

Chapter 21

Building a Data-Driven Model of Peer Review: The Case of Science Foundation Ireland



Thomas Feliciani, Pablo Lucas, Junwen Luo, and Kalpana Shankar

Abstract Research has long questioned the validity and reliability of peer review, the process for selecting manuscripts for publication and research proposals for funding. For example, scholars have shown that reviewers do not interpret evaluation criteria in the same way [1] and produce inconsistent ratings [2], and that peer review is subject to gender, ethnicity, seniority, and reputation biases [8, 11].

Keywords Peer review · Research funding · Agent-based modeling

Modeling Peer Review

Research has long questioned the validity and reliability of peer review, the process for selecting manuscripts for publication and research proposals for funding. For example, scholars have shown that reviewers do not interpret evaluation criteria in the same way [1] and produce inconsistent ratings (Bornmann et al. [2], and that peer review is subject to gender, ethnicity, seniority, and reputation biases [9, 12].

Social simulation and agent-based models in particular have proven valuable tools to study the causes of-and seek remedies for- the issues with peer review [16]. This is due to three factors: (1) the complex nature of peer review systems, which are characterized by non-linear interdependencies between applicants, reviewers, and

T. Feliciani (✉) · P. Lucas
School of Sociology and Geary Institute for Public Policy, University College Dublin, Dublin,
Ireland
e-mail: thomas.feliciani@ucd.ie

P. Lucas
e-mail: pablo.lucas@ucd.ie

J. Luo · K. Shankar
School of Information and Communication Studies and Geary Institute for Public Policy,
University College Dublin, Dublin, Ireland
e-mail: junwen.luo@ucd.ie

K. Shankar
e-mail: kalpana.shankar@ucd.ie

the institutions in which they are embedded; (2) the typically high cost and risk of testing interventions; and (3) the notorious scarcity of available data on peer review systems [8, 15].

Since 1969, when scholars have first turned to formal and computational modeling to study peer review [17], 44 modeling papers have been published on the subject. The current state-of-the-art shows two lacunae: limited model integration and limited empirical calibration and validation [4].

This paper reports on work in progress aimed at filling in these gaps. In the context of a larger, mixed-method project on the peer review process at Science Foundation Ireland (SFI), we are integrating and building on existing simulation models of peer review to compare them and better connect them to empirical reality. In our work we focus on different aspects of peer review, one of which are aggregation rules. These rules define how the assessment by different reviewers and/or on different evaluation criteria can be combined into an aggregated score, a number which captures the overall worth of a submission. We will present our ongoing work on aggregation rules as example to illustrate the typical lacunae in simulation studies, and how a mixed-methods approach can help solve them.

Aggregation Rules in Simulation Literature

Several simulation studies have explicitly modeled aggregation rules. A complete review is provided in Feliciani et al. [4]; here we mention a few examples. In some models the aggregated score can be the median of the different review scores [11]; some other models take the mean of the scores—the mean can be weighted by the reviewers' reputation [13] or complemented with information on the standard deviation of the individual scores [10].

First Issue: Limited Model Integration

The existing literature consists of an abundance of competing assumptions, alternative implementations, and unconnected findings, which result in the fragmentation of the landscape. The literature on aggregation rules is a prime example: only a few papers have compared more than one aggregation rule (the examples above have), and no one has attempted to implement aggregation rules proposed in previous work. This lack of integration among models and further development of existing models raises important concerns about their generalizability.

The way we are addressing this issue is by implementing the aggregation rules from the literature into our simulation model. By aligning these rules within a common simulation framework, we can test them against one another and find which ones (and under what conditions) are the best at maximizing common outcome

metrics of peer review, like efficacy (i.e. the capability to filter out poor quality submissions) and efficiency (i.e. reducing costs).

Second Issue: Limited Use of Empirical Calibration and Validation

Despite many researchers advocating for more empirically calibrated and validated models [7], few have incorporated empirical evidence of peer review in their work. With a few exceptions (e.g. [5]), only a few models of peer review (or of aggregation rules specifically) have been calibrated or validated.

This may be due to different reasons. One reason could be the aforementioned scarcity of data on peer review systems. A second reason could be that, even when data are available, they often are of the qualitative kind, and there is no consensus on what are good practices in using qualitative data source to calibrate and validate models.

We argue that the use of both qualitative and quantitative evidence is necessary for the study of peer review; as such data is necessary to understand the formal rules and the actual practices of the peer review process. Further to that, quantitative data can be deployed for empirical testing of models' predictions [6].

In our study of aggregation rules, we can rely on qualitative and quantitative data on the peer review process in two funding schemes at Science Foundation Ireland. We are using these data in two ways: first, to reproduce the conditions found in a real peer review process, and second, to test the effects of competing aggregation rules against empirically observable outcomes.

The data sources we have are at different levels of aggregation (funding calls, proposals, individual applicants and reviewers) with mixed quantitative and qualitative components (e.g. call documents, instructions, textual reviews, interviews with applicants, and so on). The use of qualitative data sources in particular leads to some challenges which are of common interest for modelers who work with these data types.

