Peer review as a reputation system

Jordi Sabater-Mir[1], Francisco Grimaldo[2], and Mario Paolucci[3]

[1] Artificial Intelligence Research Institute, Spanish National Research Council
Campus UAB, Bellaterra, 08193, Spain, `jsabater@iiia.csic.es`
[2] Departament d'Informàtica, Universitat de València
Av. de la Universitat s/n, Burjassot, 46100, Spain, `francisco.grimaldo@uv.es`
[3] Institute of Cognitive Sciences and Technologies, Italian National Research Council
Via Palestro 32, Roma, 00185, Italy, `mario.paolucci@istc.cnr.it`

Abstract. This work presents our first steps towards analyzing the effects a reputation system would have over science production and evaluation as an alternative to the classical peer review process. Two agent-based simulation models have been built that place a different quality control system on the process of paper production, reading and publication. An open discussion about the stylized facts that can be used to feed and validate both models is also provided.

Keywords: Peer review, reputation, agent-based simulation

1 Introduction

From its widespread adoption in the early 1970s the peer review process has ruled the quality control system for scientific publication, thus determining the evolution of science and of the people that make it. A number of investigations have therefore analyzed the pros and cons of such an examination process [2] and have eventually find it to be the least bad. Notwithstanding this, reputation systems play nowadays a key role in rating and collecting feedback for a number of domains (e.g. sales, hotel or restaurant bookings) and might also become a good alternative to traditional peer review of manuscripts [4]. In this piece of work we start exploring such an alternative by building and comparing two agent-based simulation models: one that reproduces a simplified version of the current peer review process and another that uses a reputation-based mechanism.

The interest of this comparison is manifold. First, one could argue that reputed referees should not be substituted by non-experts but it is a fact that the growth in the number of submissions threatens the sustainability of a system that already faces difficulties obtaining the cooperation of senior researchers and involves a high percentage of junior researchers. A reputation system based on the opinions of scientists after reading state-of-the-art articles may help reduce reviewing workload, an effort that could be dedicated to the production of new science. Second, drafts need to be available in order to be read and assessed but on-line repositories are more and more used and solve any physical publication space limitation. Third, the citation and reputation systems can conceptually

overlap as both recognize quality but the first one can be biased towards the aggregate reputation of the journal and develops more slowly due to the delay introduced by the publication time. Last but not least, reputation might be stronger as a drive to decide what to read since paper level scores could better match the heterogeneous quality of manuscripts published in a journal and result in a fairer visibility of them (i.e. currently, publishing a paper attracting a low number of citations in a highly ranked journal is good for you career, when did that start?).

All the previous considerations, in our honest opinion, make reputation systems deserve an opportunity to enter the game, at least to confirm - or refute - the obvious critical point that come to mind.

2   Research questions

The core question we aim to answer is how the traditional peer review system compares to a distributed, crow-sourced evaluation. Such a distributed system was not available when peer review first came out, but it could easily put into practice today. Should we promote such a change? Or instead, does peer review have substantial advantages against crowdsourcing? We think that an answer to this research question should be obtained from a computational approach, also taking into account the historical path of science.

Another related question is about how scale (in scientific papers production terms) changes the rules of the game. Is the quality filtering mechanism we are using now based on peer review the best mechanism given the amount of scientific publications that are generated and submitted every year? We hypothesize that having a reputation mechanism as the mechanism responsible for deciding about the quality of papers in an environment where the scientific production is very high has important advantages if compared with the traditional mechanism based on peer review. In a system based on reputation, the reputation of a paper becomes a measure of the scientific value of that paper for that community. This value has been assigned by those members of the community that have read the paper and decided to provide an evaluation. Following our hypothesis, the advantages of using reputation in a scientific environment with high production rates could be:

- Speed of publication. While in a classical peer review environment a paper can be delayed for its publication even more than one year, using reputation the publication would be immediate.
- The wisdom of the crowd. The paper is evaluated by many researchers, presumably with a high knowledge of the scientific area, and not only by three or four experts.

Furthermore, other important and collateral questions arise:

- In a peer review based system, the reviews are supervised and therefore they are assumed to have a minimum quality. This is not the case for the

evaluations in a reputation based system that can range from superficial evaluations to deep reviews.

- Although it is true that the publication of a paper is immediate in the case of reputation, the associated reputation value takes some time to become reliable.
- A reputation mechanism is more prone to manipulation than a supervised peer review mechanism.

