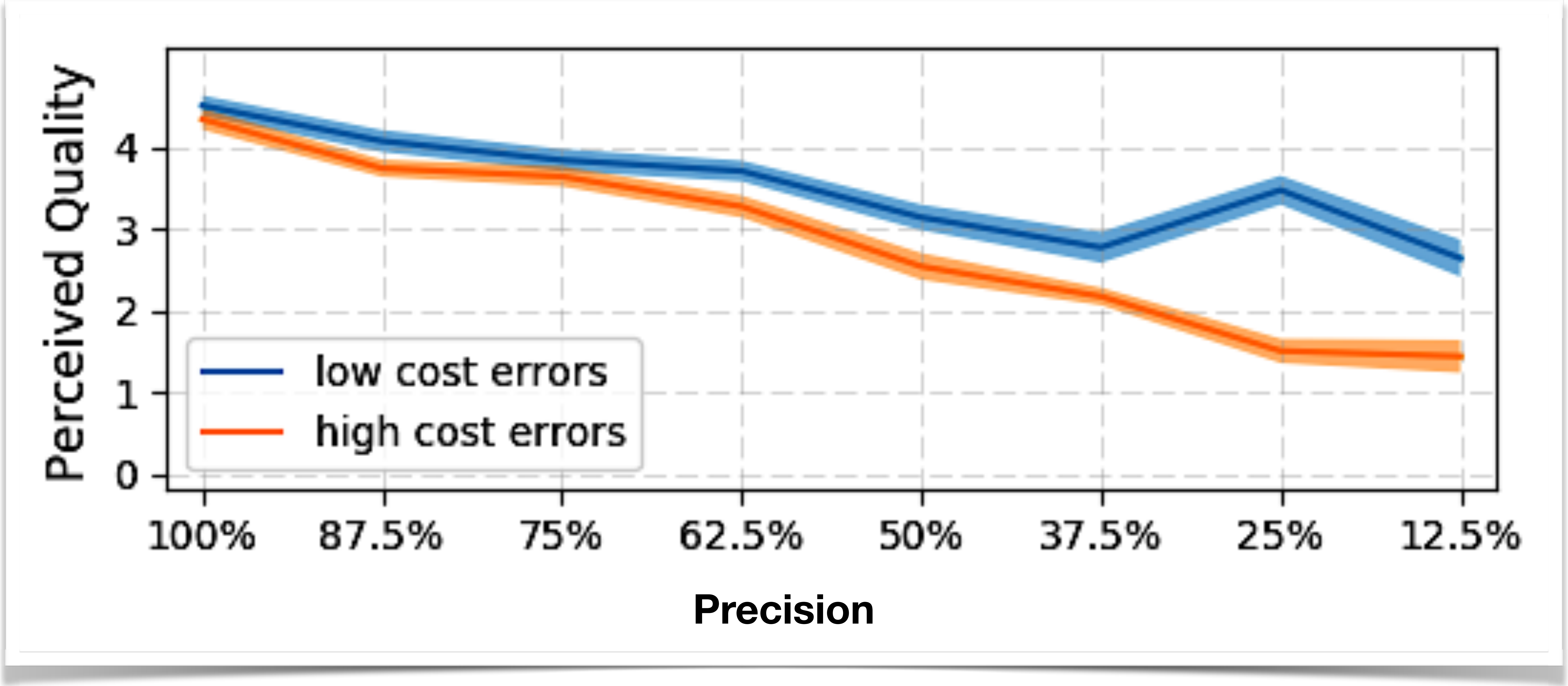
Detect and output social media posts classified as hate speech

Annotated tweets	Hate speech	Offensive but not hate speech	Not offensive
14509	2399	4836	7274
	True positives	False Positives	False positives
		Low cost errors	High cost errors

- **Low cost:** misclassifying **other types of offensive posts**
- **High cost:** misclassifying **non-offensive posts**

Fix precision, vary error types by cost.

