Some Like It Hoax:

Automated Fake News Detection in Social Networks





Eugenio Tacchini

Gabriele Ballarin

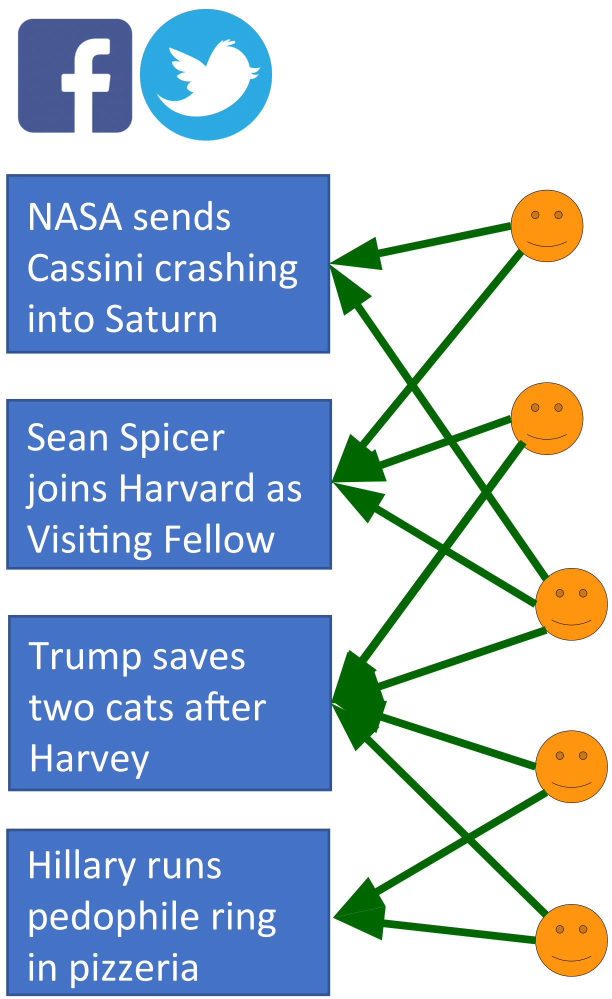
Marco Della Vedova

Stefano Moret

Luca de Alfaro

News spread through social networks...

...and some of them are fake

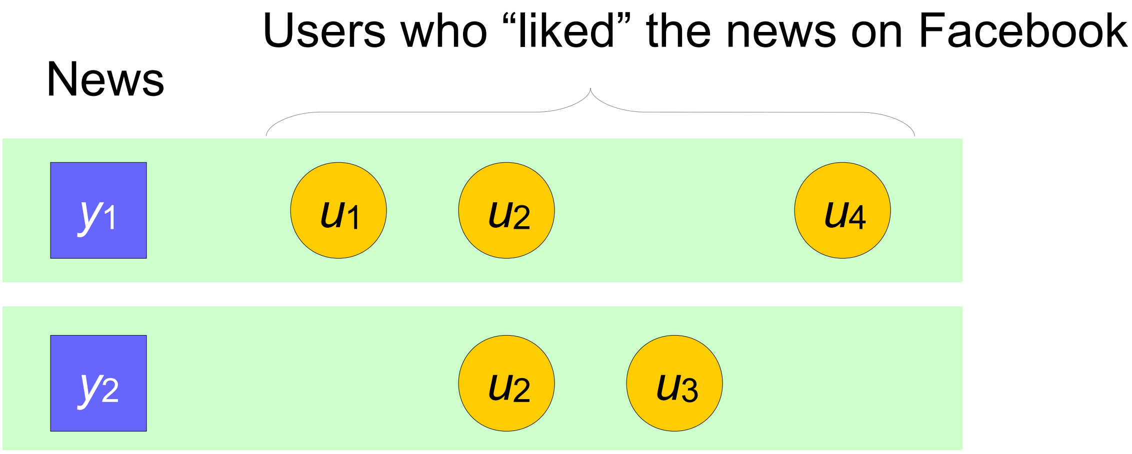


Can we identify fake news automatically?

- Text analysis is hard: many fake news read like real ones.
- Can we use social signals instead?
- Can we identify fake news on the basis of the users who share/like them?

