



## Regular article

## Hidden connections: Network effects on editorial decisions in four computer science journals



Giangiaco Bravo<sup>a,b,\*</sup>, Mike Farjam<sup>a,b</sup>, Francisco Grimaldo Moreno<sup>c</sup>,  
Aliaksandr Birukou<sup>e</sup>, Flaminio Squazzoni<sup>d</sup>

<sup>a</sup> Department of Social Studies, Linnaeus University, Växjö, Sweden

<sup>b</sup> Linnaeus University Centre for Data Intensive Sciences & Applications, Växjö, Sweden

<sup>c</sup> Department of Informatics, University of Valencia, Valencia, Spain

<sup>d</sup> Department of Economics and Management, University of Brescia, Brescia, Italy

<sup>e</sup> Springer Nature, Heidelberg, Germany

## ARTICLE INFO

## Article history:

Received 8 September 2017

Received in revised form

13 November 2017

Accepted 2 December 2017

## Keywords:

Editorial bias

Network effects

Author reputation

Peer review

Bayesian network

## ABSTRACT

This paper aims to examine the influence of authors' reputation on editorial bias in scholarly journals. By looking at eight years of editorial decisions in four computer science journals, including 7179 observations on 2913 submissions, we reconstructed author/referee-submission networks. For each submission, we looked at reviewer scores and estimated the reputation of submission authors by means of their network degree. By training a Bayesian network, we estimated the potential effect of scientist reputation on editorial decisions. Results showed that more reputed authors were less likely to be rejected by editors when they submitted papers receiving negative reviews. Although these four journals were comparable for scope and areas, we found certain journal specificities in their editorial process. Our findings suggest ways to examine the editorial process in relatively similar journals without recurring to in-depth individual data, which are rarely available from scholarly journals.

© 2017 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

## 1. Introduction

Peer review is a decentralised, distributed collaboration process through which experts scrutinise the quality, rigour and novelty of research findings submitted by peers before publication. The interaction between all figures involved, including editors, referees and authors, helps to filter out work that is poorly done or unimportant. At the same time, by stimulating a constructive dialogue between experts, this process contributes to improve research (e.g., Casnici, Grimaldo, Gilbert, & Squazzoni, 2016; Righi & Takács, 2017). This is crucial for journals, science and knowledge development, but also for academic institutions, as scientist reputation and career are largely determined by journal publications (Bornmann & Williams, 2017; Fyfe, 2015; Squazzoni & Gandelli, 2012, 2013).

Under the imperatives of the dominant “publish or perish” academic culture, journal editors are called everyday to make delicate decisions about manuscripts, which not only affect the perception of quality of their journals but also contribute to set research standards (Petersen, Hattke, & Vogel, 2017; Siler, Lee, & Beroc, 2015) by promoting certain discoveries and

\* Corresponding author at: Department of Social Studies, Linnaeus University, Växjö, Sweden.

E-mail addresses: [giangiaco.bravo@lnu.se](mailto:giangiaco.bravo@lnu.se) (G. Bravo), [mike.farjam@lnu.se](mailto:mike.farjam@lnu.se) (M. Farjam), [francisco.grimaldo@uv.es](mailto:francisco.grimaldo@uv.es) (F. Grimaldo Moreno), [aliaksandr.birukou@springer.com](mailto:aliaksandr.birukou@springer.com) (A. Birukou), [flaminio.squazzoni@unibs.it](mailto:flaminio.squazzoni@unibs.it) (F. Squazzoni).

**Table 1**

Review round distribution by journal in the original database.

Review round	Journal			
	J1	J2	J3	J4
1	4025	3233	3534	1098
2	1117	675	563	194
3	144	121	112	21
4	6	11	11	3
5	1	0	0	0
6	1	0	0	0

methods while rejecting others (Bornmann & Daniel, 2009; Lin, Hou, & Wu, 2016; Resnik & Elmore, 2016). With the help of referees, they have to make the whole editorial process as effective as possible without falling into the trap of cognitive, institutional or subjective biases, mostly hidden and even implicit (Birukou et al., 2011; Casnici, Grimaldo, Gilbert, Dondio, & Squazzoni, 2017; Lee, Sugimoto, Zhang, & Cronin, 2013; Teele & Thelen, 2017).

