Readersourcing: Crowdsourcing peer review (and other things)

Stefano Mizzaro

Department of Maths, Computer Science, and Physics
University of Udine
mizzaro@uniud.it

enetCollect, Lisbon, Portugal, 15 Mar 2019

EUROPEAN COOPERATION IN SCIENCE & TECHNOLOGY

Funded by the Horizon 2020 Framework Programme of the European Union

This publication is based upon work from COST Action CA16105, supported by COST (European Cooperation in Science and Technology).

Another Elephant in the Room

Stefano Mizzaro

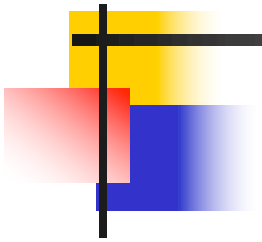
Department of Maths, Computer Science, and Physics
University of Udine
mizzaro@uniud.it

enetCollect, Lisbon, Portugal, 15 Mar 2019

EUROPEAN COOPERATION IN SCIENCE & TECHNOLOGY

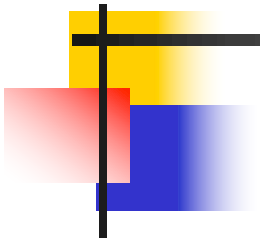
Funded by the Horizon 2020 Framework Programme of the European Union

This publication is based upon work from COST Action CA16105, supported by COST (European Cooperation in Science and Technology).



Outline

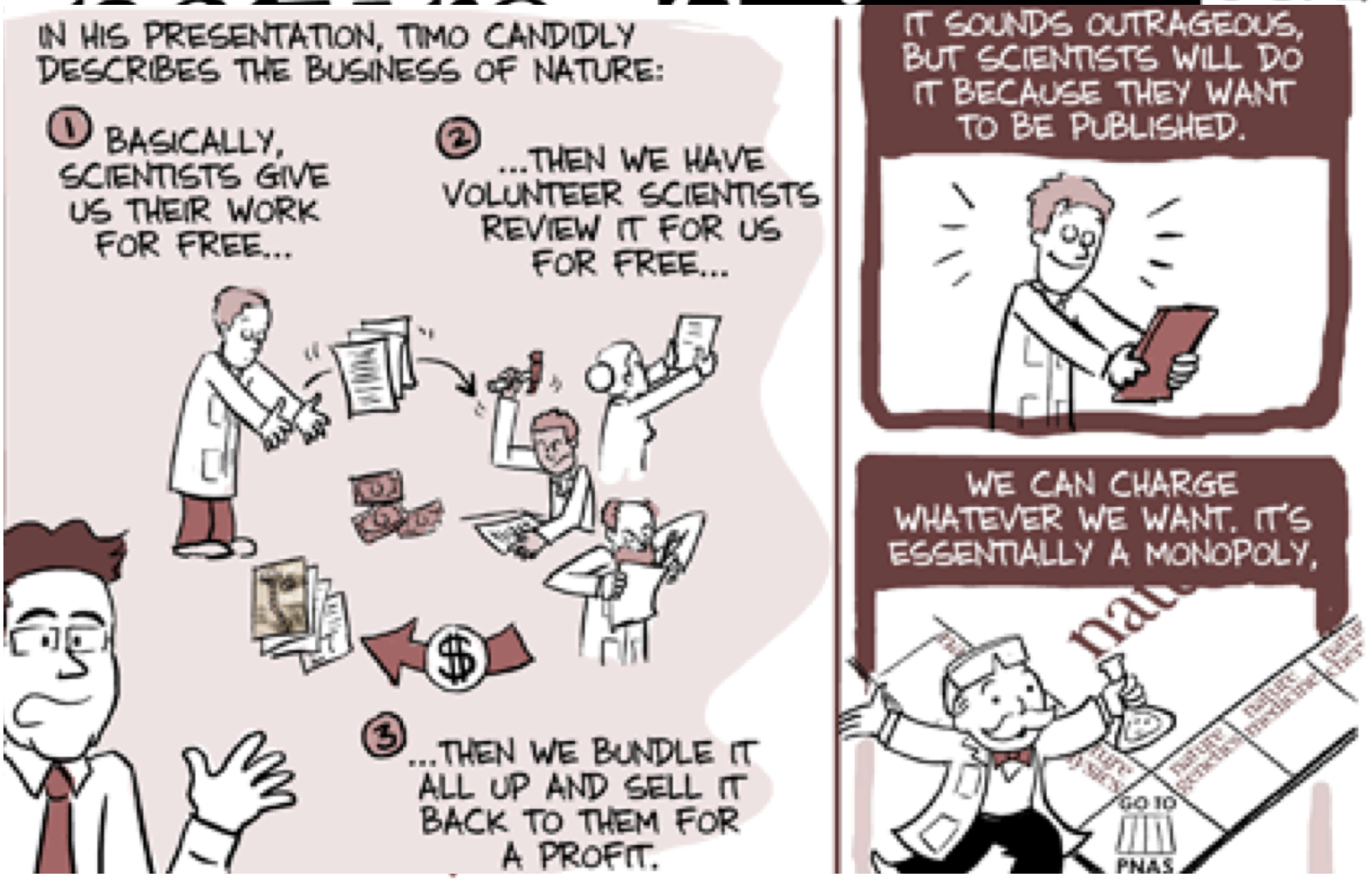
1. Readersourcing: Crowdsourcing Peer Review
2. The Quality Model
3. Peer assessment

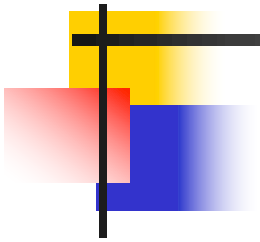


Scholarly publishing

- How do scientists work? We all know:
  - Idea, discovery, hard work, blablabla…
  - Write & submit (journal, conference, workshop,…)
  - **Peer review**
  - If accepted, publication
- Not only scientists → Scholars

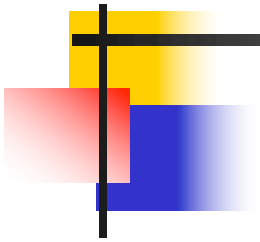
What we do…


IN HIS PRESENTATION, TIMO CANDIDLY DESCRIBES THE BUSINESS OF NATURE:
1 BASICALLY, SCIENTISTS GIVE US THEIR WORK FOR FREE...
2 ...THEN WE HAVE VOLUNTEER SCIENTISTS REVIEW IT FOR US FOR FREE...
3 ...THEN WE BUNDLE IT ALL UP AND SELL IT BACK TO THEM FOR A PROFIT.
IT SOUNDS OUTRAGEOUS, BUT SCIENTISTS WILL DO IT BECAUSE THEY WANT TO BE PUBLISHED.
WE CAN CHARGE WHATEVER WE WANT. IT'S ESSENTIALLY A MONOPOLY.
GO TO
PNAS



