Trust Evaluation through User Reputation and Provenance Analysis

Davide Ceolin, Paul Groth, Willem Robert van Hage, Archana Nottamkandath, Wan Fokkink

VU UNIVERSITY AMSTERDAM

Outline

- Problem Definition
- Trust estimates as probability distributions
- Reputation-based trust
- Provenance-based trust
- Combining reputation- and provenance-trust
- Results
- Future Work

Our Problem

Trusting crowdsourced information

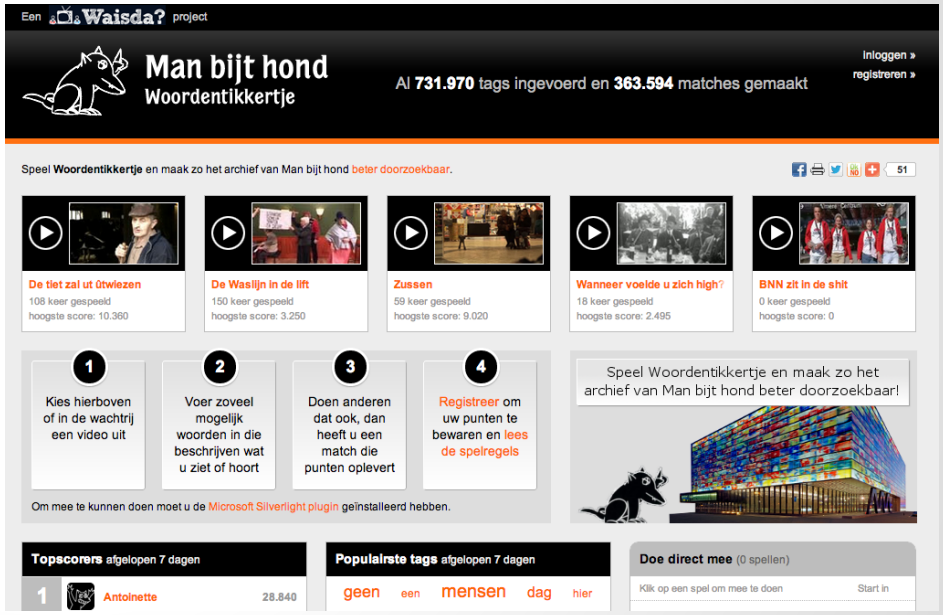
Specifically: video annotations

We can not tag videos automatically (complex, many layers,...), so we refer to the crowd, but crowdsourced annotations need to be evaluated.

Our Case Study

Waisda, a video tagging platform.
http://waisda.nl

Users score if they
tag simultaneously.



Tags are useful (e.g. to index video) but how to trust tags?