Unfortunately, although largely debated and always under the spotlight, peer review and the editorial process have been rarely examined empirically and quantitatively with in-depth, across-journal data (Batagelj, Ferligoj, & Squazzoni, 2017; Bornmann, 2011). While editorial bias has been investigated in specific contexts (e.g., Hsiehchen & Espinoza, 2016; Moustafa, 2015), the role of bias due to hidden connections between authors, referees and editors, which are determined by their reputation and position in the community network, has been examined rarely empirically (García, Rodríguez-Sánchez, & Fdez-Valdivia, 2015; Grimaldo & Paolucci, 2013; Squazzoni & Gandelli, 2013). A noteworthy exception has been Sarigöl, García, Scholtes, and Schweitzer (2017) who recently examined more than 100,000 articles published in PLoS ONE between 2007 and 2015 to understand whether co-authorship relations between authors and the handling editor affected the manuscript handling time. Their results showed that editors handled submissions co-authored by previous collaborators significantly more often than expected at random, and that such prior co-author relations were significantly related to faster manuscript handling. In these cases, editorial decisions were sped up on average by 19 days. However, this analysis could not look at whole editorial process, including rejections and referee selection, and could not disentangle editorial bias from submission authors' strategies of editors' targeting.

Our study aims to fill this gap by presenting a comprehensive analysis of eight years of the editorial process in four computer science journals. First, these journals were comparable in terms of scope and thematic areas, so providing an interesting picture of a community and its network structure. Secondly, we looked at the whole editorial process, with a particular focus on editorial decisions and referee recommendations from all submissions. While we did not have data on all characteristics of authors and submissions and so could not develop intrinsic estimates of a manuscript's quality, we used network data to trace the potential effect of authors' centrality on editorial decisions once checked for the effect of review scores. In this respect, our study offers a method to examine editorial bias without in-depth and complete data, which are rarely available from journals (Squazzoni, Grimaldo, & Marusic, 2017). More substantively, our analysis revealed that although these four journals were in the same field, their editorial processes were journal-specific, e.g., influenced by rejection rates and impact factor. Secondly and more interestingly, our findings show that more reputed authors were less penalised by editors when they did not presumably submit brilliant work.

The rest of the paper is organised as follows. Section 2 presents our dataset, including data anonymisation procedures and the construction of our main variables. Section 3 presents our findings, including a Bayesian network model that allowed us to disentangle bias throughout the editorial process, while Section 4 discusses the main limitations of our study and suggests future developments.

## 2. Dataset construction

### 2.1. Data extraction and preparation

Data were acquired by following a protocol developed by a network of scientists and publishers as part of the TD1306 COST Action *New frontiers of peer review* (hereinafter, PEERE), which allowed us to share data on peer review by protecting the interests of all stakeholders involved (Squazzoni et al., 2017).

Our data included 14,870 observations encompassing several years of editorial actions in four computer science journals (hereinafter J1–J4). Each observation corresponded to one action performed by the editor, such as asking a referee to review a paper, receiving a report, or taking an editorial decision over the paper. The time frame chosen depended on the data availability for each journal. More specifically, we used 10 years of observations for J1, 12 for J2, 7 for J3 (a more recently established outlet), and 9 years for J4. For all journals, observations were limited to January 2016. About 80% of the collected data referred to the first round of reviews (Table 1).

The target journals revolved around coherent thematic areas, though they varied in their ranking of impact. Notably, J1 was ranked in the first quartile of the 2016 Journal Citation Reports (JCR) impact factor (IF) distribution, J2 was included in the fourth quartile of the IF distribution, J3 in the third, while J4 was not indexed in the JCR. For each journal, we extracted

data on manuscripts submitted, reviewer recommendations and editorial decisions. Data were cleaned and anonymised, with scientist names and submission titles replaced by secure hash identifiers (IDs) of the SHA-256 type, which digested the original strings removing accents and non alphanumeric characters (Schneier, 1996). These automatically generated IDs allowed us to track all the entities while preserving their privacy. Note that these hash codes prevented us to disambiguate author or referee names but also that namesakes were eventually assigned the same ID whereas different spellings of the same name would end up in different IDs. Not that this only introduced a marginal distortion in our dataset. First, such homonyms were rather unlikely given that we considered a limited numbers of journals. In addition, as these journals use the same journal management system, scientists had a unique profile and only one spelling for each name was available.