Peer review criticisms…

- … too many to fit in one slide…

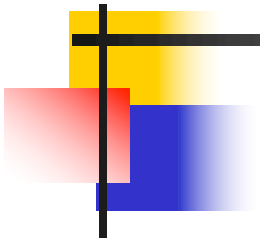


Time

- Editor: ‹‹Do you remember that paper that you submitted 1 year ago?››
- Me: ‹‹No!››
- Editor: ‹‹Oh well… Anyways, it has been rejected››

Wrong

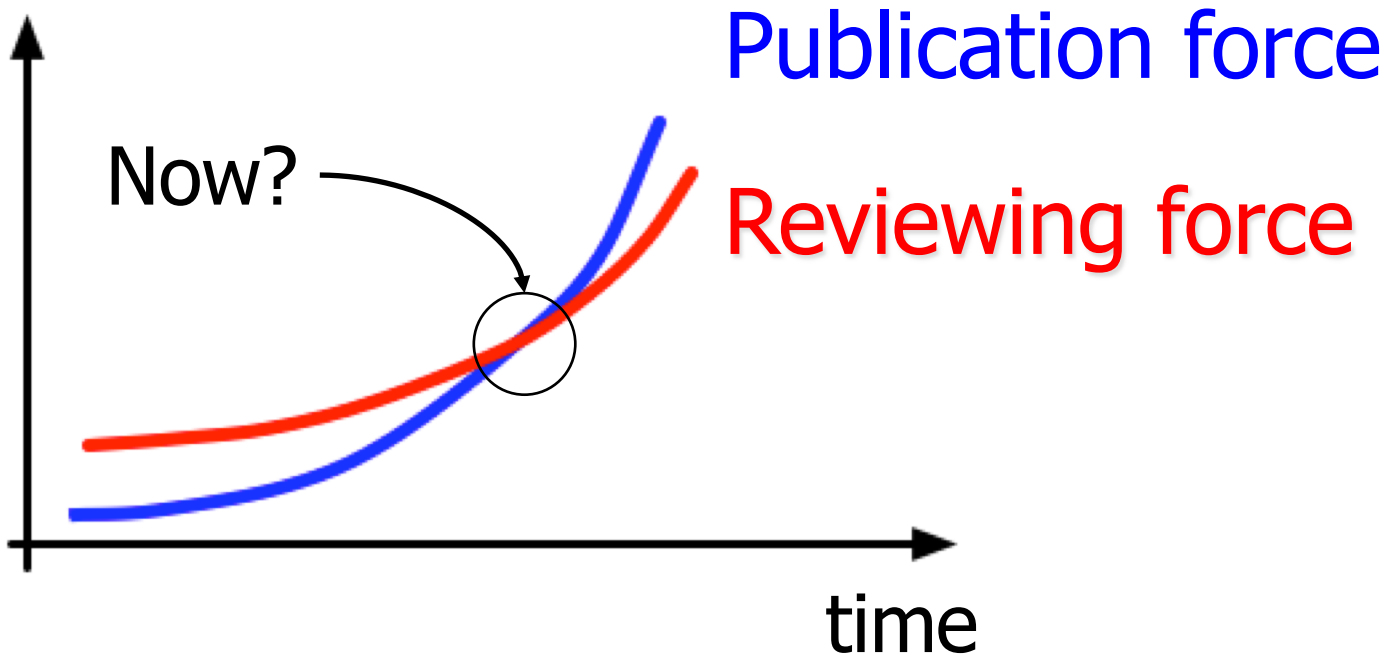
- We all know that



Impossible

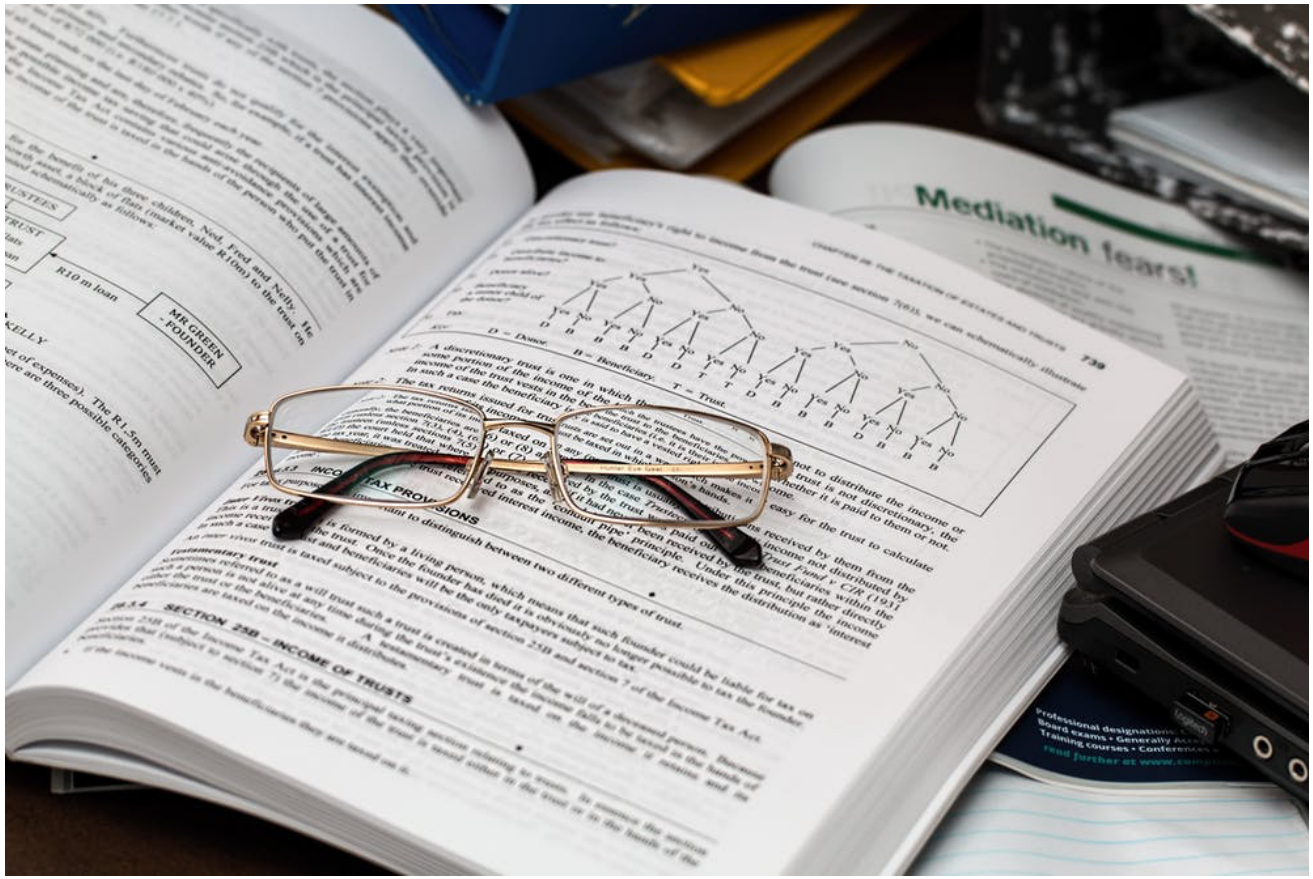
- Reviewers should re-do the experiments to see if the results hold?

