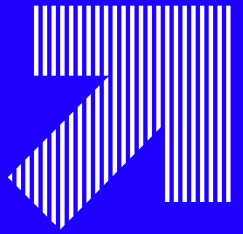
UNIVERSITÀ
DEGLI STUDI
DI UDINE

hic sunt futura



Evaluation of Crowdsourced Peer Review using Synthetic Data and Simulations

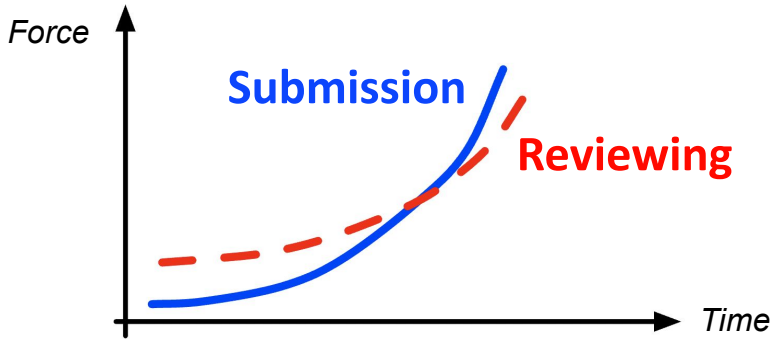
Michael Soprano, Eddy Maddalena, Francesca Da Ros, Maria Elena Zuliani, Stefano Mizzaro

21st Conference on Information and Research Science Connecting to Digital and Library Science

University of Udine, Thursday, February 20th, 2025

Scholarly Publishing and Peer Review

- **Scholarly publishing** allows for the **dissemination of knowledge**
- The process depends on **peer review**
- Researchers author **scientific articles**
- **Peers with equal expertise evaluate** their articles
- **Peer review** has been **facing challenges** for a while



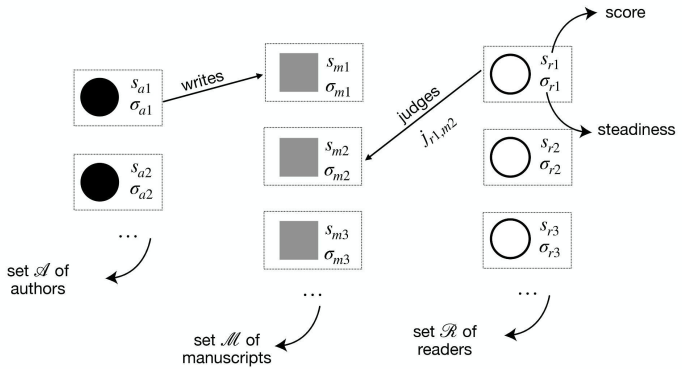
What Can Be Done?

- We have **plenty of readers**!
- **Readers read articles**, and **form opinions**
- Using **many readers instead of a few referees**
  - A form of *crowdsourcing*
- Mizzaro proposed **crowdsourcing peer review to readers**
- The **approach proposed** is called *Readersourcing*

Mizzaro, S. (2012), Journal of the American Society for Information Science and Technology. J Am Soc Inf Sci Tec, 63: 1666-1672. https://doi.org/10.1002/asi.22668

Readersourcing

- **Readersourcing** involves **three key entities**: *manuscripts*, *authors*, and *readers*
- ***Readers* assign a numerical *judgment*** to *manuscripts* written by *authors*
- **Each entity has** a **score** (s) and a **steadiness** (σ) value **based on judgment aggregation**
- The **co-determination algorithm evaluates** both **article quality** and **scholarly reputation**



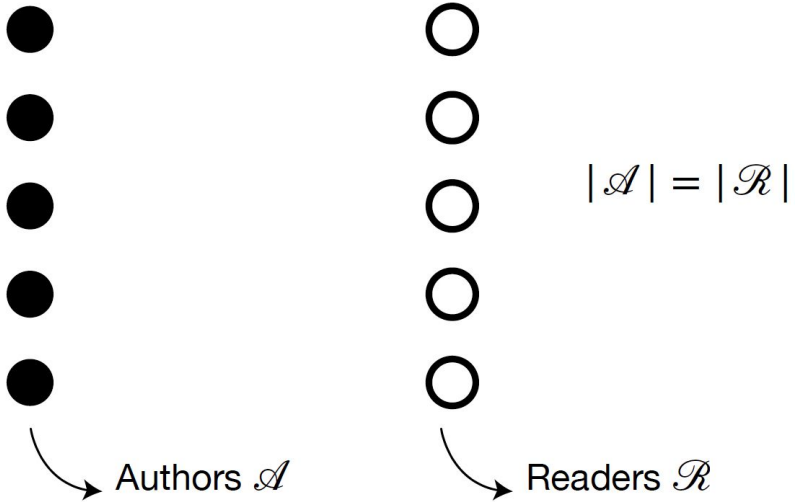
Mizzaro, S. (2003), Quality control in scholarly publishing: A new proposal. J. Am. Soc. Inf. Sci., 54: 989-1005. https://doi.org/10.1002/asi.10296

