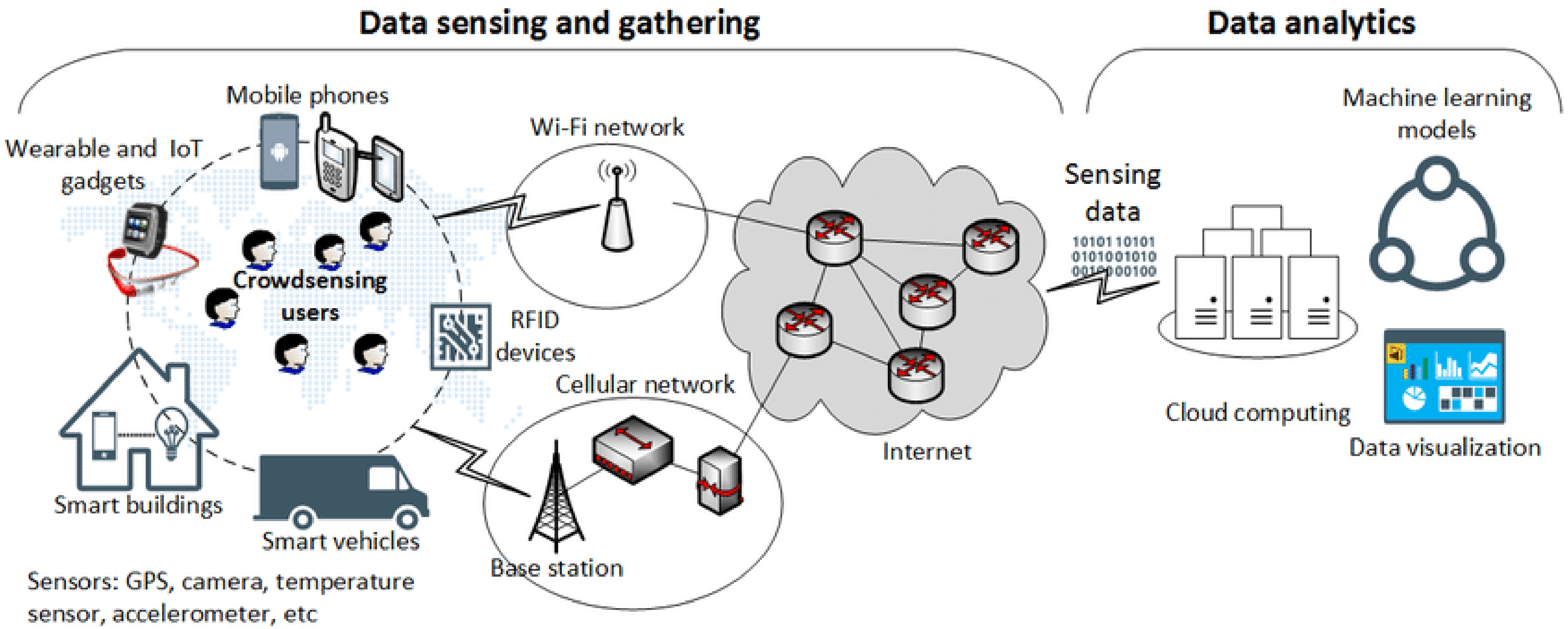
prof. Marco Avvenuti
Dept. of Information Engineering
University of Pisa
marco.avvenuti@unipi.it

USING HUMANS AS SENSORS

(SOCIAL NETWORKS AS SENSOR NETWORKS ?)

[Mobile] Crowdsensing



[Mobile] Crowdsensing

A large group of individuals having mobile devices capable of sensing and computing, collectively share data and extract information to measure, map, analyze, estimate or infer any processes of common interest

Based on the type of involvement from the users, mobile crowdsensing can be classified into two types:

<u>Participatory</u>: users voluntarily participate in contributing information	<u>Opportunistic</u>: data is sensed, collected and shared automatically *without user intervention* and, in some cases, even without the user's awareness

(Twitter) in the aftermath of emergencies

JAPAN TSUNAMI, IN MARCH 2011

HURRICANE SANDY, IN OCTOBER 2012

BOSTON MARATHON BOMBING, IN APRIL 2013

...