(Good) referees are scarce resource

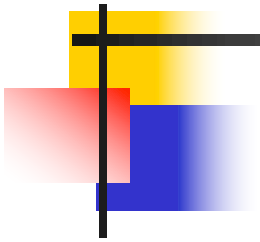


But wait a minute

- We read papers!
- What do we do with that?



Answer:



Paper ratings

- Yes, indeed we could collect ratings of scholarly papers

But wait another minute

- Readers rating papers?
- This is Peer review!
- This is Crowdsourcing peer review!

The Basic Idea

- We have plenty of **readers**!
  - They read papers
  - They have an opinion
  - They keep the opinion inside their own mind
- Quite strong reading (reviewing?!) force
- Not used at all (almost)
- Using a lot of readers in place of a few referees can be seen as **crowdsourcing**

Wait another another minute

- Peer review is not crowdsourced today
  - Still a few referees do the job (if you find them!)
- Even better (worse!): peer review **is** crowdsourced (readers read the papers!), but without exploiting the results (opinions are not logged, made public, exploited,…)

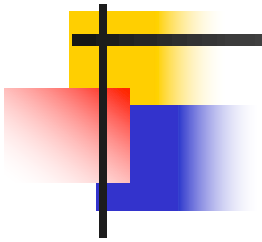
S. Mizzaro. Readersourcing – A **Manifesto**, JASIST, 63(8):1666-1672, 2012

The shoemaker's children go barefoot

- It's quite… strange that the Web tools / approaches that we developed are not used by us where they can be naturally applied…

But…

"On the Internet, nobody knows you're a dog."



Outline

1. Readersourcing: Crowdsourcing Peer Review
2. The Quality Model
3. Peer assessment

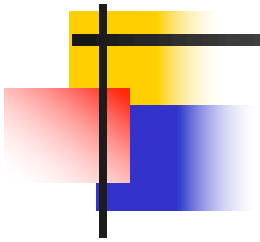
How to tell good readers from bad readers?

- If 200 PhD students say that my paper is good…
- … and 10 Experts say it is bad…
- … who should be trusted?
- Or: how to weigh appropriately good and bad readers?
- Or: how to avoid bad reviewers/readers?

Let's build on readers's reputation

In short

- Readers try to express the correct judgment…
- … because it is rewarding to be "a good reader"…
- … according to an objective measure
- 2 proposals / models:
  - [Mizzaro 2003, 2012]: Readersourcing
  - [De Alfaro & Faella 2016]: TrueReview



In short

- Readers try to express the correct judgment…
- … because it is rewarding to be "a good reader"…
- … according to an objective measure
- 2 proposals / models:
  - [Mizzaro 2003, 2012]: Readersourcing
  - [De Alfaro & Faella 2016]: TrueReview

Papers, authors, readers: 3 scores

- Each paper has a score, measuring its quality
  - Paper with high score ↔ good paper
  - High judgments by readers → high score (~ average)
- Each author has a score too
  - It changes accordingly to the scores of the papers published by the author (~ average paper score)
  - Publishing good papers → high score
- Each reader has a score too
  - Judgments by high scored readers are "heavier"
  - Reader score is a measure of its reviewing capability
- (Nothing really new so far…)

Feedback on readers

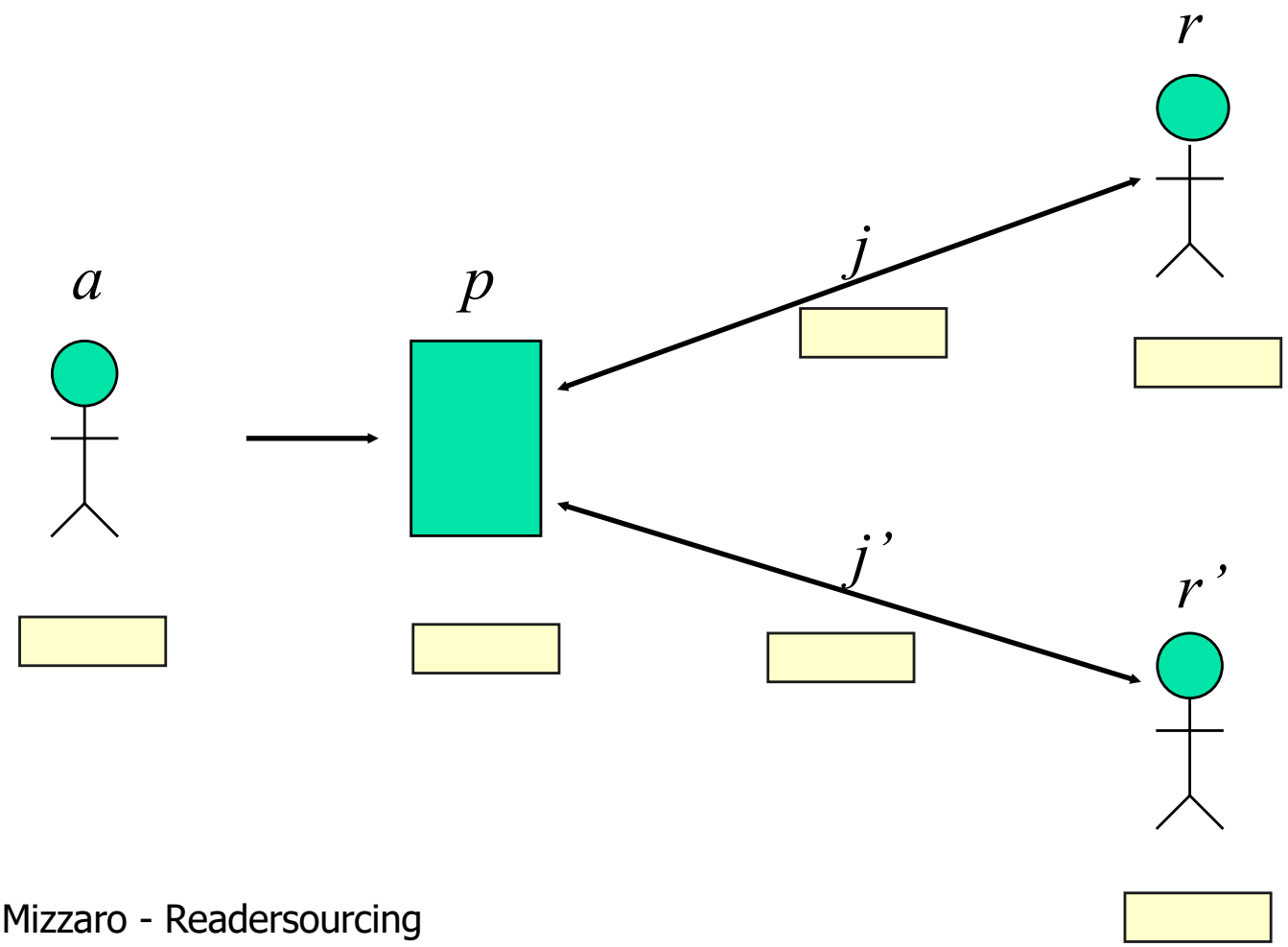
- Reader score changes
  - Accordingly to correctness of expressed judgments
  - Right judgments $\rightarrow$ higher reader score
  - Wrong judgments $\rightarrow$ lower reader score
- "Right" judgment?
  - Theoretically,
    - equal to the final paper score (the score that the paper will have at time = $+\infty$)
  - In practice,
    - the score at time = $+\infty$ is not available, but we can:
    - approximate it (with the current score)
    - revise the approximation over time as we get closer to $+\infty$