Aims

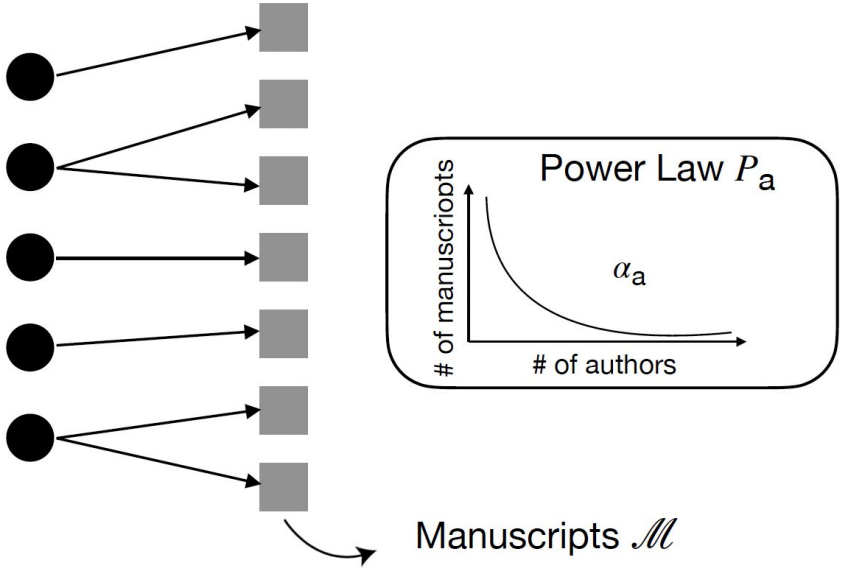
- The **effect of Readersourcing algorithm** on the **computed quantities** remains **largely unexplored**
- We evaluate Readersourcing through **simulations in a scholarly publishing scenario using a probabilistic approach**
- How effectively does **Readersourcing capture distinct aspects of judgments?**
- What is the **impact of simulation parameters on computed quantities?**