Measuring tags trustworthiness

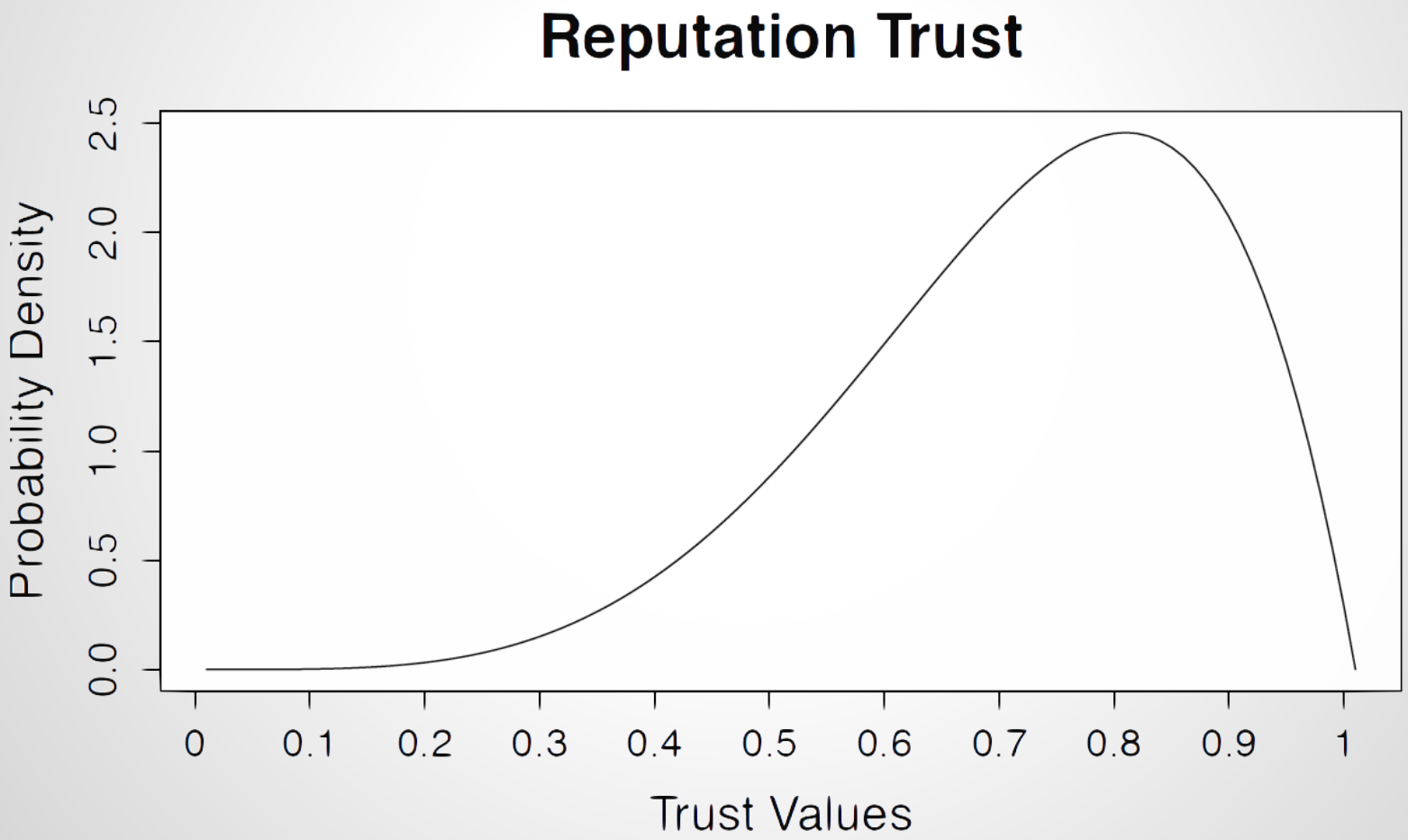
How to measure tags trustworthiness?

Let's exploit the game mechanism:

$$\#\text{matches} \approx \text{trustworthiness}$$

However, if a tag did not get any match it is not necessarily wrong.

Trust as a Beta Distribution



Reputation-based trust

A classical approach:

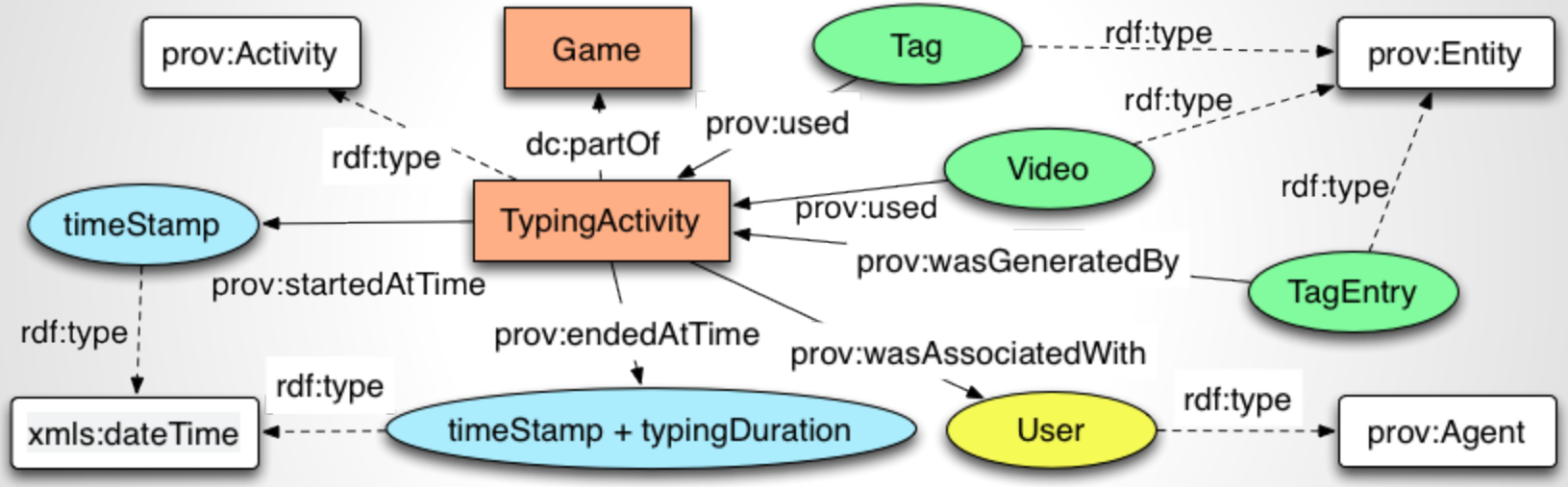
- Gather all the evidence about a user
- Estimate the Beta distribution (user reputation)
- Predict the trustworthiness of the (other) tags inserted by the user (using the Beta)