$$s_a(t) = \frac{\sum_{p \in P_a(t)} \left(\sum_{r \in R_p(t)} s_r(t_{r,p}) \cdot j_{r,p} \right)}{\sum_{p \in P_a(t)} \left(\sum_{r \in R_p(t)} s_r(t_{r,p}) \right)}$$

In one slide

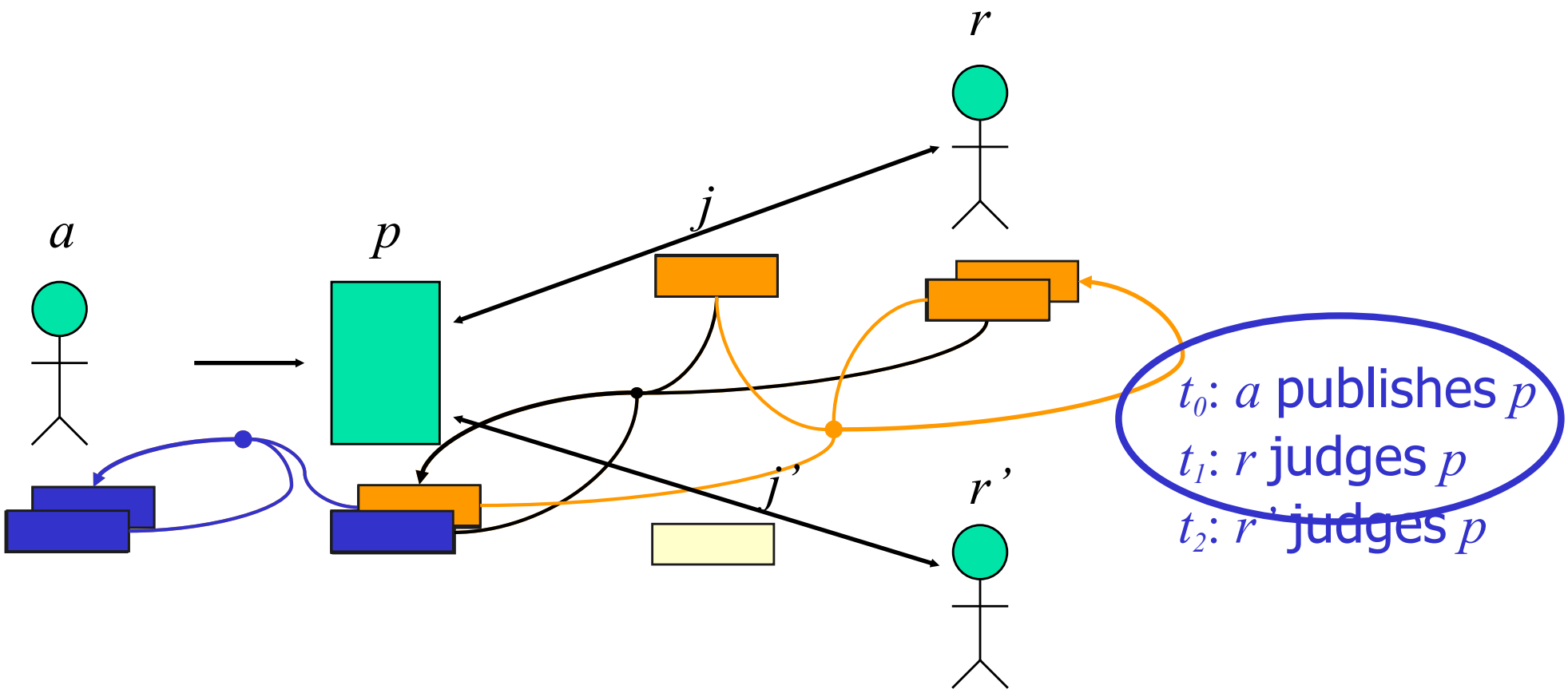
- Papers, authors, and readers have a score that measures their quality
  - (Steadiness: how stable the score is)
- Virtuous circle (hopefully)
  - Authors try to publish good papers
  - Readers try to express good/correct judgments ("they bet on the score the paper will have")
- Score of
  - Papers: which papers to read
  - Authors: "scientific productivity"
  - Readers: "scientific reputation"