We restricted our attention to submissions which had at least one active referee ID in the dataset, meaning that at least one referee was invited by the editor (i.e., the paper was not desk-rejected or retracted before review). This led to 11,516 observations on 3018 submissions. However, 4250 of these cases did not have any referee recommendations because the referee answered negatively to the editor's review request, did not send a report or the recommendation was sent later than January 2016. These observations were not considered in the analysis.

Furthermore, given the purpose of our work, we considered only observations with an unequivocal editorial decision based at least one referee recommendation. We eliminated 61 submissions for which, although the ID of one or more referee existed (and could hence be considered, for instance, in the network analysis below), no decision was recorded, along with 44 submissions with non-standard decisions (e.g., "Terminated by Editor-in-Chief" or "Skipped"). This led to a final dataset including 7179 observations on 2913 submissions. These included one review report and recommendation per observation, along with one clear editorial decision per submission.

## 2.2. Referee recommendations

Referee recommendations (which sometimes appeared as non-standard expressions in the database) were first recoded into the standard ordinal scale *accept*, *minor revisions*, *major revisions*, *reject*. In order to use the referee recommendations efficiently and test their effect on editorial decisions, we estimated a numerical score for any actual set of referee recommendations. Being this an ordinal- and not an interval-scale variable, simply computing the sum of ranks of the different recommendations was not correct. We instead decided to derive the score starting from the review distribution that we would expect if we had no priors on how common each of the four recommendations were (i.e., we assumed that they were equiprobable). In practice, we derived the set of all possible recommendations for a given number of reviews and simply counted how many were clearly better or worse than the one actually received. For instance, in case there was only one referee report and the recommendation was *accept*, three possible less favourable and no one more favourable cases existed. When the recommendation was *major revision*, we assumed there were one worse and two better cases.

More generally, we used the following procedure to calculate the review scores. We first derived the set of all possible unique combinations of recommendations for each submission (henceforth, potential recommendation set). Using this set, we counted the number of combinations that were clearly less favourable (*#worse*) or more favourable (*#better*) than the one actually received by the submission (e.g., {*accept*, *accept*} was clearly better than {*reject*, *reject*}). Note that a third group of combinations existed, which could be firmly considered either better or worse than the target (e.g., {*major revisions*, *major revisions*} was neither clearly better nor worse than {*accept*, *reject*}).

This allowed us to assign an "optimistic" estimate of the value of any actual combination of recommendations, considering the whole set of possible recommendations, or a "pessimistic" one, when considering only "clear" cases. The two resulting review scores were computed using Eqs. (1) and (2) respectively.

$$reviewScore_{optimistic} = \frac{\#worse}{\#better + \#worse} \quad (1)$$

$$reviewScore_{pessimistic} = \frac{\#worse}{\#better + \#worse + \#unclear} \quad (2)$$

Table 2 shows how the review score estimation works in case of two recommendations, while similar tables can be produced for any number of recommendations. One of the most interesting aspects of this procedure is that the resulting score is always bounded in the [0, 1] interval, which makes easy to compare papers having a different number of referees.