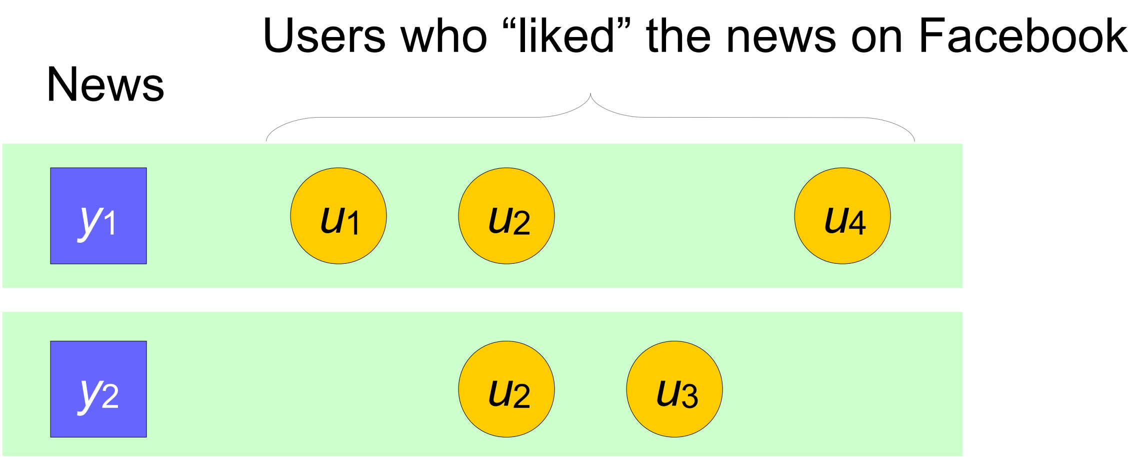
→ We answer in the affirmative, introducing techniques that can be shown to work with high accuracy

Technique 1: Logistic Regression



- Users who liked news articles are the “features”

Technique 1: Logistic Regression



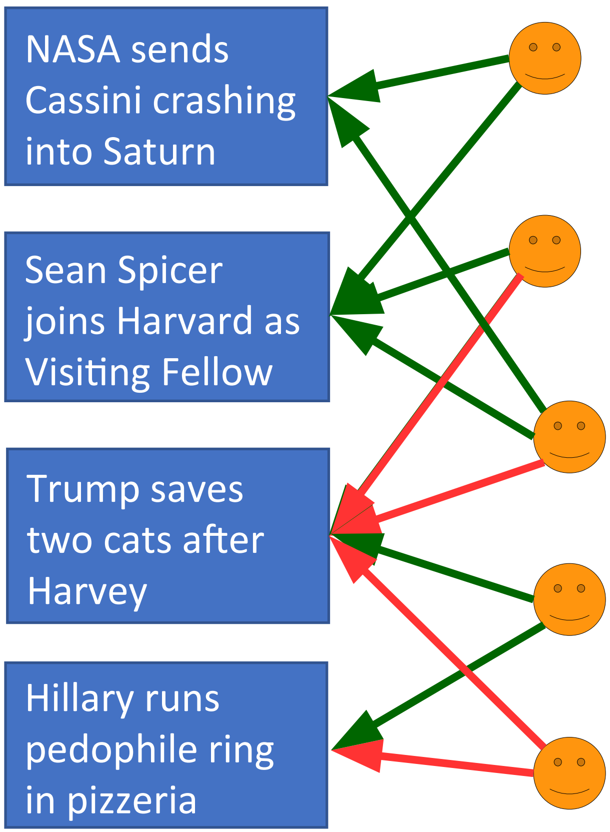
- Users who liked news articles are the “features”
- Logistic regression:

$$logit(y_1) = w_1 \cdot 1 + w_2 \cdot 1 + w_3 \cdot 0 + w_4 \cdot 1$$

$$logit(y_2) = w_1 \cdot 0 + w_2 \cdot 1 + w_3 \cdot 0 + w_4 \cdot 0$$

Train on news of known True/Fake value, and compute the user coefficients $w_1, w_2, \ldots, w_n$. Use the coefficients to classify other news.

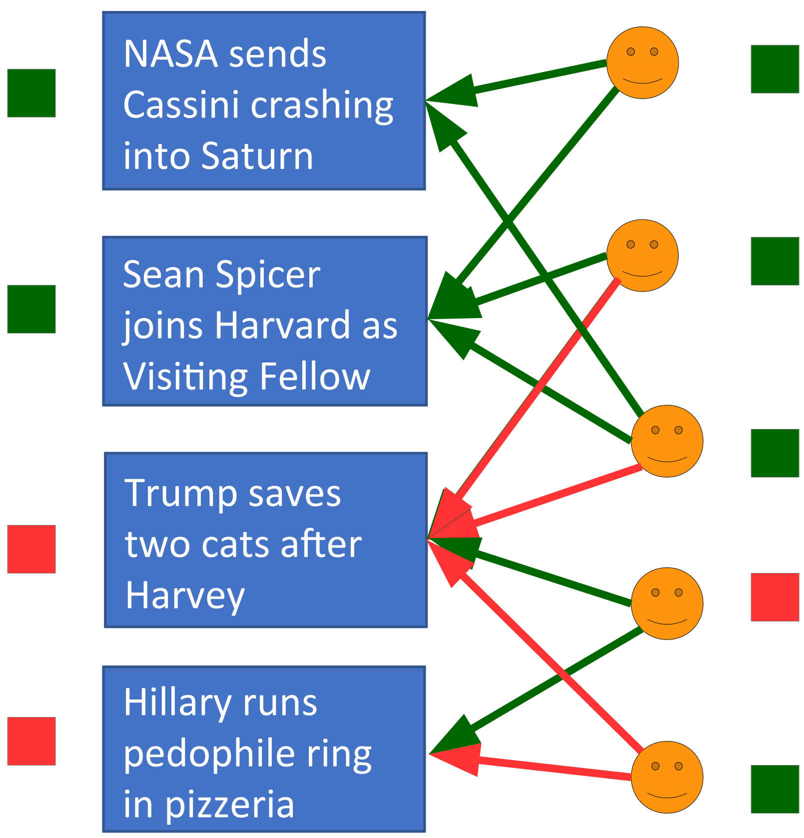
Technique 2: Harmonic Crowdsourcing



In boolean crowdsourcing:

- Users vote yes/no

Technique 2: Harmonic Crowdsourcing



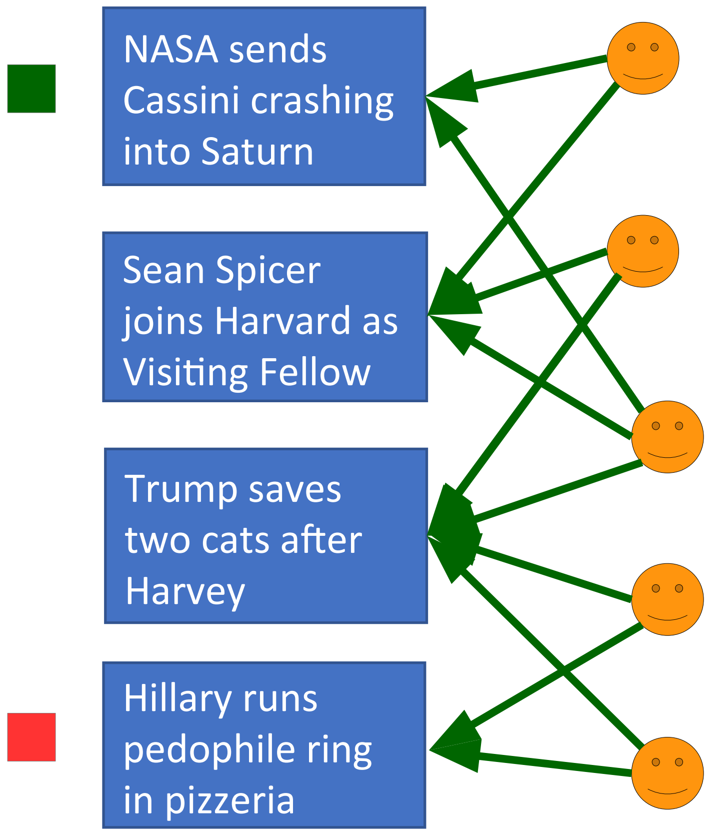
In boolean crowdsourcing:

- Users vote yes/no

The algorithms compare the votes, and reconstruct:

- Items → True, False
- Users → Truthful, Liar

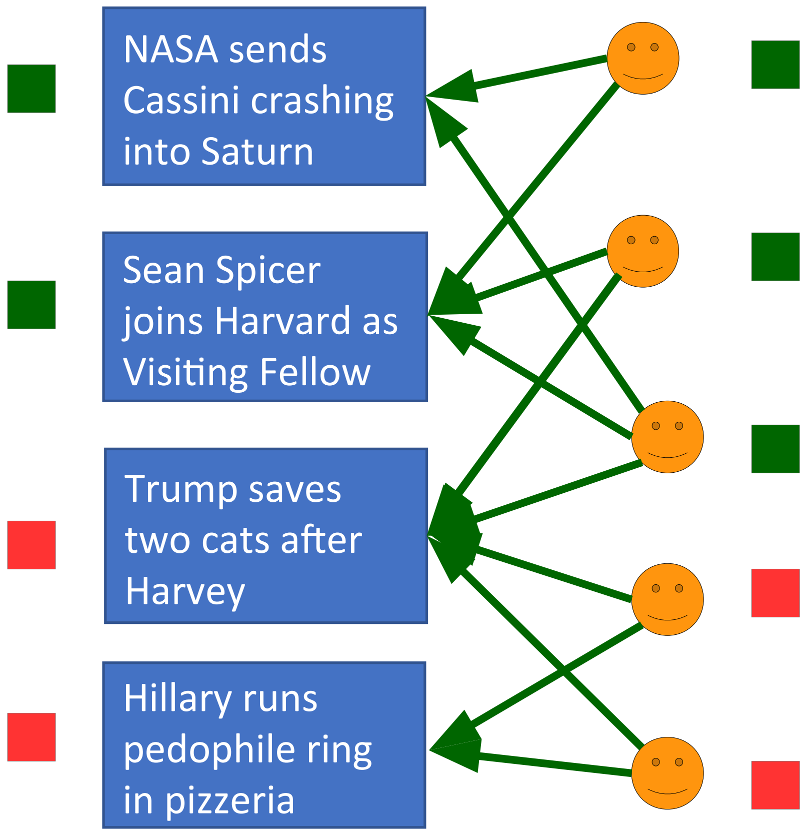
Technique 2: Harmonic Crowdsourcing



In our application:

- Users vote yes (like / share)
- The algorithms starts from a ground truth

Technique 2: Harmonic Crowdsourcing



In our application:

- Users vote yes (like / share)
- The algorithms starts from a ground truth

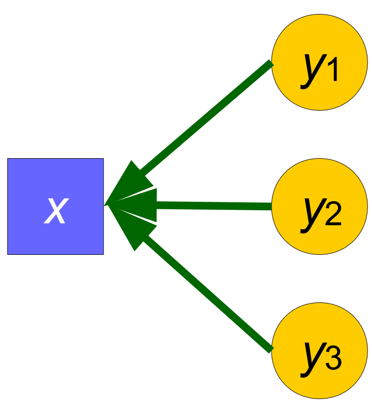
The algorithm computes:

- Items → True, False
- Users → Truthful, Liar

We use the "harmonic crowdsourcing" of [de Alfaro, Polychronopoulos HCOMP 2015]: simple, robust, scalable, and almost optimal.

Technique 2: Harmonic Crowdsourcing

- Model the truth of each node x (a user, or a news item) with a beta distribution *beta*(α_x, β_x), where:
  - α_x is the "positive" evidence for the truth of x.
  - β_x the "negative" evidence for the truth of x.
- The node x has truth $\sim$*beta*(α_x, β_x) with average value $\alpha_x / (\alpha_x + \beta_x)$
- Valuation propagation:



$$\alpha_x = \kappa + \sum_i \{ y_i - \tfrac{1}{2} \mid y_i > \tfrac{1}{2} \}$$

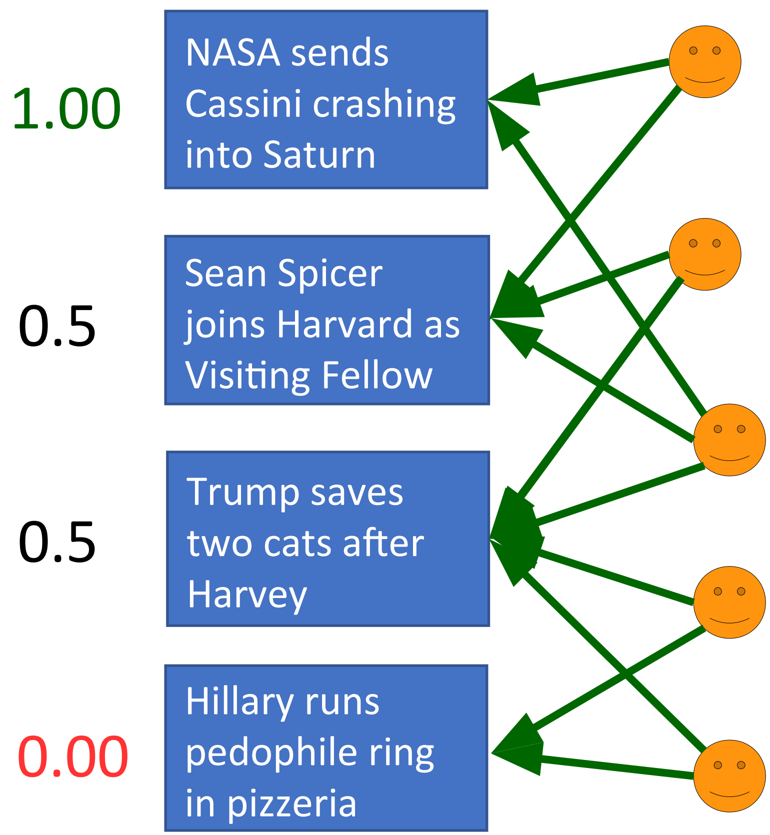
$$\beta_x = \kappa + \sum_i \{ \tfrac{1}{2} - y_i \mid y_i < \tfrac{1}{2} \}$$

$$x = \alpha_x / (\alpha_x + \beta_x)$$

κ : inertia of null belief.