A toy example

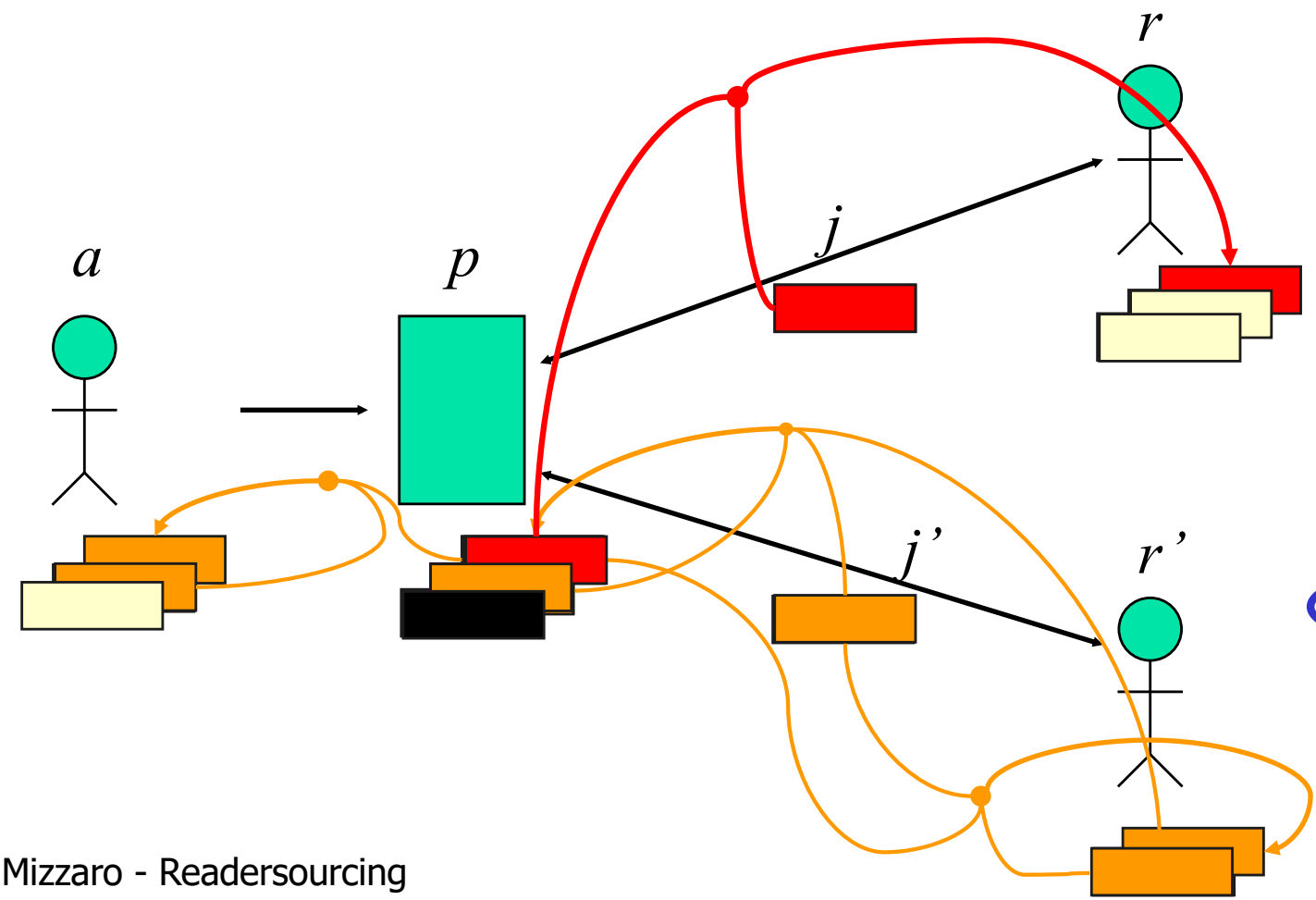


t_0: a publishes p
t_1: r judges p
t_2: r' judges p

A toy example (1/2)



A toy example (2/2)