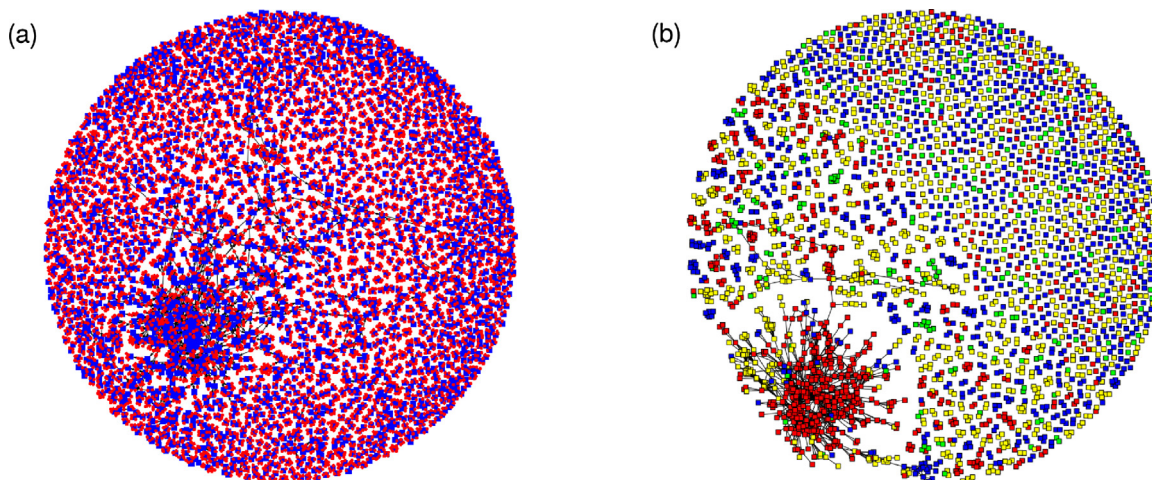
It is worth noting that the ranks produced by our "optimistic" and "pessimistic" scores do not change in Table 2, nor do they for a different number of referee recommendations (note that, given (1) and (2),  $reviewScore_{pessimistic} \leq reviewScore_{optimistic}$  as long as  $\#unclear \geq 0$ ). Furthermore, given that in all our analyses the two scores led to qualitatively similar estimates, from now on we reported only results obtained by the "optimistic" score only, called *review score*.

Finally, we estimated a referee *disagreement score* as the number of referee recommendations that should be changed to reach a perfect agreement among the referees, divided by the number of referees so that comparability across papers with a different number of reviews could be achieved. For instance, in case of three reviews recommendations such as {*accept*, *accept*, *minor revisions*}, the disagreement score would be 1/3; in case of {*accept*, *major revisions*, *minor revisions*}, it would be 2/3.

**Table 2**

Review score estimation for all possible combinations for a two-recommendation set.

Recommendations	Potential recommendation set			Review score	
	#better	#worse	#unclear	Pessimistic	Optimistic
{accept, accept}	0	9	0	1.00	1.00
{accept, minor revisions}	1	8	0	0.89	0.89
{accept, major revisions}	2	6	1	0.67	0.75
{minor revisions, minor revisions}	2	5	2	0.56	0.71
{minor revisions, major revisions}	4	4	1	0.44	0.50
{accept, reject}	3	3	3	0.33	0.50
{major revisions, major revisions}	5	2	2	0.22	0.29
{minor revisions, reject}	6	2	1	0.22	0.25
{major revisions, reject}	8	1	0	0.11	0.11
{reject, reject}	9	0	0	0.00	0.00



**Fig. 1.** Author–paper network. (a) Complete affiliation network. Papers are drawn as blue squares, authors as red circles. (b) Bipartite projection on papers. Colours indicate journals in which papers were submitted. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of the article.)

### 2.3. Network centrality measures

We looked at submissions as the target of our network analysis as these could be considered as “objects” that linked all the figures involved although in different roles, such as authors, referees and editors. We wanted to establish the position of each paper in the authors and referees’ networks. As both the author(s) and the referee(s) of a given paper had a direct relation with it, a tripartite *affiliation network*<sup>1</sup> exists in principle, where both authors and referees are “affiliated” (i.e., hold links) with the paper they wrote or reviewed. Nevertheless, a single researcher could act both as author and referee over time. For sake of simplicity, we considered two separate bipartite networks, the one including submissions and authors and the one including submissions and referees, and separately derive centrality measures for each of them. Both networks were derived from the dataset following the procedures below.