MAJOR EARTHQUAKES AROUND THE WORLD

OTHER NATURAL DISASTERS SUCH AS HURRICANES AND FLOODING

MAN-MADE DISASTERS SUCH AS RIOTS, CIVIL UNREST

(Twitter) in the aftermath of emergencies

EMERGENCY RESPONSE

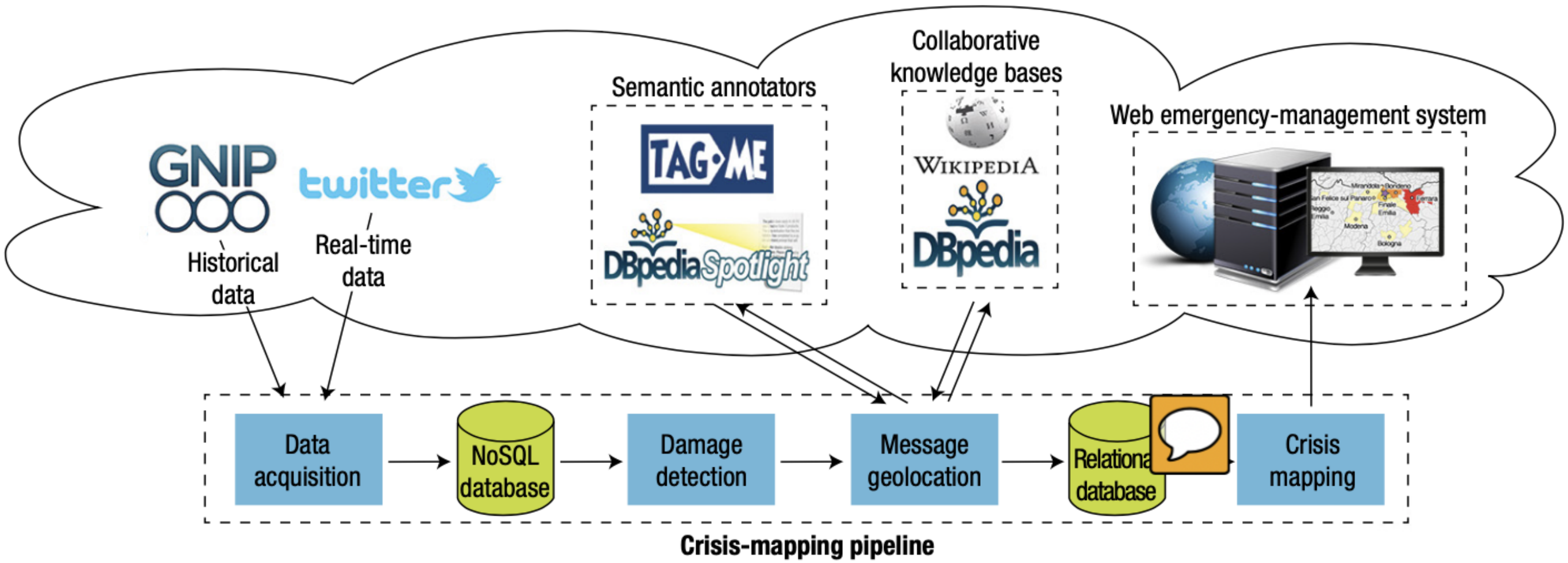
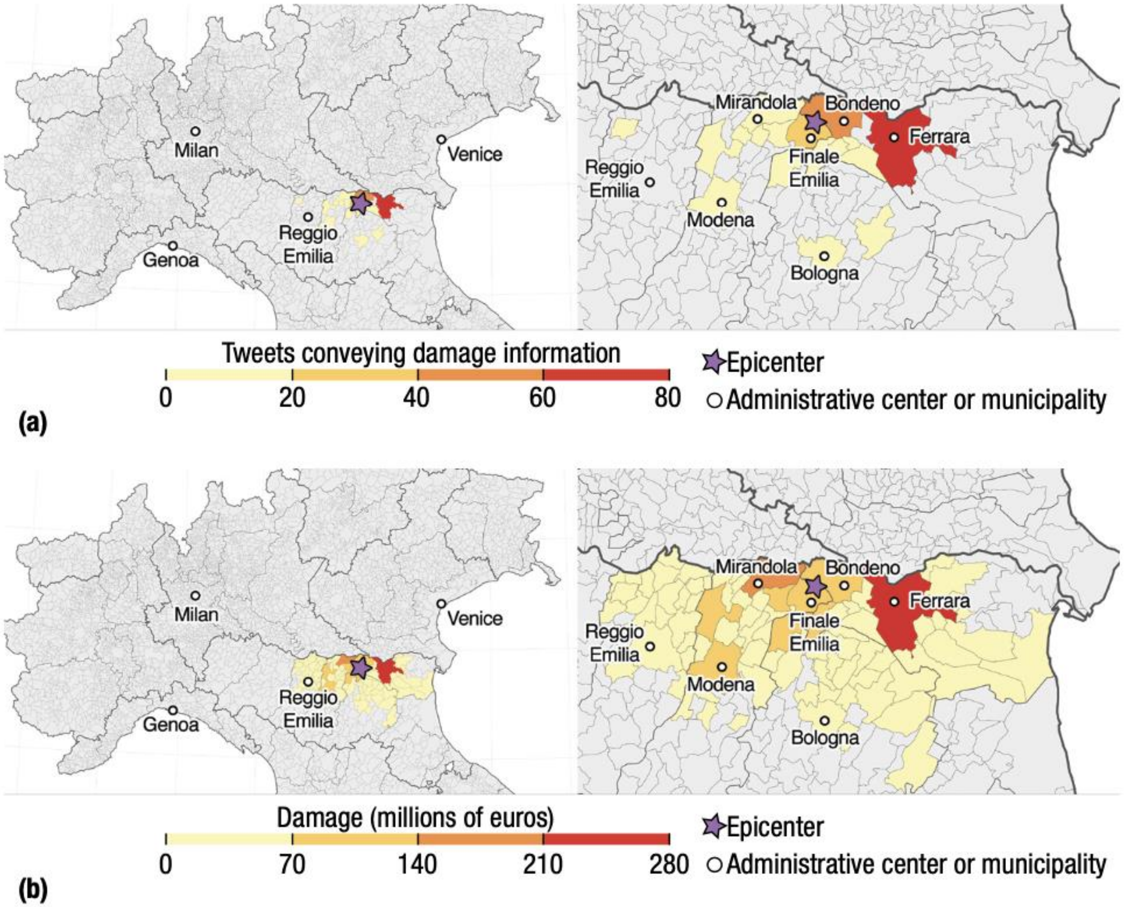