t_0: a publishes p

t_1: r judges p

t_2: r' judges p

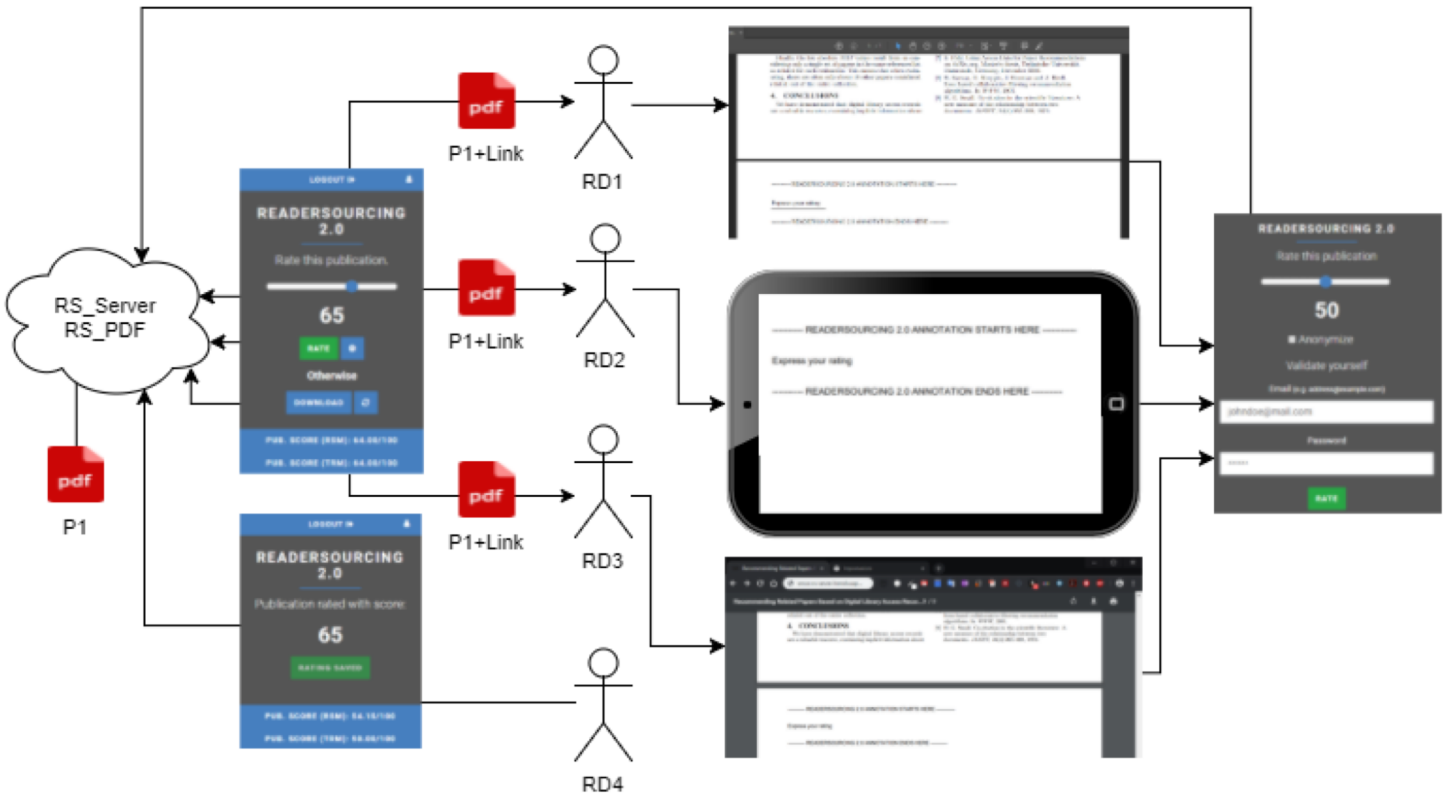
www.readersourcing.org

- Not only in theory!
  - (Still in beta)
    - (Well, alpha)
      - (almost!)

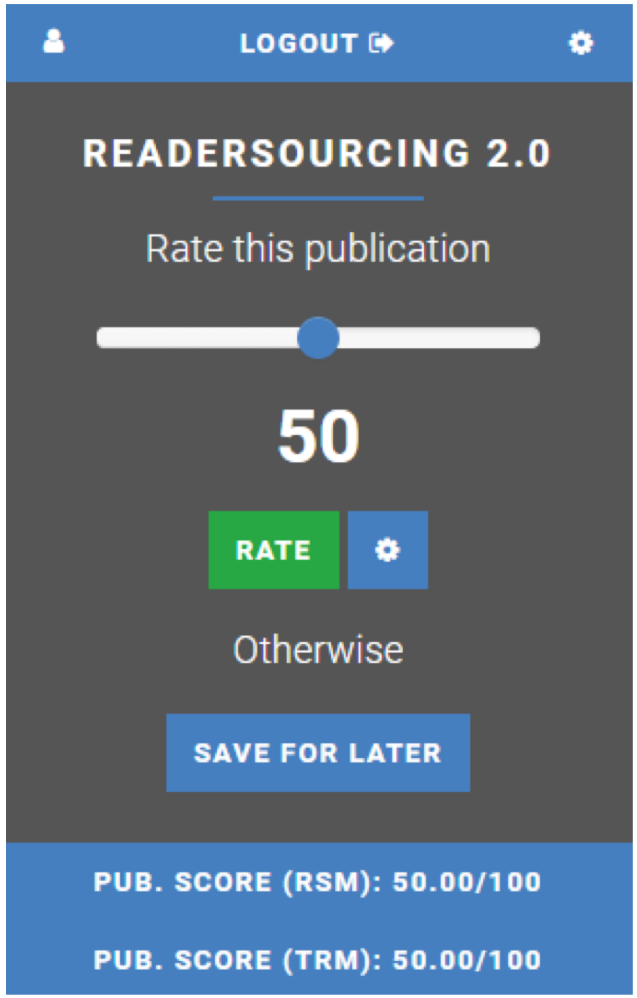
Michael Soprano

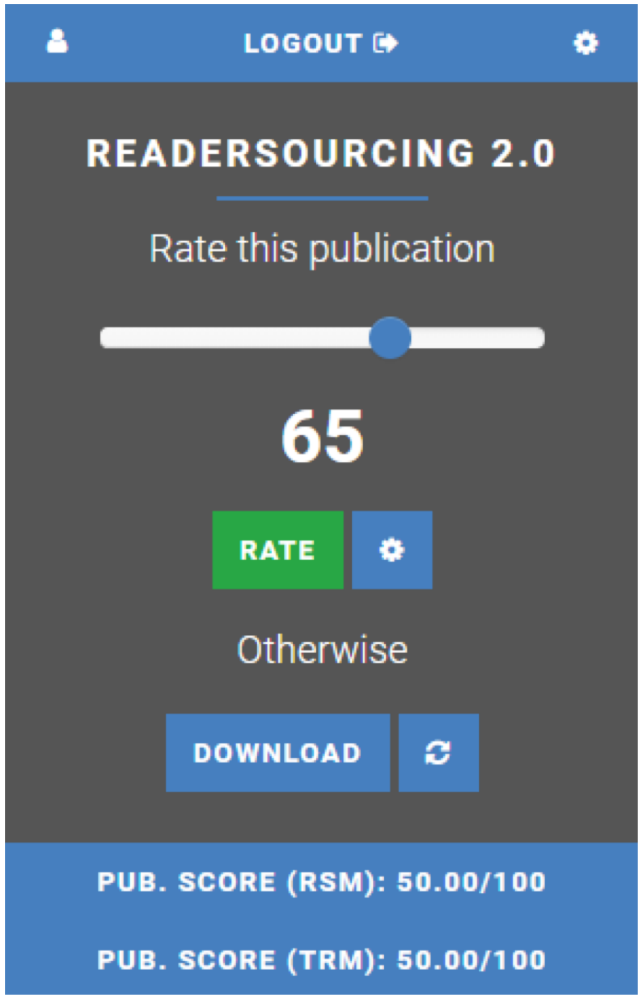
- An **independent**, third-party, **non-profit**, **academic/scientific** endeavour, aimed at **quality rating** of **scientific/scholarly literature**
- Collaboration with SISSA-Medialab