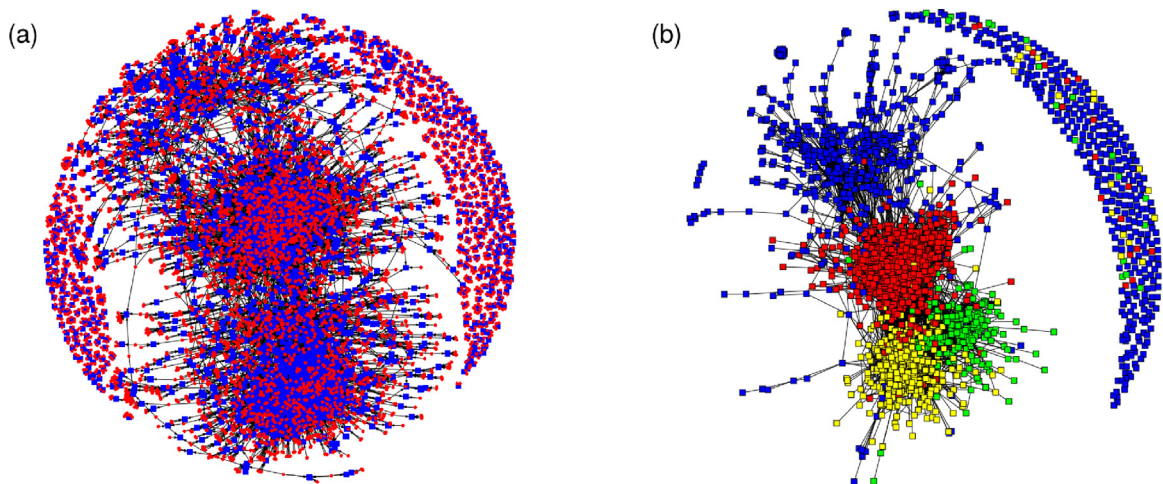
#### 2.3.1. Author–paper affiliation network

Authors and papers included in the dataset form a network where authors hold a directed link to the paper(s) they wrote. This resulted in a bipartite network with 10,049 nodes (7031 authors and 3018 papers) and 9275 links (Fig. 1a). Two-thirds of the papers in the network had three or less authors and 10.7% only one author. In addition, only 18% of the authors submitted more than one paper to the journals included in the database, which resulted in a very low network density (0.00009).

The bipartite projection of the network on papers (also known as co-membership network) considered papers as linked if they shared at least one author. It also had low density (0.00009), with over 40% of papers which were isolated. On the other hand, we found a large cluster of well connected papers, most of them submitted to two journals (red and yellow in Fig. 1b). We estimated the position of the paper in the network by using standard centrality measures, such as the degree and eigenvector centrality, the latter often used in the analysis of affiliation networks (Faust, 1997). The degree of a given

<sup>1</sup> This should not be confused with the academic affiliation of authors and referees. Following the standard definition used in network analysis, here authors and referees are “affiliated” to the paper they respectively wrote or reviewed.





**Fig. 2.** Referee–paper network. (a) Complete affiliation network. Papers are drawn as blue squares, authors as red circles. (b) Bipartite projection on papers. Colours indicate journals in which papers were submitted. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of the article.)

node simply reports the number of links it holds, while the eigenvector centrality is a more complex measure taking into account not only the position of the node in the network but also the centrality of the other nodes to which it is connected.<sup>2</sup> The corresponding statistics were computed and saved for all nodes in the paper projection, i.e., for each paper included in the dataset.

### 2.3.2. Referee–paper affiliation network

Similarly, we represented referee and papers as a network where referees held a directed link to the paper they reviewed. After applying this rule to the dataset, we found a bipartite network with 8546 nodes (5528 referees and 3018 papers) and 11,516 links (Fig. 2a). Only 11.9% papers in the network were reviewed by only one referee, and 52.7% by three or more. Furthermore, 33% of the referees reviewed more than one paper. Here, the network density was higher than in the author–paper network case (0.0002).

In this case, the projection of the network on papers considered two papers linked when they were reviewed by at least a common referee. Results showed that this network had a higher density than the original bipartite network (0.08), with only 8.6% of the papers that were isolated. Most of them were actually included in a very large component, even if the four separate journal cluster were still identifiable (Fig. 2b). As before, referee centrality measures were computed and saved for all nodes in the paper projection.

## 3. Results

### 3.1. Descriptives

The final dataset aggregating first round reviews and editorial decisions for each submission included 2957 observations. Papers were submitted to four journals: 29.2% to the first one (J1), 33.0% to J2, 30.3% to J3, and 7.5% to J4. Table 3 shows the corresponding frequency distribution of referee recommendations (a) and editorial decisions (b).

Results indicate that editorial decisions and the combined scores of referee recommendations were aligned (Table 4). It is also worth noting that the disagreement among referees was lower when the submission was accepted by the editor, whereas it significantly increased in case of major revisions ( $t = 3.10$ ,  $p = 0.004$ , differences between “accept” and the other groups were not significant). Consistently with Casnici et al. (2017), reports linked to rejection and major revision decisions were predominantly longer than in case of acceptance and minor revisions, so indicating the willingness of referees to justify their opinion with details or help authors improving their work.