FIGURE 1. Architectural overview of our crisis-mapping system, which has four main components. Once the system acquires tweets, it stores them in a NoSQL database and then parses them for mentions of damage and location information, and finally stores them in a relational database. Using the combination of damage and location information, it then creates a crisis map, coloring areas with the most damage.

M. Avvenuti et al., “Impromptu Crisis Mapping to Prioritize Emergency Response,” *Computer*, vol. 49, no. 5, 2016.

(Twitter) in the aftermath of emergencies



After <u>30 minutes</u> of the heartquake

After <u>30 days</u> of the heartquake

M. Avvenuti et al., "Impromptu Crisis Mapping to Prioritize Emergency Response," *Computer*, vol. 49, no. 5, 2016.

Humans as Sensors (HaS)

- The huge availability of social network content suggests that the largest “sensor network” yet might be *human*
- An extension of participatory sensing: *utilizing social networks as sensor networks*
- Human sources represent sensors
- The occasional observations they make about the physical world represent (data) *claims*

HaS: a Binary Model of Human Sensing

Let consider only participatory sensing of the physical environment that is *external to the human sensor* (e.g., «Shooting erupts on Liberty Square!»)

Subjective reporting such as «It is an inspiring day!» and personal emotions «I am depressed» are not considered

In this case, the physical environment has a unique state, leading to a unique ground truth, according to which human observations are either true or false (e.g., either there was shooting on Liberty Square or not)

HaS: a Reliable Sensing Problem

Observations of the physical world may be true or false, and hence are viewed as *binary claims*

The sensing problem is to determine which claims are correct (i.e., separate data from noise), given that:

the *reliability* of human sensors is generally *unknown a priori*	the *provenance* of reported observations is *uncertain* (i.e., individuals may report observations made by others as their own)

HaS: the Model of Human Sensing

Let model human participants as *sources of unknown reliability* generating *binary measurements of uncertain provenance*

Based on the above model, let formulate a rigorous estimation-theoretic problem to *optimize filtering of correct observations* in a *maximum likelihood* sense

Let demonstrate that this model enables the reconstruction of ground truth from noisy human observations in practice

1: Unknown Reliability

Unlike physical sensors whose characteristics, calibration, and failure modes are known, we do not, in general, know the reliability of human observers and hence cannot assume it in our problem formulation

Also, humans are much broader in what they can observe, albeit less accurate

Different individuals have different reliability, expressed as the probability of producing true claims

2: Binary Observations

- The physical world is a collection of mention-worthy facts (e.g., «Main Street is flooded», «The BP gas station on University Ave is out of gas», «Police are shooting people on Market Square»)
- Human sensors report some of the facts they observe
- Because the observation reported can either be true or false, claims can be thought of as measurements of different *binary variables*
- Reliable sensing is to infer which of the reported human observations match ground truth in the physical world

Crash blocking lanes on I-5S @ McBean Pkwy in Santa Clarita
BREAKING NEWS: Shots fired in Watertown; source says Boston Marathon terror bomb suspect has been pinned down
The police chief of Afghanistan's southern Kandahar province has died in a suicide attack on his headquarters.
Yonkers mayor has lifted his gas rationing order. Fill it up!

3: Uncertain Provenance

When Bob tweets that «Main Street is flooded», even if we authenticate Bob as the actual source of the tweet, we do not know if Bob truly observed that first-hand or heard it from Sally

It is not unusual for a person to report observations they received from others as if they were his/her own

Such rumor spreading behavior has no analogy in correctly functioning physical sensors

From a sensing perspective, this means that errors in «measurements» across «sensors» may be non-independent, as one erroneous observation may be propagated by other sources without being verified