Hate Speech: Error Types



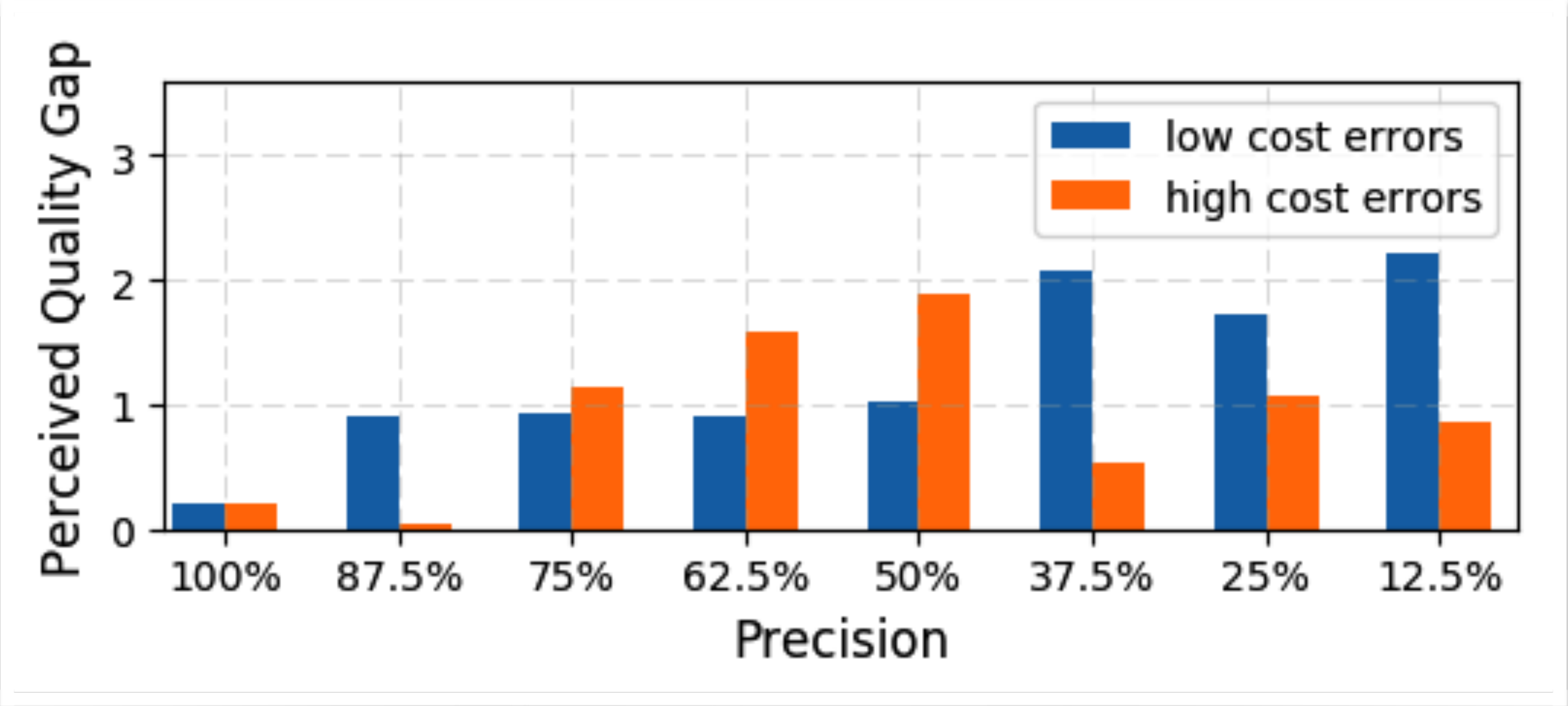
Hate speech detection:

- **Low cost:** misclassify **other types of offensive posts** as **hate speech**
- **High cost:** misclassify **non-offensive posts** as **hate speech**

User Traits: Stance & Experience

Total annotations: 8 (precision points) **X 2** (error types) **X 5** (annotations) **X 6** (samples)
Were you ever the direct target of hate speech?
Yes, unfortunately
No, but I've experienced other forms of online harassment
No, never
How important it is to moderate hate speech content on social media?
I think this is increasingly necessary
Some form of moderation is needed, but I also worry about free speech rights
It is not necessary. If you don't like what folks say, do not engage with them

User Traits & Their Perception of Performance



Those targeted by hate speech online appear to apply a broader definition of what constitutes hate speech.

