Open Peer Review Model

Model documentation

Federico Bianchi & Flaminio Squazzoni

Building on a previous model by Squazzoni & Gandelli (2013), we simulated an idealized scientific community of N scientists, who alternatively performed the roles of authors and reviewers of manuscripts for scholarly journals. When authors, scientists submit manuscripts for publication, while when reviewers, they evaluate a randomly assigned submitted manuscript. At each iteration of the model, half of the scientists are randomly assigned as authors, half as reviewers.

For the sake of simplicity, we assumed that each author submits only one manuscript at each iteration step and that each manuscript is reviewed by one reviewer, who is required to assign a score to the manuscript. At the end of each iteration, a fixed proportion P of submissions are selected for publication based on the best scores assigned by reviewers.

We assumed that both authoring and reviewing are costly activities. Each scientist is provided with a variable amount of resources ri, which they consume to perform their tasks. Each scientist is provided with a small, fixed amount of resources (E) at the beginning of each iteration. Furthermore, at the end of each iteration, scientists are assigned extra-resources in case their submitted manuscripts are eventually selected for publication. This was to model a productivity-based resource allocation policy, mitigated by access to minimal resources to which all scientists are entitled independent of their performance, such as research infrastructure provided by their organizations, access to libraries, online repositories, laboratories, etc.

We assumed that each submitted manuscript has an intrinsic quality value. This depends on each author's *expected quality* ($\bar{q_i}$), which is calculated at the beginning of each iteration as a function of the author's resources, as follows:

$$\bar{q_i} = \frac{vr_i}{vr_i + 1}$$

where v is a fixed rate at which authors' expected quality varies according to the increase of author's resources. Then, the actual *submission quality* (q_m) of a manuscript is randomly drawn from a normal distribution $N(\bar{q_i}, \sigma = B\bar{q_i})$, with B as the author bias factor, defined as an exogenously fixed parameter.

This was to model possible variability of quality across different submissions by the same author, due to idiosyncratic factors. Moreover, we assumed that the submission quality also determines the amount of resources consumed to submit a manuscript: submitting better manuscripts requires authors to invest more resources.

When reviewers, scientists estimate the quality of their assigned manuscript. In order to model reviewing bias, the estimated quality ($s_{i,m}$) of a manuscript m by reviewer i depends on the distance between i's expected quality as author and the submission quality of the manuscript, as in the following equation:

$$s_{i,m} = \frac{r_i(1 + q_m - \bar{q_i})}{2}$$

Equation 2 reflects the assumption that the cost of reviewing increases with the distance between a reviewer's skills and the ideal skills needed to properly review the manuscript due to its quality. Similar to how we model authors' resource consumption, we assumed that the amount of resources consumed for reviewing are also determined by the quality of the review, hence by $s_{i,m}$. Therefore, if reviewers are assigned a manuscript whose quality is close to that of their own submissions, they consume 50% of their resources to deliver their review. As a result, fewer resources are consumed if they are assigned manuscripts of lower quality, while more resources are consumed if they are assigned manuscripts of higher quality. Moreover, reviewing expenses also depend proportionally on the reviewer's level of resources. This was to mirror the fact that top scientists usually require less time and effort for reviewing, as they are more experienced and skilled than average scientists. However, we linked effort and value as a proportion as when reviewing similar manuscripts, more productive scientists consumed more resources than average scientists because their time was more costly.

In order to keep the model simple and focus on the scientists' behavior, we did not explicitly model a publishing landscape with different journals having different standards and selectivity (see Kovanis et al. 2016). We simply assumed a selective scholarly journal market. In our model, authors of unpublished manuscripts lose all resources invested in the submission process. On the contrary, if published, author resources are multiplied by a value *M*, ranging between 1 (for most resourceful scientists), to 1.5 (for least resourceful scientists). This was to model different marginal benefits yielded from resource gains for scientists depending on their level of resources (ideally reflecting the number of their previous publications). The assumption was that the marginal utility of each publication is higher at the beginning of a scientist's career than in later stages.

More precisely, at the end of each iteration, all n published authors are ranked by level of resources according to an A_i sequence, where A_0 represents the published author with the lowest level of resources and A_{n-1} represents the published author with the highest level of resources. Defining g as the highest multiplier value possible, we calculated the multiplier M for scientist i as follows:

$$M(A_i) = g - \frac{i}{n-1}(g-1) = \frac{g(n-i-1)+i}{n-1}$$