HaS: a Solution Architecture

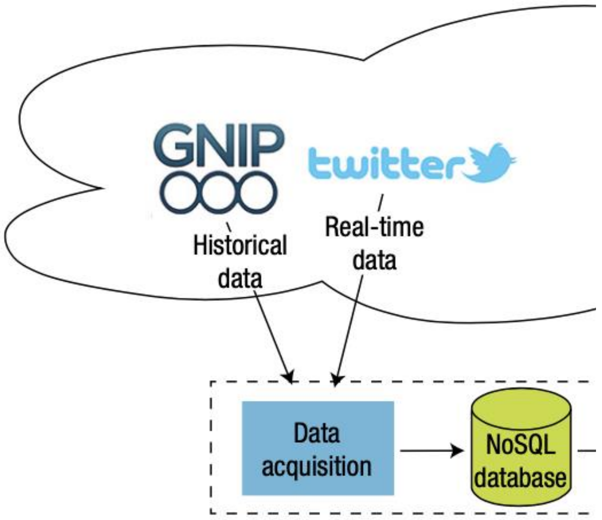
To enable reconstruction of ground truth information from data reported by human sources, we need to:

1. collect data from the «sensor network»
2. structure the data for analysis
3. understand how sources are related
4. use this collective information to estimate the probability of correctness of individual observations

1: Data Collection

In principle, we can collect data from any participatory sensing front end/social media platform. Here, we suppose to collect from (ex)Twitter

Tweets were collected through a long-standing query via the Streaming Twitter API to match given keywords and, optionaly, an indicated geographic region on a map. These can either be *and*ed or *or*ed



2.1: Data Structuring

- A *distance function* `distance(t1,t2)` is computed for every pair of reported observations
- This function takes two reported observations as input and returns a measure of similarity between them. The more dissimilar the observations, the larger the distance
- In the case of data collection from Twitter:
  - individual tweets = individual observations
  - *A*, *B* vectors of token occurrence in tweets
  - let use a *cosine similarity function* to return a measure of similarity based on the number of matching tokens in the two inputs

$$\text{similarity} = \cos\theta = \frac{A \cdot B}{\|A\|\|B\|}$$

dot product

vector magnitude

```
doc_1 = "Data is the oil of the digital economy"
doc_2 = "Data is a new oil"

# Vector representation of the document
doc_1_vector = [1, 1, 1, 1, 0, 1, 1, 2]
doc_2_vector = [1, 0, 0, 1, 1, 0, 1, 0]
```

	data	digital	economy	is	new	of	oil	the
doc_1	1	1	1	1	0	1	1	2
doc_2	1	0	0	1	1	0	1	0

$$A \cdot B = \sum_{i=1}^{n} A_i B_i$$

$$= (1 * 1) + (1 * 0) + (1 * 0) + (1 * 1) + (0 * 1) + (1 * 0) + (1 * 1) + (2 * 0)$$

$$= 3$$

$$\sqrt{\sum_{i=1}^{n} A_i^2} = \sqrt{1 + 1 + 1 + 1 + 0 + 1 + 1 + 4} = \sqrt{10}$$

$$\sqrt{\sum_{i=1}^{n} B_i^2} = \sqrt{1 + 0 + 0 + 1 + 1 + 0 + 1 + 0} = \sqrt{4}$$

$$cosine\ similarity = cos\theta = \frac{A \cdot B}{|A||B|} = \frac{3}{\sqrt{10} * \sqrt{4}} = 0.4743$$

2.2: Data Structuring

- We first build a graph where vertices are individual observations (tweets) and links represent similarity among them
- We then cluster the graph, causing similar observations to be clustered together
- We call each such cluster a **claim**, defined as a piece of information that *several* sources reported

2.3: Data Structuring

- We can now construct a **Source-Claim graph, *SC***, in which:
  - each source, *Si*, is connected to all claims they made (i.e., clusters they contributed to)
  - each claim, *Cj* , is connected to all sources who espoused it (i.e., all sources of tweets in the corresponding cluster)
- We say that *SiCj = 1* if source *Si* makes claim *Cj*
- Each claim is `true`, if it is consistent with ground truth in the physical world; otherwise, it is `false`

3.1: Sources Relationship

- *Uncertain provenance*: sources may have reported either their own observations or observations they heard from others
- We assume the existence of a latent **social information dissemination graph, *SD***, that estimates how information might propagate from one person to another:

1. **Epidemic Cascade network (*EC*)**: each distinct observation is modeled as a cascade and the time of contagion of a source describes when the source mentioned this observation