Why It Matters?

Understand behavioral phenomena of societal importance, third-party interventions and other policy questions

We need new ways to evaluate computational systems

We need to develop and apply techniques that reduce the effects of data biases

Research Paths & Impact

Types of Contributions

A more efficient solution to a known problem

An interesting solution to a known problem

Introduces a new problem (e.g., the solution matters less)

Makes the community contributions more accessible

Clarifies the trade-offs and gaps of existing approaches

- includes negative results, meta-analyses, etc.

Some Rationales

Impact more important than being faithful to a topic

- I decided the topic of my thesis with 3 months before defense

Do what others don't want to do

Optimize for relevance within the application domain

Focus on what can be done, not on what cannot be done

Illuminating a problem as important as solving a problem

Some Rationales

Impact more important than being faithful to a topic

- ~~I decided the topic of my thesis with 3 months before defense~~

Do what others don't want to do

Optimize for relevance within the application domain

Focus on what can be done, not on what cannot be done

Illuminating a problem as important as solving a problem

Impact Can Have Many Flavors

Some Types of Impact

Scientific impact/make a breakthrough

Popularize a methodological approach

Enable others to do more/better work

- release data, tools, surveys, etc.

Great teaching material

Policy impact/news coverage

Collaboration Is Important

Collaborators

Advisor(s)

Outside mentors

- from internships
- shared interests
- visiting professors

Lab mates

Students

Collaborators

Advisor(s)

Outside mentors

- from internships
- shared interests
- visiting professors

Lab mates

Students

Look for diversity in backgrounds, and respect their perspective.

Finding Mentors

Remember

- they are busy
- they have their own priorities/interests

Align your goals with their interests

Do not overestimate their benefits

They can help with more than research

- networking, support, endorsement, etc.

Finding Mentors

Remember

- they are busy
- they have their own priorities/interests

Align your goals with their interests

Do not overestimate their benefit

They can help with more

- networking, support … , etc.

But, keep them accountable for the things they commit to.

Be a Mentor

Take their interests into account

Advisor shoes

Put them first

- review their work ⇒ will help you write better.
- do dry-runs with them ⇒ will help you make better presentations.

You learn better when you explain it to others

Be a Mentor

Take their interests into account

Advisor shoes

Put them first

- review their work ⇒ will help you write better.
- do dry-runs with them ⇒ will help you
  presentations.

You learn better when yo ... ers

Offer help, do not expect them to always know what they need.

Receiving Feedback: A Few "Do not(s)"

Don't be critical of how people give you (solicited) feedback

- "You do not know how to give feedback"

Don't get defensive and argumentative

- "You are wrong"
- "You did not understand it"

Don't try to figure it out on the spot, take time

Receiving Feedback: A Few "Do not(s)"

Don't be critical of how people give you (solicited) feedback

- "You do not know how to give feedback"

Don't get defensive and argumentative

- "You are wrong"
- "You did not understand it"

Don't try to figure it out on the spot, take time

Be grateful!

Give Credit, Always

Err on the side of giving others more credit, than less credit

Ignoring it, likely the easiest way to make people resentful

It's easy to think your contributions are more valuable than of others

KEEP
CALM
AND
QUESTION
EVERYTHING

Email:
alexandra@aolteanu.com

Web:
aolteanu.com

Twitter:
@o_saja