Use of Qualitative Data Sources

The first challenge concerns the formalization of the model—that is, the initial phase of model building where the modeler translates an informal description of a process into a formal system. In our case, formalization means translating reviewer guidelines and SFI regulation and guidance documents into code for the agent-based model.

The literature offers few examples of protocols or methods to produce code based on some kinds of qualitative data. One example is the Engineering Agent Based Social Simulation framework (EABSS), which demonstrates how model development can

be driven by a focus group [14]. Others had interview data as starting point: interviews can be used to draw cognitive maps, and cognitive maps are implemented in the simulation environment to guide agents' behavior [3].

A second challenge concerns the use of qualitative data for empirical validation of a simulation model. To our knowledge, the only way of testing numerically-expressed model predictions against qualitative data (e.g. a report by the chair of a SFI sitting panel) is to convert the qualitative data into quantities. For textual inputs (like reviews) we can do the conversion with a combination of manual coding and computational methods (e.g. natural language processing)—this, at least, is the approach we are taking to translate textual reviews into input for the empirical calibration of a simulation model of aggregation rules.

Conclusion

By taking our ongoing study of aggregation rules in peer review as an example, we have illustrated two common gaps in the modeling literature, why it is important that we address them, and how we can do it. One of the two gaps concerns the insufficient interface between models and the real world: we have argued that the use of mixed data sources can alleviate the issue, and we have summarized our strategies for implementing diverse data sources in the empirical calibration and validation of a simulation model.

To conclude, our modeling work on the peer review process at SFI has two ambitious objectives: (1) to test competing assumptions and modeling strategies, and (2) to pioneer the integration of qualitative evidence into a simulation model, for which standards have yet to emerge.

Acknowledgements This material is based upon works supported by the Science Foundation Ireland under Grant No.17/SPR/5319.

References

1. H. Abdoul, C. Perrey, P. Amiel, F. Tubach, S. Gottot, I. Durand-Zaleski, C. Alberti, Peer review of grant applications: criteria used and qualitative study of reviewer practices. *PLoS ONE* **7**(9), e46054 (2012)
2. L. Bornmann, R. Mutz, H.-D. Daniel, A reliability-generalization study of journal peer reviews: a multilevel meta-analysis of inter-rater reliability and its determinants. *PLoS ONE* **5**(12), e14331 (2010)
3. S. Elsayah, J.H.A. Guillaume, T. Filatova, J. Rook, A.J. Jakeman, A methodology for eliciting, representing, and analysing stakeholder knowledge for decision making on complex socio-ecological systems: from cognitive maps to agent-based models. *J. Environ. Manag.* **151**, 500–516 (2015)
4. T. Feliciani, J. Luo, L. Ma, P. Lucas, F. Squazzoni, A. Marušić, K. Shankar, A scoping review of simulation models of peer review. *Scientometrics* (2019)

5. F. Grimaldo, M. Paolucci, J. Sabater-Mir, Reputation or peer review? The role of outliers. *Scientometrics* **116**, 1421 (2018)
6. S. Hassan, J. Pavón, L. Antunes, N. Gilbert, Injecting data into agent-based simulation, in *Simulating Interacting Agents and Social Phenomena*, eds. by K. Takadama, C. Cioffi-Revilla, G. Deffuant (2010), pp. 177–191
7. P. Hedström, G. Manzo, Recent trends in agent-based computational research: a brief introduction. *Sociol. Methods Res.* **44**(2), 179–185 (2015)
8. C.J. Lee, D. Moher, Promote scientific integrity via journal peer review data. *Science* **357**(6348), 256–257 (2017)
9. C.J. Lee, C.R. Sugimoto, G. Zhang, B. Cronin, Bias in peer review. *J. Am. Soc. Inform. Sci. Technol.* **64**(1), 2–17 (2013)
10. J.D. Linton, Improving the peer review process: capturing more information and enabling high-risk/high-return research. *Res. Policy* **45**(9), 1936–1938 (2016)
11. Lyon A, Morreau M, The wisdom of collective grading and the effects of epistemic and semantic diversity, in *Theory and Decision* (2017), pp. 1–18
12. H.W. Marsh, U.W. Jayasinghe, N.W. Bond, Improving the peer-review process for grant applications: reliability, validity, bias, and generalizability. *Am. Psychol.* **63**(3), 160–168 (2008)
13. S. Righi, K. Takács, The miracle of peer review and development in science: an agent-based model. *Scientometrics* **113**(1), 587–607 (2017)
14. P.-O. Siebers, F. Klügl, What software engineering has to offer to agent-based social simulation, in *Simulating Social Complexity*, eds. by B. Edmonds, R. Meyer (2017), pp. 81–117
15. F. Squazzoni, E. Brezis, A. Marušić, Scientometrics of peer review. *Scientometrics* **113**(1), 501–502 (2017)
16. F. Squazzoni, K. Takács, Social simulation that “peers into peer review. *J. Artif. Soc Soc. Simul.* **14**(4) (2011)
17. A.L. Stinchcombe, R. Ofshe, On journal editing as a probabilistic process. *Am. Sociol.* **4**(2), 116–117 (1969)