3.2: Sources Relationship

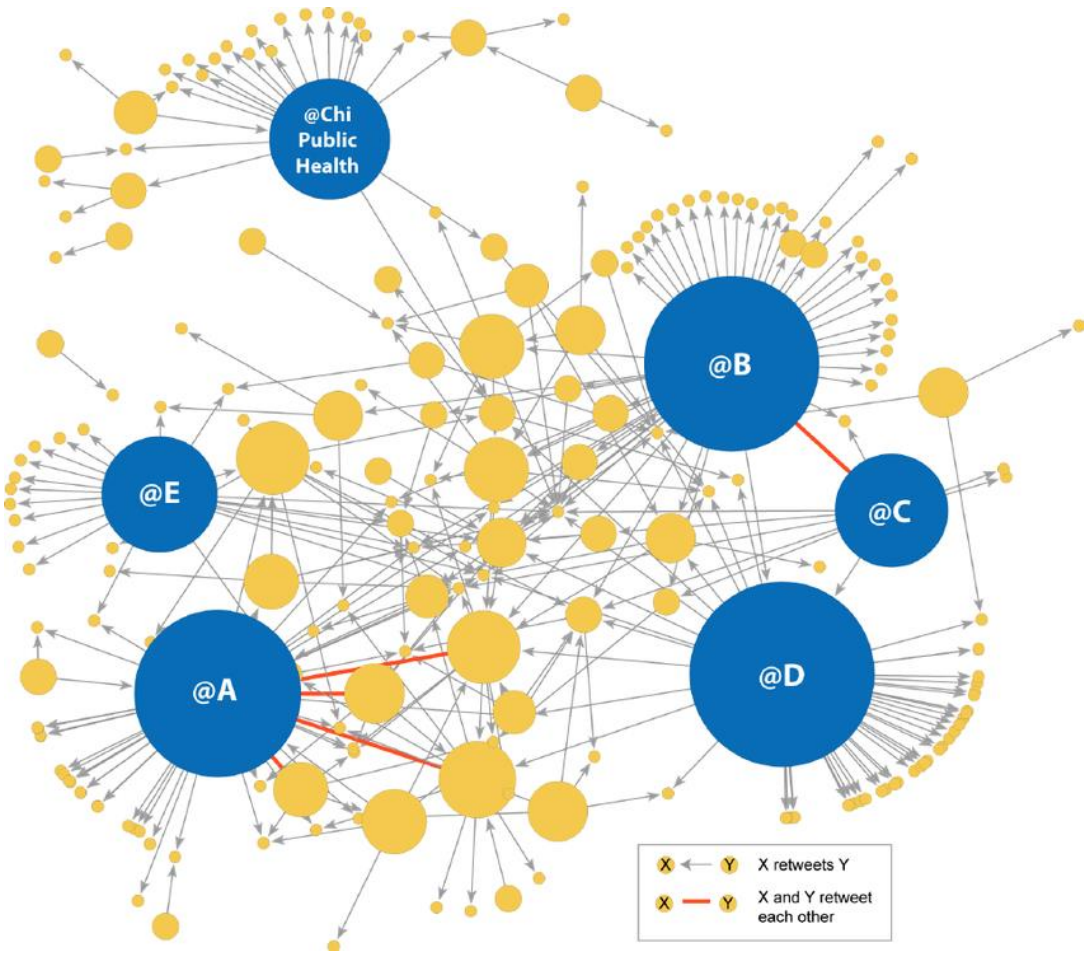
2. **Follower-Followee network (*FF*)**: a directed link (*Si*,*Sk*) exists in the social graph from source *Si* to source *Sk,* if *Sk* is a follower of *Si*
3. **Re-Tweeting network (*RT*)**: a directed link (*Si*,*Sk*) exists in the social graph if source *Sk* retweets some tweets from source *Si*
4. **Combined network (*RT+FF*)**: a directed link (*Si*,*Sk*) exists when either *Sk* follows *Si* or *Sk* retweets what *Si* said

3.2: Sources Relationship

Note: Red nodes correspond to authors of retweeted tweets. The rest are retweeters distinguished by their in-degree centrality (the lower, the lighter).

Examples of following/followers social graphs of retweeted tweets

3.2: Sources Relationship



Retweet network with nodes sized by outdegree, or number of times the Twitter user was retweeted during the week.

4.1: The Estimation Problem

With inputs computed, the next stage is to perform the analysis that estimates correctness of claims

For each claim, *Cj,* we need to determine if it is true or false

Tweets are analysed using a sliding window: Let the total number of claims computed from tweets received in the last window be N

4.2: The Estimation Problem

Voting: claims with more support are more believable

- For each C_j, where $1 \leq j \leq N$, count all S_i, where $S_iC_j = 1$
- This solution is Suboptimal:
  - **different sources have different degrees of reliability**, hence, their «votes» do not have the same weight
  - **sources may not be independent**. When a source simply repeats what they heard from others, their «vote» does not add to the credibility of the claim

4.3: The Estimation Problem

Likelihood Estimation: compute the *probability of correctness* of claims taking into account *unknown source reliability* and *uncertain provenance:*

- Given graphs *SC* and *SD* what is the likelihood that claim *Cj* is true, for each *j*?
- Formally, we compute the conditional probability that *Cj* is true given *SC* and *SD*:

$$\forall j,\ 1 \leq j \leq N : P(Cj = 1 \mid SC, SD)$$

Once this probability is computed, forward to the user those tweets that meet a specified (user configurable) probability of correctness.