The two networks presented a similar underlying structure although with significant density differences. Degrees in both the author–paper and referee–paper networks were exponentially distributed with coefficients  $< 1$ , meaning that a majority of papers had no or just a few links, while a small number of papers had many more links (up to 55 for the author–paper and to 183 for the referee–paper network). The fact that the density of the network was higher for the referee–paper network did

<sup>2</sup> In a preliminary stage of the research, we also computed the closeness and betweenness centrality measures for each node. A more comprehensive index of centrality was then derived from these correlated measures through principal component analysis. However, this index was neither more informative nor less skewed than the much simpler degree number. We therefore decided to use the latter in our analysis.

**Table 3**

Frequency distribution (%) of referee recommendations (a) and editorial decisions (b) by journal.

(a) Referee recommendation	Journal			
	J1	J2	J3	J4
Accept	4.4	2.4	6.1	10.4
Minor revisions	23.1	18.4	26.8	29.9
Major revisions	39.8	32.8	30.1	32.0
Reject	32.7	46.4	37.0	27.8

(b) Editorial decision	Journal			
	J1	J2	J3	J4
Accept	1.2	0.1	1.2	3.6
Minor revisions	10.4	7.2	18.4	27.7
Major revisions	38.5	26.7	30.2	37.3
Reject	49.9	66.1	50.1	31.4

**Table 4**

Average review scores (not considering uncertain cases), referee disagreement, referee report length and review time by editorial decision.

Editorial decision	Review score	Disagreement	Report length (characters)	Review time (days)
Accept	0.89	0.19	1729	34
Minor revisions	0.71	0.24	2602	36
Major revisions	0.39	0.32	3636	45
Reject	0.09	0.21	3358	36

**Table 5**Correlations (Spearman's  $\rho$ ) between the author degree and the number of referees and degree of referees, respectively, with 95% bootstrap confidence intervals.

	Number of referees	Degree of referees
Journal 1	0.13 [0.06, 0.19]	0.20 [0.13, 0.27]
Journal 2	0.27 [0.22, 0.33]	0.11 [0.04, 0.17]
Journal 3	0.06 [−0.01, 0.12]	0.02 [−0.04, 0.08]
Journal 4	0.10 [−0.03, 0.23]	0.12 [−0.01, 0.24]

not depend on a larger number of referees per paper. The average number of referee per paper (1.83) was actually smaller than the number of authors (2.33). The higher density was instead due to the larger share of referee who reviewed multiple papers, which is higher than the one of the authors who submitted two or more of their works in the dataset journals.

### 3.2. Referee and editorial bias

We used network measures and other variables presented above to analyse the potential sources of bias in the review process. The first step was to look at the editor's choice of referees. At least two decisions were potentially biased here. First, depending on the characteristics of the author, the editor could invite different referees (Ganguly & Mukherjee, 2017), who, in turn, could produce systematically different recommendations. Secondly, also depending on authors' characteristics, the editor could invite a different number of referees. Therefore, these decisions could reveal certain biases of the editorial process, which could treat certain authors differently due to their characteristics, e.g., reputation.

To check the first of this points, we estimated the correlation between characteristics of authors and referees of each paper derived from the affiliation networks showed in Figs. 1b and 2b. Given that the estimated centrality measures were highly correlated, with the degree of authors and referees showing the highest variance, we focused on this measure when discussing authors' and referees' properties. Table 5 shows that the degree of authors and referees was positively (though weakly) correlated in J1, J2 and possibly J4 (see also Section 3.3).

Given that the degree could reflect job experience and seniority, this would indicate that editors preferentially selected experienced referees for paper written by experienced authors. Otherwise, this could indicate that experienced referees accepted to review preferably papers from presumably "important" authors. We did find a similar pattern with correlations between the degree of authors and the number of referees (Table 5), where small correlations can be found in all but the third journal. Although weak, the fact that a certain degree of correlation is present is surprising given that, at this stage, we did not include in the analysis any other source of information able to reduce the variance of the data.

