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

Boosting Engagement of Serbian Universities in Open Science (BE-OPEN)
Study Visit
25 – 27 September 2017

Agenda
VENUE:
Palazzo Garzolini di Toppo Wassermann, Via Gemona 9
Sala del Consiglio, 1st floor

DAY 1: Monday, 25 September 2017
9.30 - 10.00 Registration of participants
10.00 - 10.15 M. Lenisa and M.C. Nicoli (University of Udine), Opening
10.15 - 11.15 T. Margoni (University of Glasgow), T.
11.15 - 11.45 Coffee break
11.45 - 12.45 E. Giglia (University of Torino), “Open Access and Open Science in Italy: Goals, Achievements, and Hindrances”
12.45 - 13.00 Discussion
13.00 - 14.30 Lunch
15.30 – 16.00 Discussion

DAY 2: Tuesday 26 September 2017
9.00 - 10.00 J. Piščanc and M. Rossi (University of Trieste), “10 years of Open Access at University of Trieste: from IR to CRIS integrated with Open Research Data”
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

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.
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

```
1 + 1 = 3
```
Reviewers should re-do the experiments to see if the results hold?
(Good) referees are scarce resource

Publication force

Now?

Reviewing force

time
But wait a minute

- We read papers!
- What do we do with that?
Answer:
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 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,...)

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...
"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 $\iff$ good paper
  - High judgments by readers $\rightarrow$ high score ($\sim$ average)

- Each **author** has a **score** too
  - It changes accordingly to the scores of the papers published by the author ($\sim$ average paper score)
  - Publishing good papers $\rightarrow$ 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 $+\infty$
\[ s_{a}(t) = \sum_{p \in P_{a}(t)} \left( \sum_{r \in R_{p}(t)} s_{r}(t_{r}, p) \cdot j_{r, p} \right) \frac{\sum_{p \in P_{a}(t)} \left( \sum_{r \in R_{p}(t)} s_{r}(t_{r}, p) \right)}{\sum_{p \in P_{a}(t)} \left( \sum_{r \in R_{p}(t)} s_{r}(t_{r}, p) \right)} \]
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)

$t_0$: $a$ publishes $p$

$t_1$: $r$ judges $p$

$t_2$: $r'$ judges $p$
A toy example (2/2)

$t_0$: $a$ publishes $p$
$t_1$: $r$ judges $p$
$t_2$: $r'$ judges $p$
Not only in theory!
- (Still in beta)
  - (Well, alpha)
    - (almost!)

An independent, third-party, non-profit, academic/scientific endeavour, aimed at quality rating of scientific/scholarly literature

Collaboration with SISSA-Medialab

Michael Soprano

www.readersourcing.org
Architecture

[Diagram of a system architecture involving different components such as RS_Server, RS_PDF, P1, RD1, RD2, RD3, RD4, and related processes.]
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

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
  - ...

Mizzaro - Readersourcing 40
Summary

1. Scholarly publishing, Peer review
2. Readersourcing: Crowdsourcing Peer Review
3. The Quality Model
5. Peer assessment (Professor)
6. Language learning (Native speakers)
Conclusions –
Take home message

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


...  

[www.readersourcing.org](http://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 Nazi
- Carla Piazza
- Ivan Scagnetto
- Walter Vanzella
- Luca Vassena
- Paolo Zandegiacomo Riziò
- Some Referees (!?)
- ...