5.1: Expectation maximization (unknown reliability)

- Let m be the total number of sources (i.e., «sensor») from which we have data
- Let us describe each source, $Si, 1 \leq i \leq m$, by two parameters, neither of which are known in advance:
  - the odds of true positives, $a_i = P(SiCj = 1 | Cj = 1)$
  - the odds of false positives, $b_i = P(SiCj = 1 | Cj = 0)$,
- Let d the unknown expected ratio of correct claims in the system, $d = \mathrm{P}(Cj = 1)$
- Let $\vartheta = [a_1 \ldots a_m b_1 \ldots b_m d]$ be the vector of the above unknowns

5.2: Expectation maximization (uncertain provenance)

Maximun Likelihood Estimation (MLE): an optimization approach

- finds the values of the unknown ϑ that maximize the probability of observations, *SC*, given the social network *SD:*

$$P(SC|SD,\vartheta)$$

- as $P(SC|SD,\vartheta)$ depends on which claims are true and which are false, let us introduce the vector *Z* where element $z_j = 1$ if *Cj* is true and zero otherwise
- assuming independent claims, we can rewrite the expression we want to maximize as follows:

$$P(SC|SD,\vartheta) = \sum_{z} P(SC,z|SD,\vartheta)$$

5.3: Expectation maximization

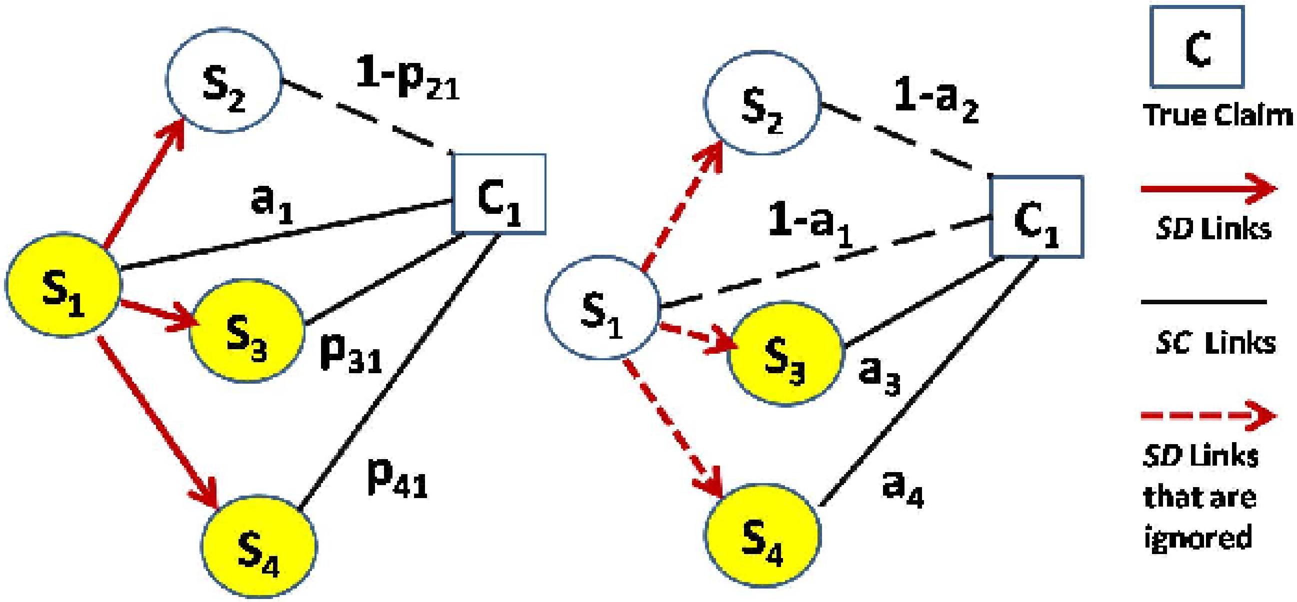
- Solution to the maximization of MLE can be found using algorithms described in literature (e.g., the iterative Expectation Maximization algorithm, page 5), *provided the notion of uncertain provenance of social networks is incorporated*
- **Intuition**: considering the SD graph,
  - if the parent does not make the claim, then the children act as if they are independent sources
  - If the parent makes the claim, then each child repeats it with a given *repeat probability*

5.4: Expectation maximization

- This can be done by merging graphs *SC* and SD and dividing them into subsets, *SCj*, each one describing which sources supported the *Cj* and which did not
- Then, computing the probability that source *Si* makes clain Cj given that its parent *Sk* makes the claim (*repeat ratio):*

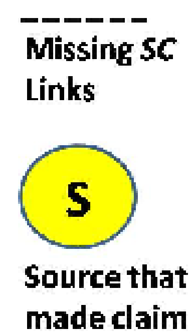
$$p_{ik} = \frac{\text{number of times Si and Sk make same claim}}{\text{number of claims Sk makes}}$$

Example



Example 1: Parent S_1 makes claim C_1. Each descendant, i, repeats it with probability p_{i1} or not to with probability $1\text{-}p_{i1}$.

Example 2: Parent S_1 does not make claim C_1. Its descendants are modeled as independent nodes. They make claim C_1 with probability a_i or chose not to with probability $1\text{-}a_i$.