3 The model

To ground the discussion and provide computational answers to our research questions, we have built a simple model of paper production. On the top of paper production, we apply either a simulated peer review process, or a simulated reputation mechanism. Both of those mechanisms drive the reading process. Scientists, in one case, read papers as prioritized by journal's quality. In the other case, they read papers as prioritized by their reputation.

Following the approach used by other paper production models [3], we consider as main entities: the scientist (in the role of author and reader) and the paper (possibly co-authored and available to be read). Scientists and papers are generated with a random initial *quality*. In our NetLogo implementation, we use random generators of positive quality following either a *zipf* or *exponential* distribution. Scientists are also characterised by their *productivity*, that is, the number of papers that they produce in a simulation step. The same value drives also the number of papers read by a scientist. Being capable of reading only a limited amount of papers, scientist must choose. The choice is driven by different mechanisms in the two submodels detailed below.

The main loop of creation/publishing/reading can be summed up in the following steps

1. co-authorship creation
2. paper generation (with quality update for authors)
3. paper review (only for peer review)
4. paper reading (and evaluation, in the reputation case).

Paper production is simulated as the main activity of scientists. In our model, a group of scientists is put together to produce a paper, to simulate co-authorship. We implemented both a completely *random* co-authorship and a co-author extraction routine that privileges previous co-authors. Once the paper is produced, its quality is calculated as the average between the quality of its authors and the quality of the "idea" presented in that paper that is randomly generated (zipf or exponential). Once the quality of the paper is attributed, the authors' quality is modified as if attracted towards the paper's value. In other words, working with better scientists or having a great idea improves a scientist' quality. The opposite is also true.

In the peer review scenario we introduce the *journal* entity. A journal has a target quality against which submissions are evaluated. The process of evaluation

is performed by a set of scientists as peer reviewers. The reviewers try to ascertain if the submitted paper has a quality that satisfies the journal's request, with a noise that is inversely proportional to their own quality. Thus, produced papers are randomly submitted to journals but only papers that receive a majority of accept recommendations from the assigned reviewers will reach the published state. In this scenario, scientists use the quality of the journal in a weighted lottery over published and unread papers to extract the next article to read.

In the reputation scenario papers do not need to get through the gates of peer review. Every paper written is immediately accessible for reading. Scientists use the paper's reputation values in a weighted lottery over all unread papers to extract the next article to read. The reputation is initialized as the mean reputation of the papers' authors, but gradually substituted by the average of evaluations as the manuscript is read. When reading a paper, scientists evaluate its quality with a noise inversely proportional to their own quality. The evaluation of a paper after reading is similar to the one used in peer review, but, as we assume papers get read with less attention, with a substantially larger noise.

4   Discussion

We have proposed and described a model of paper publication to be studied under the effects of two different mechanism, peer review and reputation.

We plan to validate the results of the model by reference to stylized facts [1], as Lotka's law and others. We are going to use stylized facts both as inputs and outputs to the simulations. On the one hand, they will inform parameters choices and distribution choices as input to the system. On the other hand, they will also be used to compare the measures that we aim to obtain as results from the simulation, for example the number of reads per paper.

Creating a model is a matter of choice between what is included and what is left out. As we went through the process of model design, we did several of these choices, comparing also with existing models. Validating the choice of mechanisms, however, is still an open issue in the simulation community, and we are looking for advice both on the specific issue of paper creation models, and on the methodology used for such a choice.

Acknowledgements

References

1. M. Meyer. Bibliometrics, Stylized Facts and the Way Ahead: How to Build Good Social Simulation Models of Science? *Journal of Artificial Societies and Social Simulation*, 14(4), 2011.

2. M. Paolucci and F. Grimaldo. Mechanism change in a simulation of peer review: from junk support to elitism. *Scientometrics*, 99(3):663–688, 2014.
3. N. Payette. Agent-Based Models of Science. In A. Scharnhorst, K. Börner, and P. van den Besselaar, editors, *Models of Science Dynamics*, Understanding Complex Systems, pages 127–157. Springer Berlin Heidelberg, 2012.
4. I. Pinyol and J. Sabater-Mir. Computational trust and reputation models for open multi-agent systems: a review. *Artificial Intelligence Review*, 40(1):1–25, 2013.