After looking at the editors, the second step was to focus on the work of referees. Given that referees were not randomly assigned to submissions, a potential source of bias here could be due to referee experience (measured as degree in the referee–paper network). Therefore, we tried to test whether the referees' position in the network could predict their recommendations. We estimated an ordinary least squares (OLS) model using the mean degree of the referees for a given paper, together with the number of referees and the journal ID as control variables and calculated if this predicted the cor-

**Table 6**

OLS estimates for the review score model.

	Coefficient	95% CI	Std. coefficient <sup>a</sup>
(Intercept)	0.047	[−0.002, 0.097]	–
Referee degree	−0.017	[−0.030, −0.005]	−0.063
Number of referees	0.078	[0.063, 0.093]	0.218
Journal 2	−0.018	[−0.049, 0.013]	–
Journal 3	0.102	[0.073, 0.131]	–
Journal 4	0.129	[0.090, 0.168]	–
$R^2$	0.084		
$F(5, 2907)$	53.460		

<sup>a</sup> When and how coefficients of categorical variables should be standardized represents a controversial issue (e.g., Gelman, 2008). We hence decided to proceed by only standardising continuous variables. As a consequence, the table does not show standardized coefficients for the “Journal” variable.

**Table 7**

Logistic model estimates on the probability of rejection.

	Coefficient	95% CI	Std. coefficient <sup>a</sup>
(Intercept)	3.468	[2.770, 4.185]	–
Review score	−11.840	[−12.725, −10.999]	−3.243
Number of referees	−0.350	[−0.573, −0.129]	−0.268
Author degree	−0.302	[−0.522, −0.093]	−0.161
Referee degree	−0.037	[−0.251, 0.179]	−0.116
Disagreement	1.255	[0.716, 1.799]	0.287
Author degree × disagreement	0.567	[0.089, 1.074]	0.130
Referee degree × disagreement	−0.320	[−0.821, 0.178]	−0.073
Journal 2	0.327	[−0.072, 0.728]	–
Journal 3	−0.029	[−0.416, 0.359]	–
Journal 4	−0.729	[−1.235, −0.219]	–
Pseudo- $R^2$ (McFadden)	0.553		

<sup>a</sup> When and how coefficients of categorical variables should be standardized represents a controversial issue (e.g., Gelman, 2008). We hence decided to proceed by only standardising continuous variables. As a consequence, the table does not show standardized coefficients for the “Journal” variable.

responding review score. Table 6 shows that referees with a higher degree tended preferably to assign lower review scores. However, the corresponding effect is rather small: results from a model including only the referees’ degree as predictor (i.e., excluding the controls) indicated that this variable alone explained about 2% of the review score variance. In addition, given that in general referees had assigned different papers by editors, we cannot exclude at this stage that the difference in recommendations was due to a difference in quality of the submissions instead of revealing a severe attitude of more experienced referees (on this, more detail in Section 3.3). Also notice that submissions with a larger number of referees tend to obtain higher scores.

The last step was to look at the editorial decision. Ideally, the only element influencing the editor’s decision should be the review score. We tried to predict the editorial decision through a model including the review score as a fixed effect. To look at potential sources of bias, in case of author and referee degrees, we tested the number of referees and their disagreement. We also added interaction terms between referees’ disagreement and referee and author degrees. This was to understand if editors could be influenced more likely by extra-elements, such as the degree (i.e., experience) of authors or referees, when dealing with contradictory reviews. Finally, given that the distribution of editorial decisions was different across our four journals, we added dummies for each journal but the first as predictor. Being the editorial decision an ordinal variable, using OLS regression was no longer appropriate. We estimated our model using two different strategies: (i) logistic regression, and (ii) ordered logistic regression with a cumulative link function.

As Table 3 showed that around half of the papers were rejected in the first review round, to efficiently use our data we split the editorial decision variable at this level. A logistic model was then used to predict whether a paper was rejected vs. invited for revisions or accepted. As expected, the review score resulted the strongest predictor for the editorial decision (Table 7). Furthermore, consistently with Table 3, journal dummies showed significant differences in the rejection rate. More interestingly, the fact that editorial decisions were influenced by the authors’ degree – both as pure effect and in interaction with disagreement – suggests that editors interpreted referee recommendations differently depending on the author’s reputation in the scientific community. In addition, having more and more experienced referees decreased the probability of being rejected independently of the other factors.