Algorithm 1 Expectation Maximization Algorithm

Initialize θ with random values between 0 and 1
Estimate the dependent ratio (i.e., p_{ig}) from source dissemination graph SD based on Equation (7)
while $\theta^{(n)}$ does not converge **do**
  for $j = 1 : N$ **do**
    compute $Z(n, j)$ based on Equation (11)
  end for
  $\theta^{(n+1)} = \theta^{(n)}$
  for $i = 1 : M$ **do**
    compute $a_1^{(n+1)}, ..., a_m^{(n+1)}, b_1^{(n+1)}, ..., b_m^{(n+1)}, d^{(n+1)}$ based on Equation (12)
    update $a_1^{(n)}, ..., a_m^{(n)}, b_1^{(n)}, ..., b_m^{(n)}, d^{(n)}$ with $a_1^{(n+1)}, ..., a_m^{(n+1)}, b_1^{(n+1)}, ..., b_m^{(n+1)}, d^{(n+1)}$ in $\theta^{(n+1)}$
  end for
  $n = n + 1$
end while
Let Z_j^c = converged value of $Z(n, j)$
Let a_i^c = converged value of a_i^n; b_i^c = converged value of b_i^n; d^c = converged value of $d^{(n)}$
for $j = 1 : N$ **do**
  if $Z_j^c \geq 0.5$ **then**
    C_j^* is true
  else
    C_j^* is false
  end if
end for
Return the claim classification results.

- The input is the source claim graph *SC* from social sensing data and the source dissemination graph *SD* estimated from social network
- The output is the maximum likelihood estimation of source reliability and claim correctness

Evaluation

Baseline: Voting

Data credibility is estimated by the number of times the same tweet is collected from the human network. The larger the repetition, the more credibility is attributed to the content

Voting: it counts both retweets and original tweets as full votes

Voting-NoRT: it only counts the original tweets

Evaluation

Baseline: Expectation Maximization (EM)

Data credibility is estimated by maximazing the likelihood of tweets collected from the human network

Regular-EM: it assumes that all sources constitute independent observers

Regular-EM-AD: it removes dependent sources using some heuristic approaches from social networks (*admission control)*

Trace	Sandy	Irene	Egypt Unrest
Start Date	11/2/2012	8/26/2011	2/2/2011
Time duration	14 days	7 days	18 days
# of tweets	12,931	269,308	93,208
# of users twitted	7,583	207,562	13,836
# of follower-followee links	37,597	3,902,713	10,490,098
# of users crawled	704,941	2,510,316	5,285,160

TABLE II. STATISTICS OF THREE TRACES

Methodology

Choose the event type of interest

Select keywords and, possibly, geographic location

Crawl Twitter using the available API

Log collected tweets

Apply tweet clustering to determine claims

Build *SC* and *SD* graphs

Apply MLE filtering

Validation Method

The output of filtering was manually graded to determine match with ground truth. Due to man-power limitations, manually grading only the 150 top ranked claims using the following rubric:

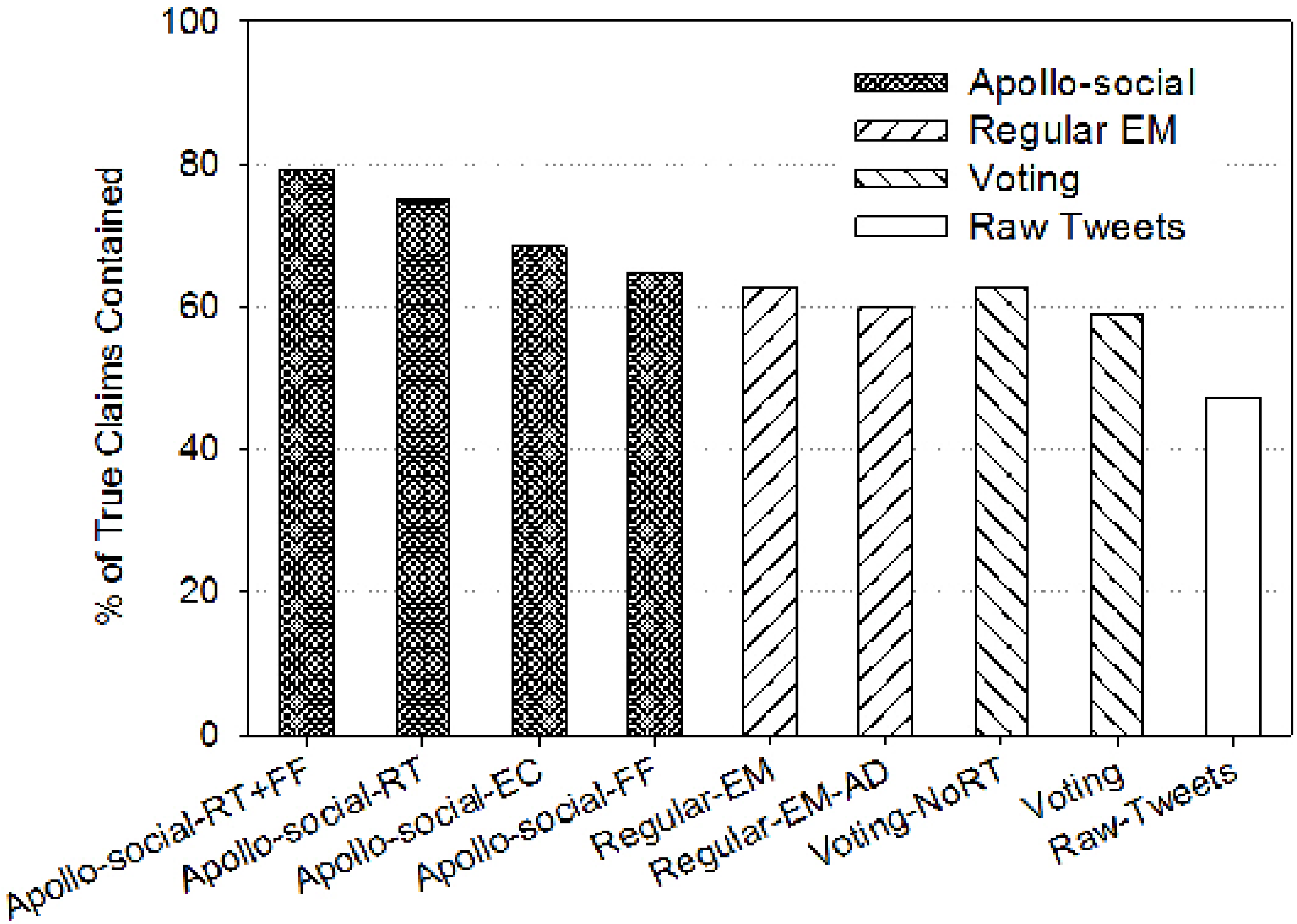
True claims: Claims that describe a physical or social event that is generally observable by multiple independent observers and corroborated by sources external to the experiment (e.g., mainstream news media)