Our hypothesis

We believe we can go beyond the classical reputation-based trust:

Who ≤ Who + How

(How) Provenance-based trust

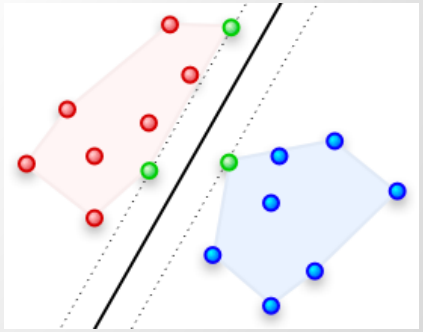


- “How”-provenance can be used as a stereotype for tags authorship
- May be less precise than reputation but more easily available

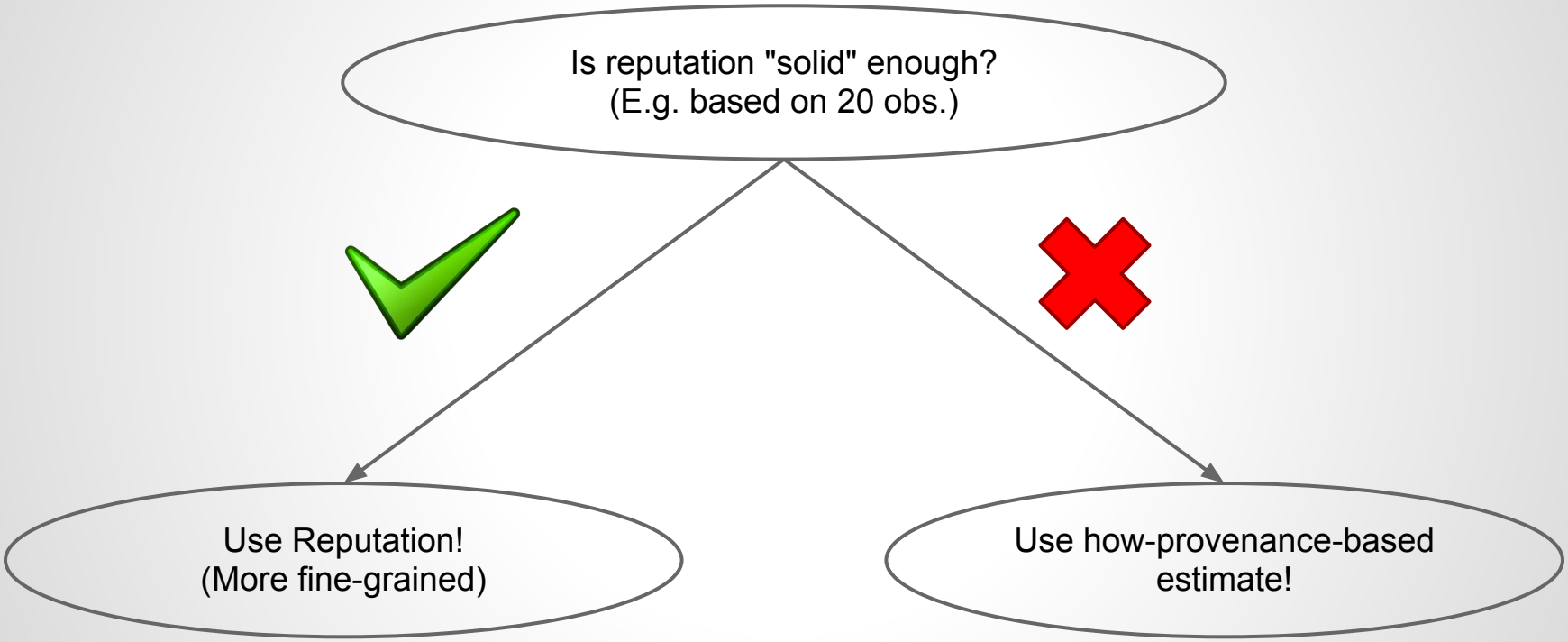
Computing provenance-based trust

“How-provenance” is composed by several “features” (timestamp, typingDuration, ...)