Architecture

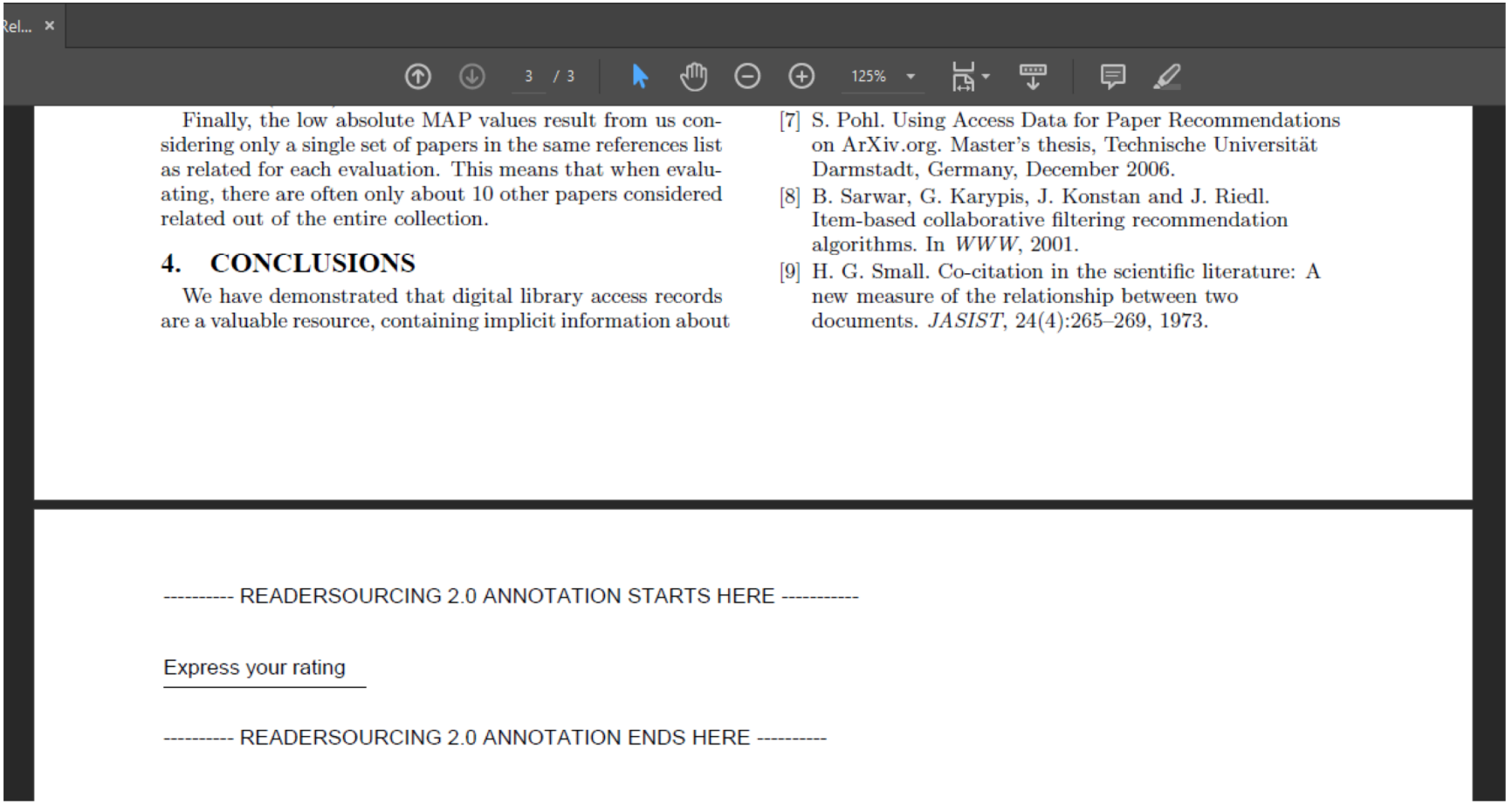


Screenshots





Screenshots



Finally, the low absolute MAP values result from us considering only a single set of papers in the same references list as related for each evaluation. This means that when evaluating, there are often only about 10 other papers considered related out of the entire collection.

4. CONCLUSIONS

We have demonstrated that digital library access records are a valuable resource, containing implicit information about

[7] S. Pohl. Using Access Data for Paper Recommendations on ArXiv.org. Master's thesis, Technische Universität Darmstadt, Germany, December 2006.

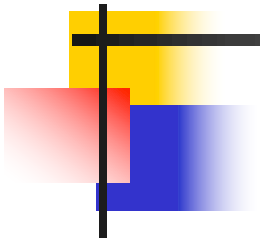
[8] B. Sarwar, G. Karypis, J. Konstan and J. Riedl. Item-based collaborative filtering recommendation algorithms. In *WWW*, 2001.

[9] H. G. Small. Co-citation in the scientific literature: A new measure of the relationship between two documents. *JASIST*, 24(4):265–269, 1973.

---------- READERSOURCING 2.0 ANNOTATION STARTS HERE ----------

Express your rating

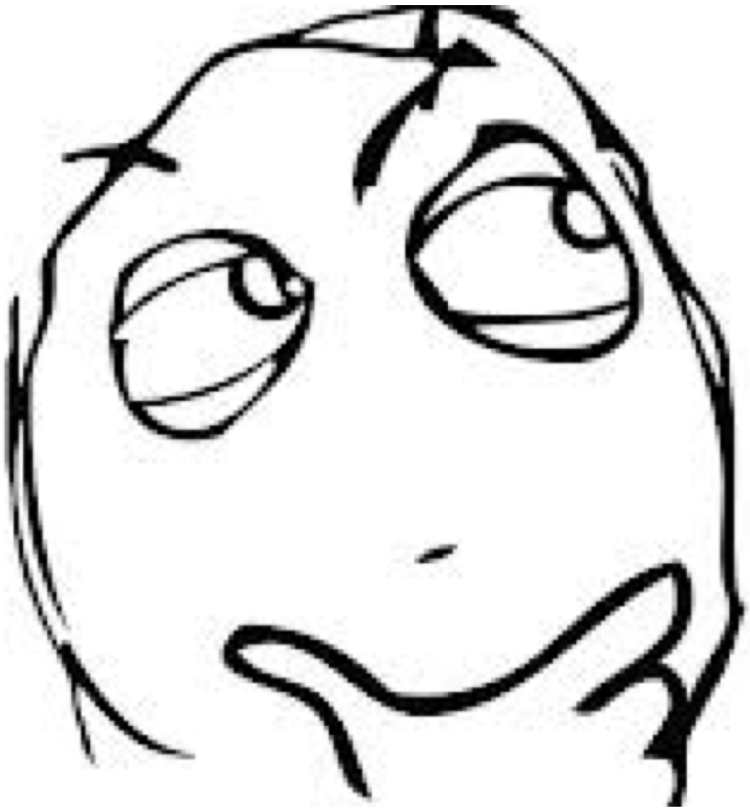
---------- READERSOURCING 2.0 ANNOTATION ENDS HERE ----------

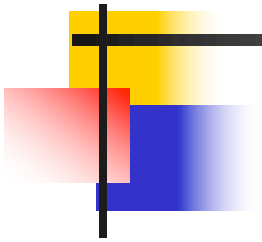


Several questions

- Is it efficient? (spoiler: yes)
- Does it converge?
- What about lazy readers?
- What about lobbies?
- Does it work??