Unconfirmed claims: Claims that do not meet the definition of true claims. May include true claims that cannot be verified

#	Media	Tweet found by Apollo-social	Tweet found by Regular EM
1	Rockland County Executive C. Scott Vanderhoef is announcing a Local Emergency Order restricting the amount of fuel that an individual can purchase at a gas station.	Rockland County Orders Restrictions on Gas Sales - Nyack-Piermont, NY Patch http://t.co/cDSrqpa2	**MISSING**
2	New York City Mayor Michael Bloomberg has announced that the city will impose an indefinite program of gas rationing after fuel shortages led to long lines and frustration at the pump in the wake of superstorm Sandy.	Gas rationing plan set for New York City: The move follows a similar announcement last week in New Jersey to eas... http://t.co/nkmF7U9I	RT @nytimes: Breaking News: Mayor Bloomberg Imposes Odd-Even Gas Rationing Starting Friday, as Does Long Island http://t.co/eax7KMVi
3	New Jersey authorities filed civil suits Friday accusing seven gas stations and one hotel of price gouging in the wake of Hurricane Sandy.	RT @MarketJane: NJ plans price gouging suits against 8 businesses. They include gas stations and a lodging provider.	**MISSING**
4	The rationing system: restricting gas sales to cars with even-numbered license plates on even days, and odd-numbered on odd days will be discontinued at 6 a.m. Tuesday, Gov. Chris Christie announced on Monday.	# masdirin City Room: Gas Rationing in New Jersey to End Tuesday # news	RT @nytimes: City Room: Gas Rationing in New Jersey to End Tuesday http://t.co/pYIVOmPo
5	New Yorkers can expect gas rationing for at least five more days: Bloomberg.	Mayor Bloomberg: Gas rationing in NYC will continue for at least 5 more days. @eyewitnessnyc #SandyABC7	Bloomberg: Gas Rationing To Stay In Place At Least Through The Weekend http://t.co/mmqqjYRx

TABLE III. GROUND TRUTH EVENTS AND RELATED CLAIMS FOUND BY APOLLO-SOCIAL VS REGULAR EM IN SANDY

Precision of the Top Claims



Observations and Limitations

The scheme tends to reduce the number of introspective (e.g., emotional and opinion) tweets, compared to tweets that presented descriptions of an external world

The reason may be that emotions and slogans tend to be retweeted and hence, tend to be suppressed by the scheme

In contrast, external facts are often observed independently by multiple sources, and hence not suppressed

Distortions and biases of human sensors are quite persistent in terms of direction and amplitude, unlike white noise distortions. In a sense, *humans exaggerate information in predictable ways*

Observations and Limitations

The FF, RT and RT+FF schemes can be improved by building information propagation models that account for:

- topic (e.g., sensational, emotional, and other news might propagate along different models)
- expertise
- relationship reciprocity
- mutual trust
- personal bias (e.g., the claim that «police is beating up innocent demonstrators», versus «demonstrators are attacking the police»)

References

- D. Wang et al., "Using humans as sensors: An estimation-theoretic perspective," IPSN-14 Proceedings of the 13th International Symposium on Information Processing in Sensor Networks, Berlin, 2014, pp. 35-46.
- Avvenuti M, Cimino MG, Cresci S, Marchetti A, Tesconi M., "A framework for detecting unfolding emergencies using humans as sensors". *Springerplus*. 2016;5:43.