Simulation Flow

Step 1: Generate authors and readers

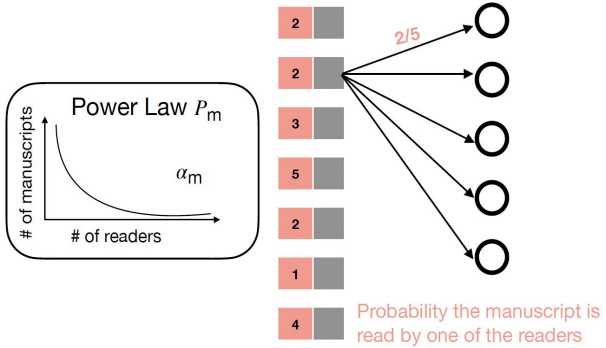


Step 2: Authors publish manuscripts following P_a

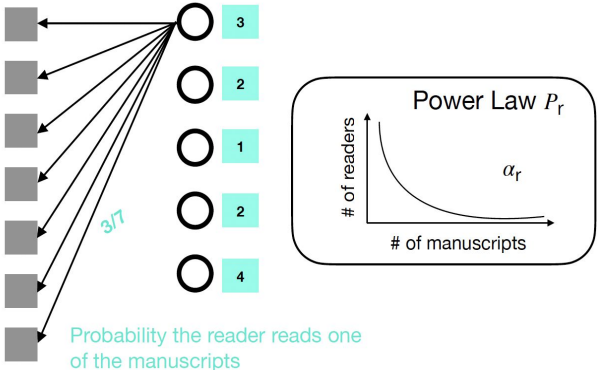


Simulation Flow

Step 3: Manuscripts are read following P_{m}

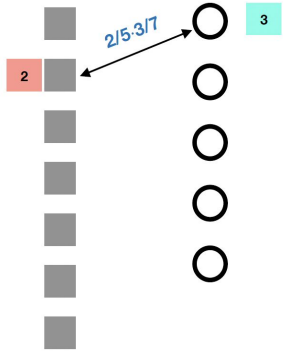


Step 4: Readers read following P_{r}



Step 5: The association between readers and manuscripts follows $P_{\text{r,m}}$

Joint probability between P_{r} and P_{m}

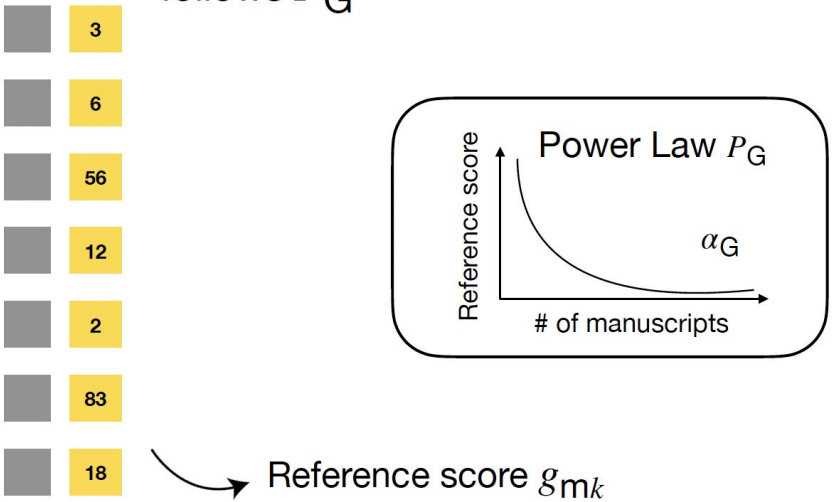


The **joint probability** is the product of the two probabilities, assuming **event independence**

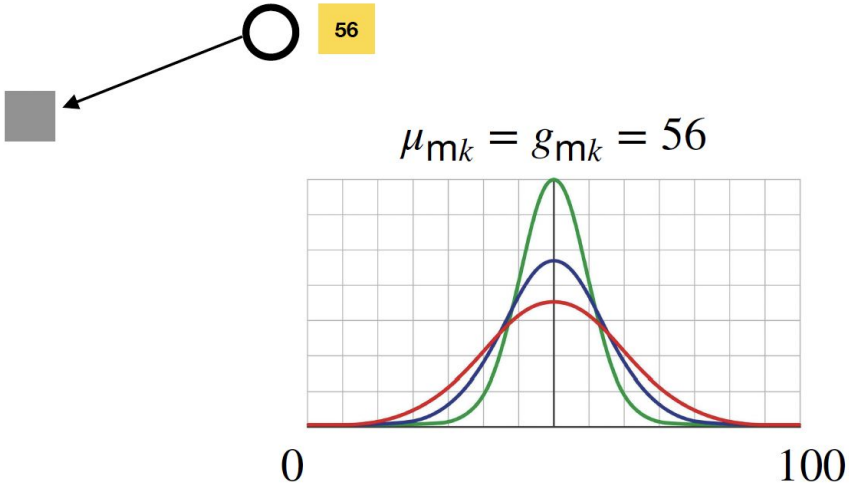
$$P(\text{manuscript } k \text{ is read by reader } l) = P(\text{reader } l \text{ reads any manuscript}) \times P(\text{manuscript } k \text{ is read by any reader})$$

Simulation Flow

Step 6: Manuscripts have a reference score that follows P_G



Step 7: Judgments follow a Gaussian distribution



Experimental Setup

- We generate a **set of simulation configurations** by empirically selecting **discrete values for six parameters**
- The **total number of configurations** is given by
  - 1 population size x 81 α values x 4 st. dev. values = 324
  - We run each configuration 10 times to account for stochasticity
- We **run around 3,000 simulations**, targeting 250 authors/readers
- We analyze **synthetic judgments** and **six computed quantities** in Readersourcing (score and steadiness for each entity)

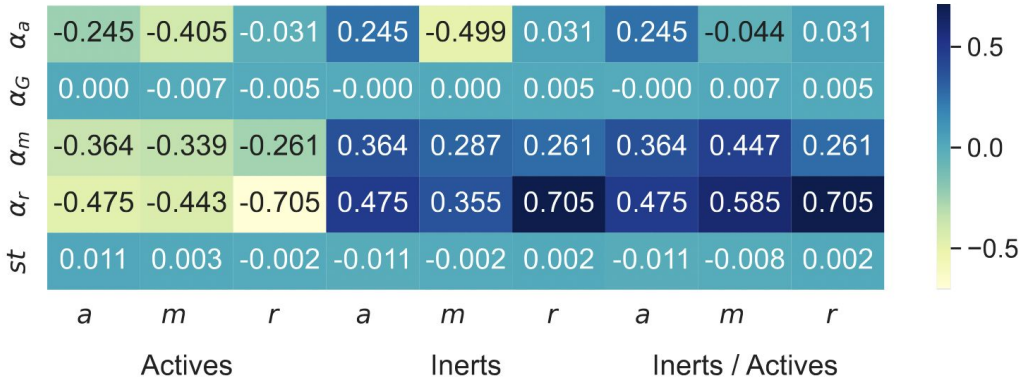
Active and Inert Entities

- The **probabilistic model may generate entities that do not "participate"**, we called them *inerts*
  - Unread manuscripts, readers who do not provide judgments, ...
- We first **quantify inerts** and then **exclude them** from the simulation flow