Technique 2: Harmonic Crowdsourcing

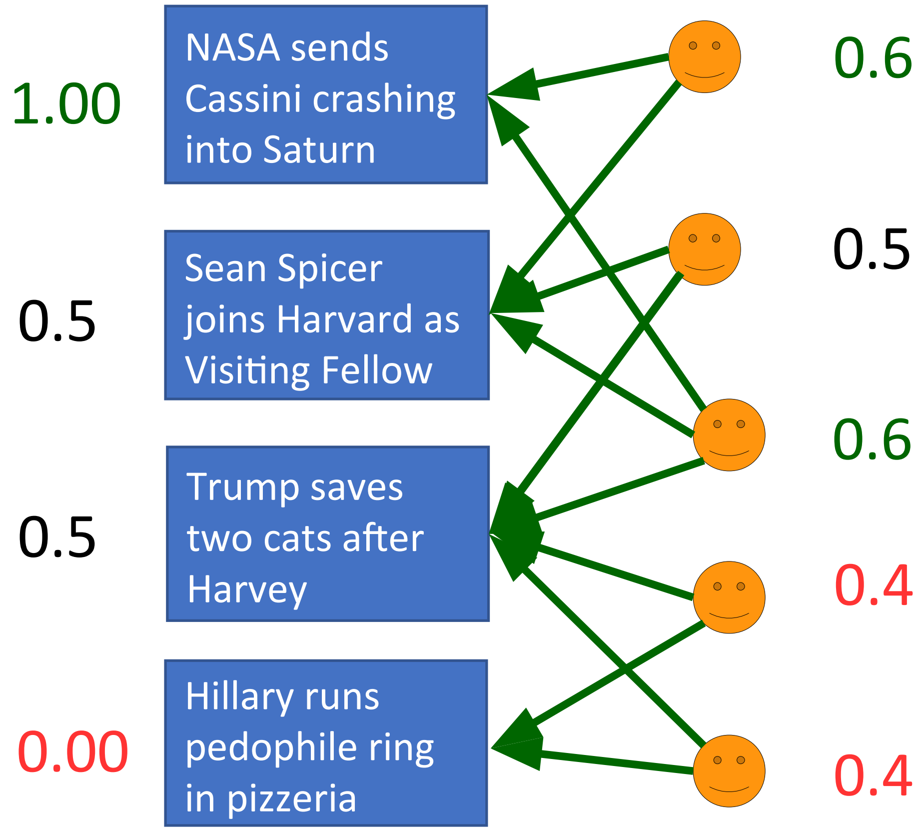
Propagation:



Label ground truth with 1, 0; the rest with 0.5

Technique 2: Harmonic Crowdsourcing

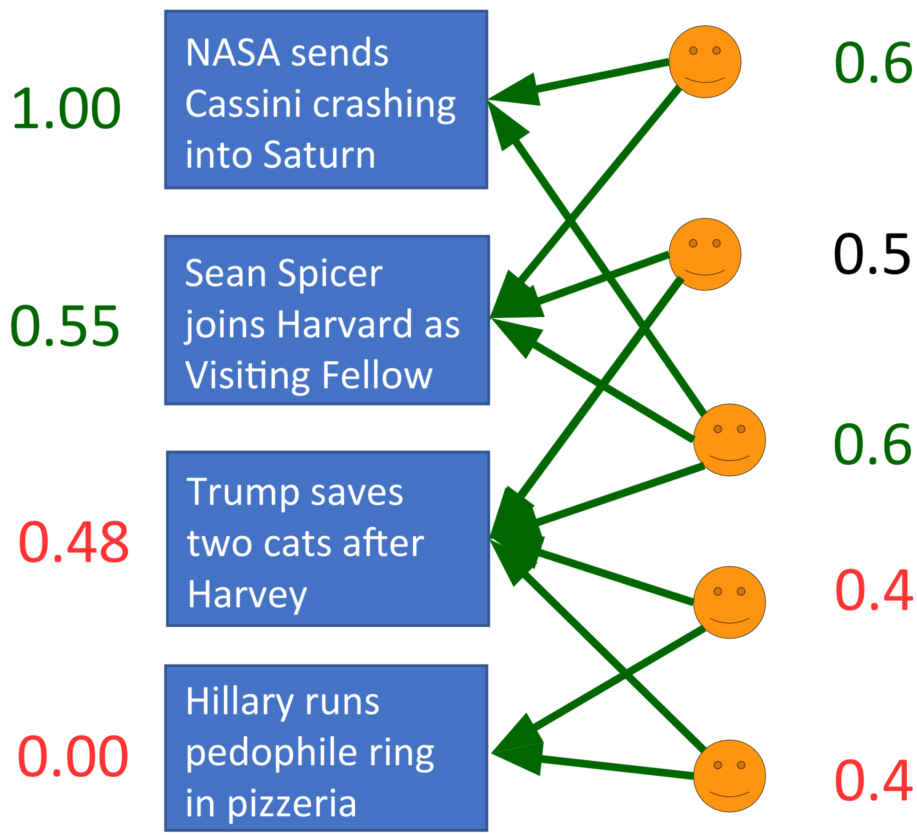
Propagation:



Iterate: items → users

Technique 2: Harmonic Crowdsourcing

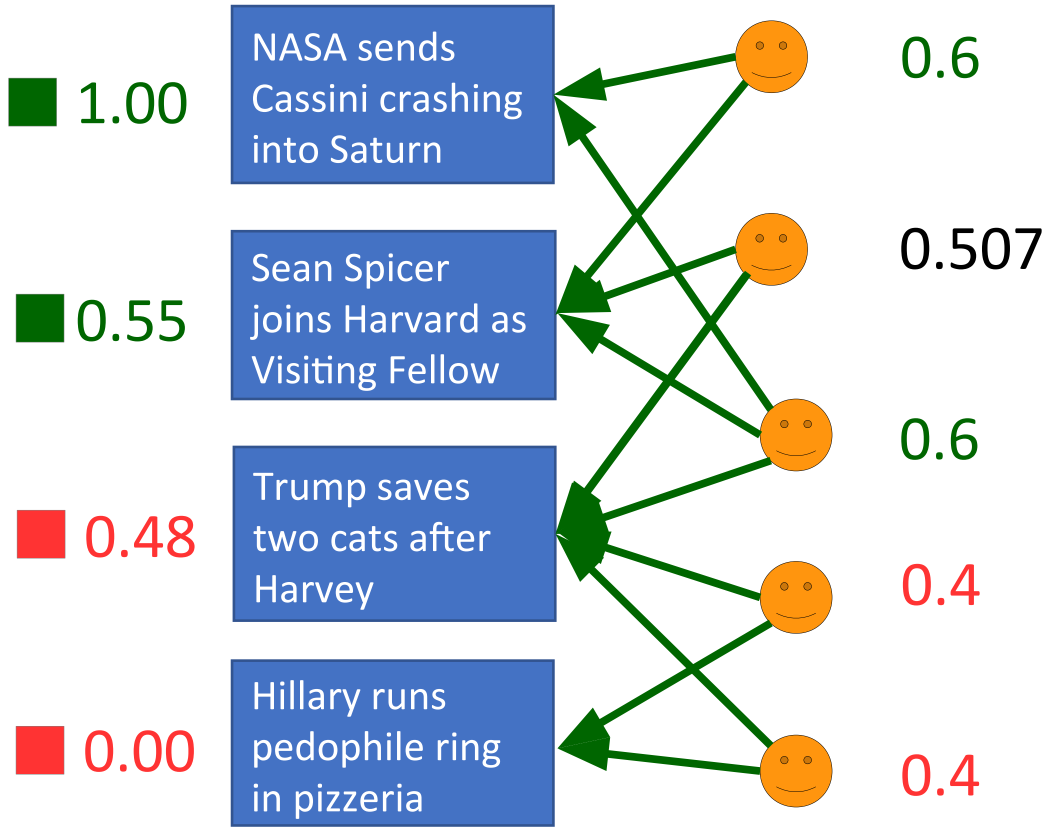
Propagation:



Iterate: users → items

Technique 2: Harmonic Crowdsourcing

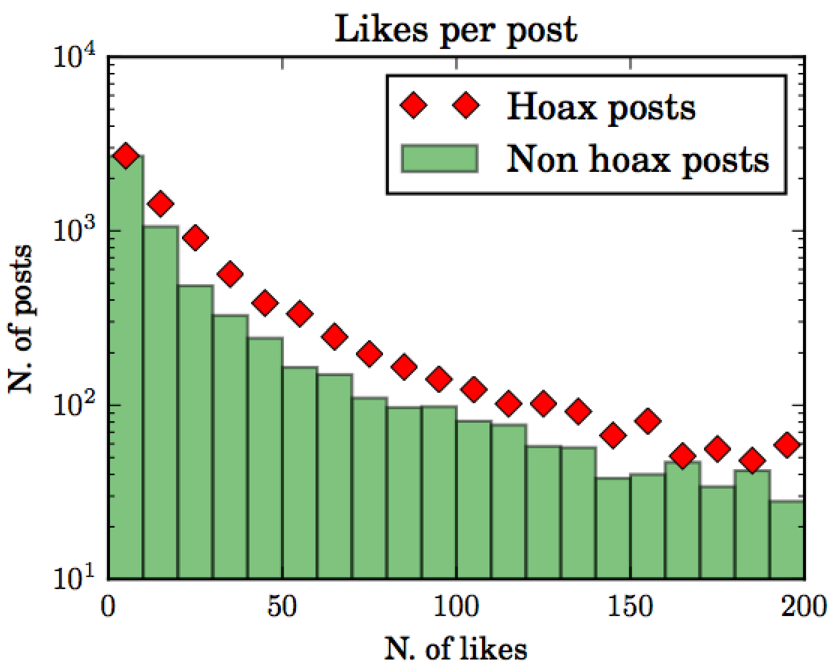
Propagation:

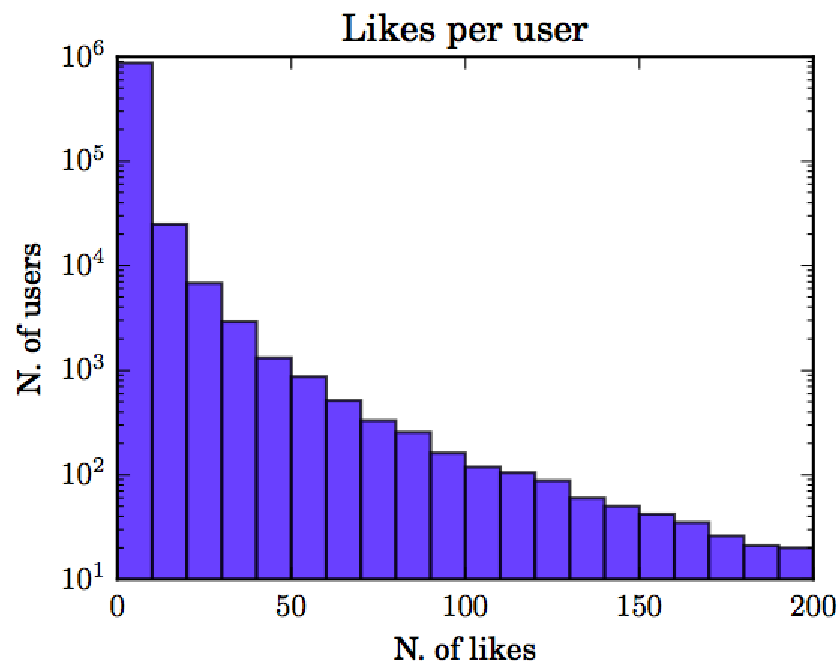


Iterate: items → users … fixpoint is soon reached.

The Dataset

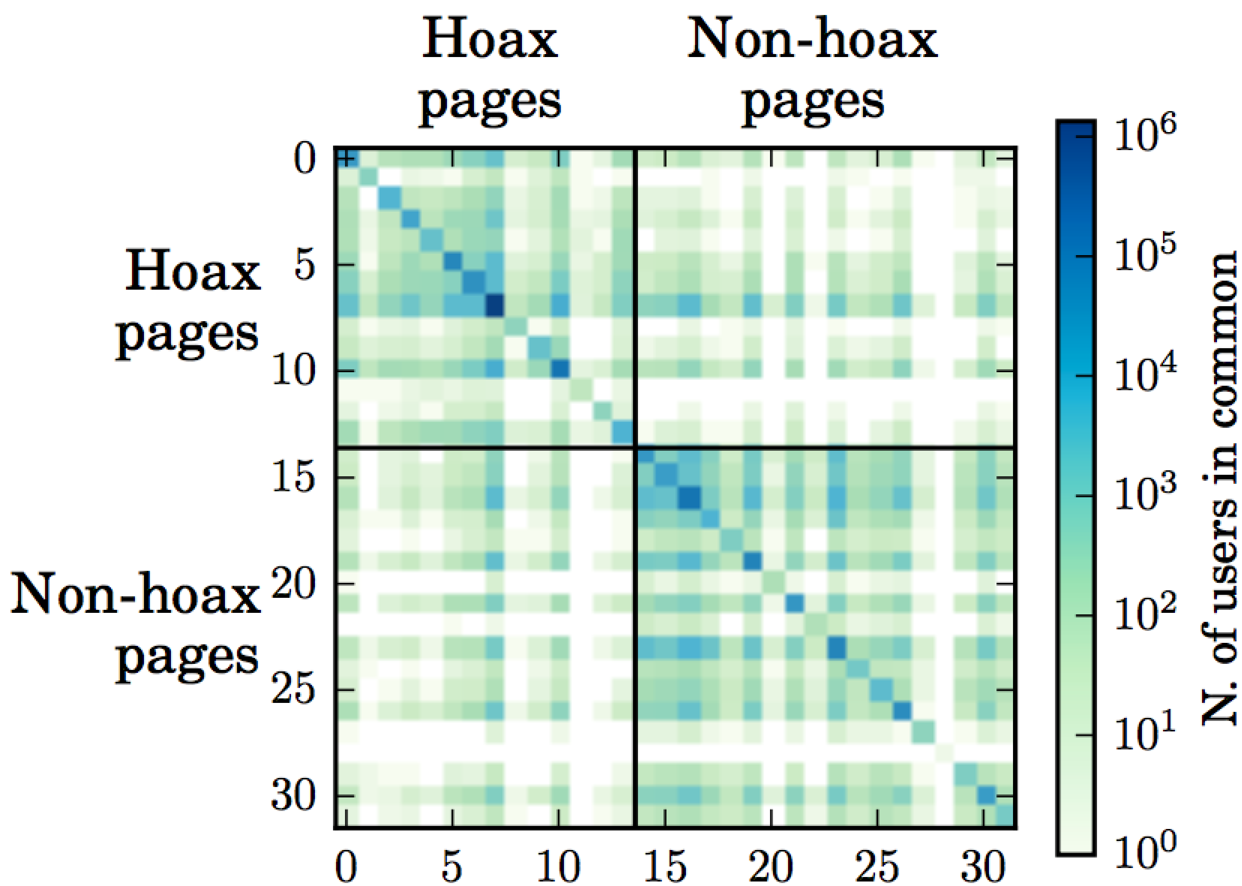
- All posts in a set of Facebook pages from Jul 1-Dec 31, 2015 on science vs. conspiracy.
- The pages are from [Science vs Conspiracy: Collective Narratives in the Age of Misinformation. Bessi et al, PLOS ONE, 2105]