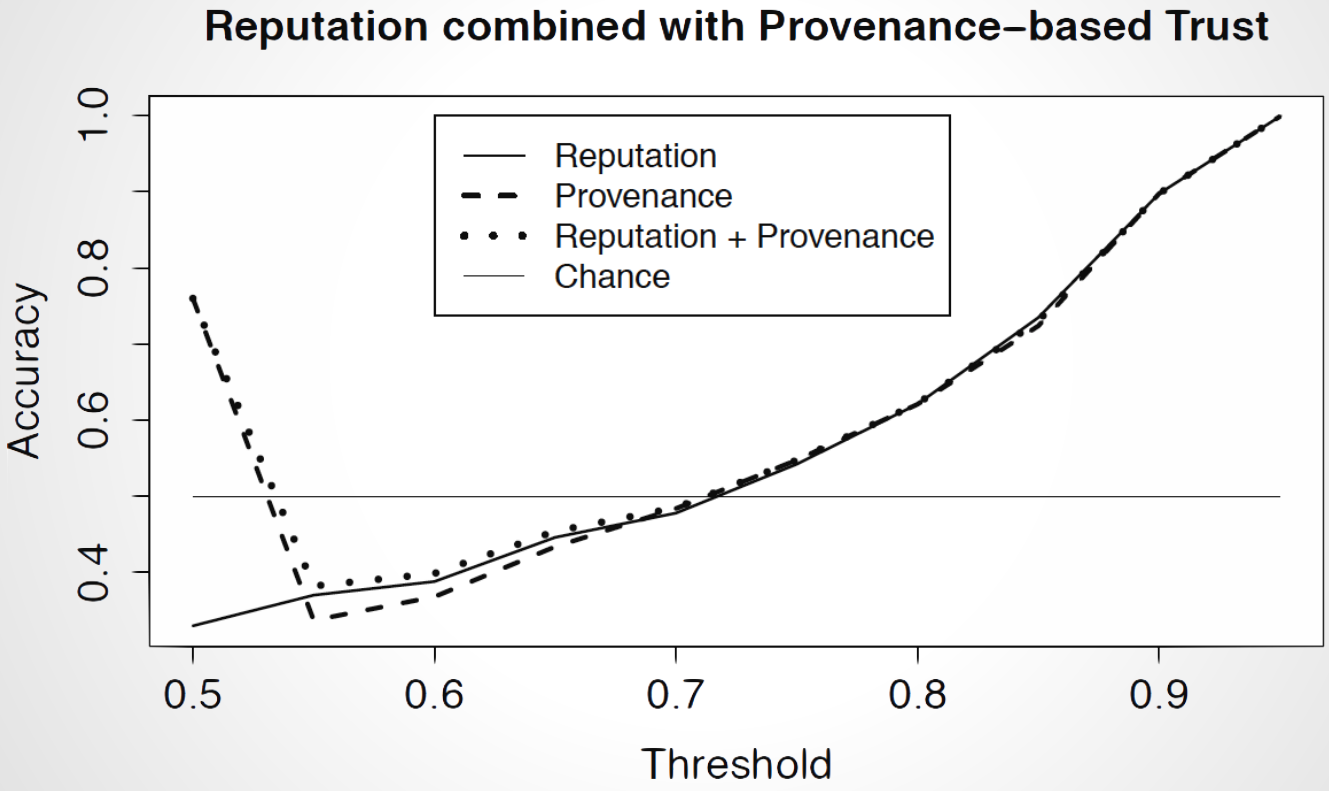
We use Support Vector Machines to learn the model and make our estimates.



Combining reputation- and provenance-trust



Results



$$\text{Accuracy} = \frac{\#\{\text{tag: Real\_Trust(tag)}>\text{Threshold} = \text{Pred\_Trust(tag)} > \text{Threshold}\}}{\#\text{tags}}$$

Discussion

Who =* How

Who ≤* Who + How 

How ≤* Who + How

(At least in our case study)

=* statistically not different
≤* statistically different (less)

Future Work

- We plan to apply this approach to other case studies as well
- A generic platform for producing trust assessments based on annotated provenance graphs is under development
- We also plan to merge provenance-based estimates with semantic similarity-based ones in order to provide a tool for a wide range of applications.

Thank you!
Questions?

d.ceolin@vu.nl

VU UNIVERSITY AMSTERDAM