Statistic	Active Authors	Active Manuscripts	Active Readers
mean	214	1844	162
min	175	899	115
25%	208	1557	150
50%	216	1811	163
75%	223	2110	176
max	241	3340	206

Effect of the Probabilistic Approach

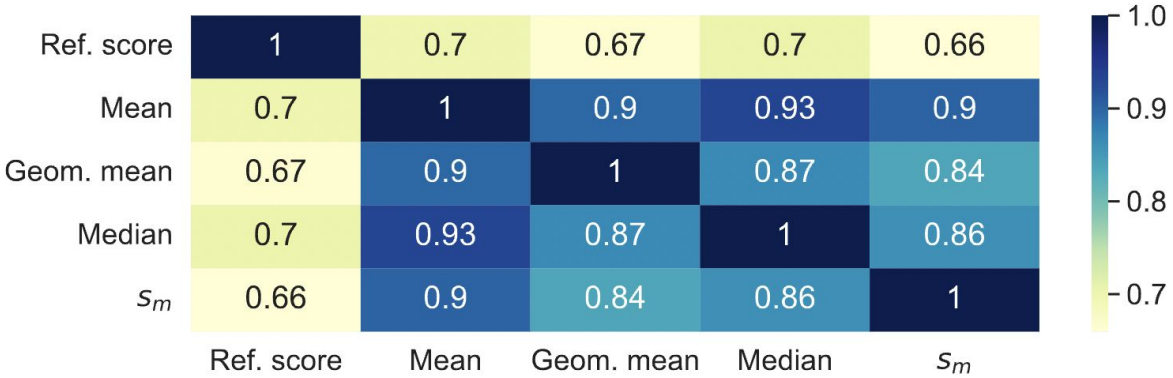
- **Some simulation parameters influence** the number of **active and inert entities**
- **Understanding these effects helps tune parameters** to minimize inerts and refine experimental designs



Kendall's τ correlation values w.r.t to the number of active and inert entities, and the inert-to-active ratio

Comparison with Simpler Aggregation Strategies

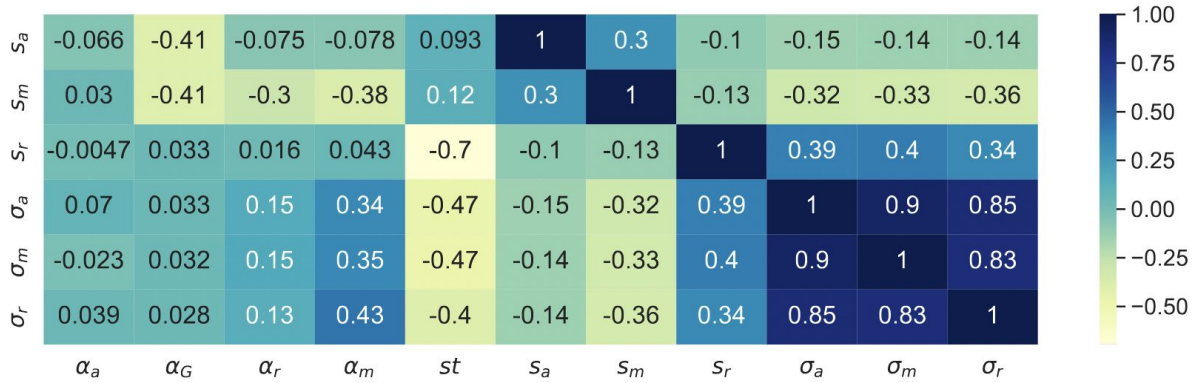
- **Weak correlation with the reference score** suggests an **alternative judgment dimension**
- **Strong correlation with aggregation strategies** reflects **overall judgment trends**
  - "Readersourcing captures a different perspective"



Kendall's τ correlation values w.r.t. the reference score, aggregated judgments, and the manuscript score

Impact of Model Components on Quantities

- Strong steadiness correlation indicates **system-wide consistency**
- Feature importance analysis links **higher manuscript scores to more consistent judgments**
- ANOVA shows **standard deviation affects score steadiness**



Kendall's τ correlation values w.r.t. simulation parameters, scores, and steadiness values

Conclusions

- We design a **simulation flow to evaluate Readersourcing**
- Readersourcing provides **insights beyond simple aggregation strategies** and shows **system-level properties**
- Future work includes **comparisons with other models** and **simulating more complex scenarios**

Acknowledgments

This research is partially supported by the **PRIN 2022 Project** – “MoT—The Measure of Truth: An Evaluation-Centered Machine-Human Hybrid Framework for Assessing Information Truthfulness” – Code No. 20227F2ZN3, CUP No. G53D23002800006 Funded by the European Union – Next Generation EU – PNRR M4 C2 I1.1