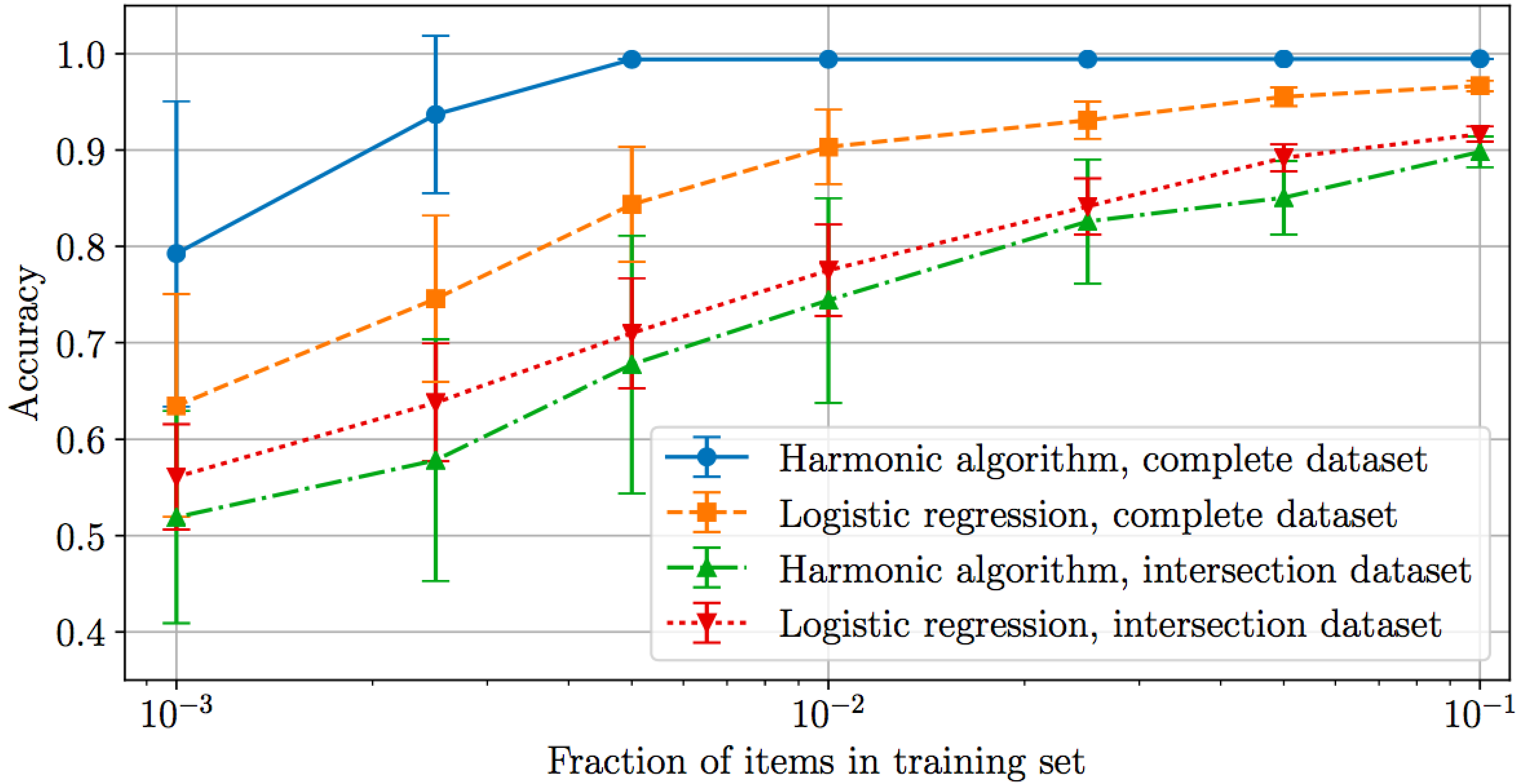


We consider two sub-datastes:

- **Complete dataset:** 15,500 posts, 909,236 users, 2.3M likes
- **Intersection dataset:** only the users who liked *both* hoax and non-hoax posts: 10,520 posts, 14,149 users, 118k likes.



Results



Accuracy
1.0
0.9
0.8
0.7
0.6
0.5
0.4
10^{-3}
10^{-2}
10^{-1}
Fraction of items in training set
Harmonic algorithm, complete dataset
Logistic regression, complete dataset
Harmonic algorithm, intersection dataset
Logistic regression, intersection dataset

Conclusions

- Our emphasis is in obtaining high accuracy with as small as possible a training set.
  - Labeling even a small percentage of the daily news by hand is a large task, so the smaller the labeled fraction required, the better.
- We obtain an accuracy of 99% even when < 0.5% of the dataset is used for training.
  - The harmonic algorithm is very efficient in spreading knowledge across the “likers” graph.
- Even on the artificial *intersection* dataset, consisting only of people who liked both hoax and non-hoax posts, the accuracy is high: 90% with 10% in training set, ~75% with 1% in training set.