Why are you telling me this?





Outline

1. Readersourcing: Crowdsourcing Peer Review
2. The Quality Model
3. Peer assessment

Peer review vs. assessment

- Students assess other students

- 1) Author submits paper
- 2) Peers review paper

- 1) Student "submits" the answer to an exercise
- 2) Other students evaluate answer

- It is the same model
- Plus, **The Professor** establishes the real evaluation

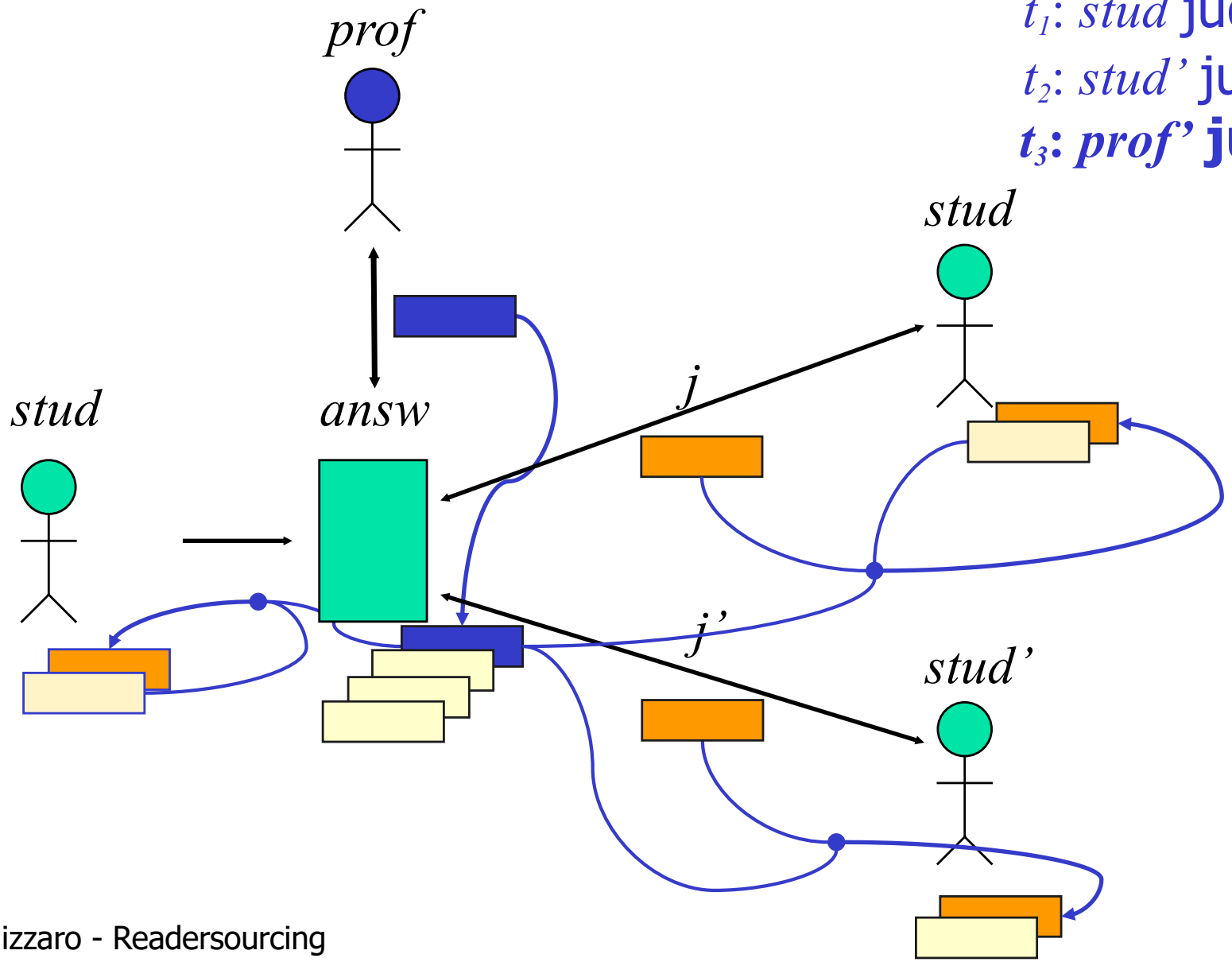
The Professor knows the answer!

t_0: *stud* publishes *answ*

t_1: *stud* judges *answ*

t_2: *stud'* judges *answ*

t_3: *prof'* judges *answ*



And in language learning…

- Native speakers!
- Many "almost professors!"
  - MWE = Multi Word Expression
  - The Grand Gurus
  - …
- Language learning looks the ideal environment

- I'm open to collaborations!
  - Busuu, ELSA, Wormingo, Tile Attack, vocabulary trainer, …
  - (EU-) projects
  - …

Summary

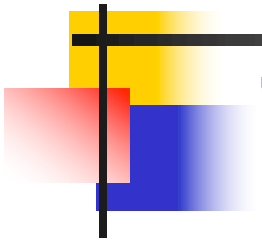
1. Scholarly publishing, Peer review
2. Readersourcing: Crowdsourcing Peer Review
3. The Quality Model
4. http://www.readersourcing.org
5. Peer assessment (Professor)
6. Language learning (Native speakers)

Conclusions – Take home message

References

- S. Mizzaro. Quality Control in Scholarly Publishing: A New Proposal, *Journal of the American Society for Information Science and Technology*, 54(11):989-1005, 2003
- A. Cusinato, V. Della Mea, F. Di Salvatore, S. Mizzaro. QuWi: Quality Control in Wikipedia. In WICOW 2009: 3rd Workshop on Information Credibility on the Web @ 18th WWW Conference
- Dialogue on a Midspring's Night Dream in Dagstuhl, http://etuttounmagnamagna.blogspot.com.au/2009/03/dialogue-on-midsprings-night-dream-in.html
- S. Mizzaro. Readersourcing – A Manifesto, *Journal of the American Society for Information Science and Technology*, 63(8):1666-1672, 2012
- …
- www.readersourcing.org

- (just ask me for a copy)



Thanks

- Paolo Coppola
- Vincenzo Della Mea
- Massimo Di Fant
- Luca Di Gaspero
- Marco Fabbrichesi
- Andrea Fusiello
- Stefania Gentili
- Stevan Harnad
- Paolo Massa
- Marco Mizzaro
- Elena Nazzi
- Carla Piazza
- Ivan Scagnetto
- Walter Vanzella
- Luca Vassena
- Paolo Zandegiacomo Riziò
- Some Referees (!?)
- …