To check the robustness of our results and fully exploit our data on all the possible levels of the editorial decision, we tested a ordered logistic model with a cumulative link function to predict whether a paper was accepted, invited for resubmission with minor revision, with major revision, or rejected. Table 8 shows that most estimates, including the effect of the author’s degree, were qualitatively similar to what we obtained from the simpler logistic model above. However, the effect of some factors slightly varied. Most notably, (i) the effect of the number of referee strongly decreased, with a corresponding confidence interval (CI) that included zero in the new model; (ii) the positive effect of the interaction between

**Table 8**

Cumulative-link ordered logistic estimations predicting paper acceptance (reference category), minor revision, major revision or rejection.

	Coefficient	95% CI	Std. coefficient <sup>a</sup>
Review score	–11.388	[–12.021, –10.779]	–3.119
Number of referees	–0.169	[–0.349, 0.010]	–0.129
Author degree	–0.204	[–0.363, –0.045]	–0.113
Referee degree	0.006	[–0.176, 0.189]	–0.109
Disagreement	1.667	[1.219, 2.121]	0.382
Author degree × disagreement	0.368	[–0.028, 0.776]	0.085
Referee degree × disagreement	–0.463	[–0.892, –0.036]	–0.106
Journal 2	0.321	[–0.007, 0.652]	–
Journal 3	–0.099	[–0.413, 0.215]	–
Journal 4	–0.780	[–1.160, –0.399]	–
$\theta(\text{Accepted} \text{Minor revisions})$	–12.163	[–12.989, –11.336]	–
$\theta(\text{Minor revisions} \text{Major revisions})$	–7.413	[–8.077, –6.748]	–
$\theta(\text{Major revisions} \text{Reject})$	–2.834	[–3.399, –2.270]	–
Pseudo- $R^2$ (McFadden)	0.509		

<sup>a</sup> When and how coefficients of categorical variables should be standardized represents a controversial issue (e.g., Gelman, 2008). We hence decided to proceed by only standardising continuous variables. As a consequence, the table does not show standardized coefficients for the “Journal” variable.

the author degree and disagreement was also reduced close to zero; (iii) the negative effect of the interaction between the referee degree and disagreement increased (in absolute terms) and the corresponding CI no longer included zero. Since only 14% of papers in the dataset received an *accept* or *minor revision*, we did not consider other logistic models to further test if the estimates of the explanatory variables are consistent (proportional) across different thresholds for the editorial decision.

### 3.3. A comprehensive analysis of the review process

Although interesting, the models above could not provide a comprehensive picture of the review process yet. Furthermore, the authors' degree (which possibly reflected experience and seniority) could systematically affect the quality of a paper. In other words, even if we observed a significant relation between this variable and the review score, this could simply reflect the fact that authors with a higher degree were more likely to submit papers of higher quality. To get a more comprehensive picture of the whole process and better distinguish between biased and unbiased paths leading to the editorial decision, we trained a *Bayesian network* (Friedman, Geiger, & Goldszmidt, 1997) to estimate the probability of a paper rejection on the basis of all the variables previously considered.<sup>3</sup>

Bayesian networks model interdependencies among a set of variables as connections in a directed acyclic graph, where nodes represent random variables – e.g., the review score – and edges represent conditional dependencies – e.g., how much the review score had an effect on the rejection probability of a paper. We opted for this method for two reasons. First, the structure of a Bayesian network is learned inductively from the data, with no need for the researchers to provide prior assumptions on the relevant causal effects. Secondly, once learned the network could be used to derive probabilities of the event of interest given a set of conditions on the other variables – e.g., how likely is the rejection of a paper given that its author has a network degree higher than a certain value.

Fig. 3 shows the structure of the Bayesian network inductively learned from the data through maximum likelihood estimation.<sup>4</sup> The algorithm is fully data-driven and used the training set (80% of the data) to learn both the network structure and the direction of each edge. Non-significant paths were automatically excluded. The only external constraint on the network structure was that there could be no link from *reject* to any of the other nodes. Note that, to increase the readability of the figure, the *journal* variable, which significantly affected all other nodes, was not included.

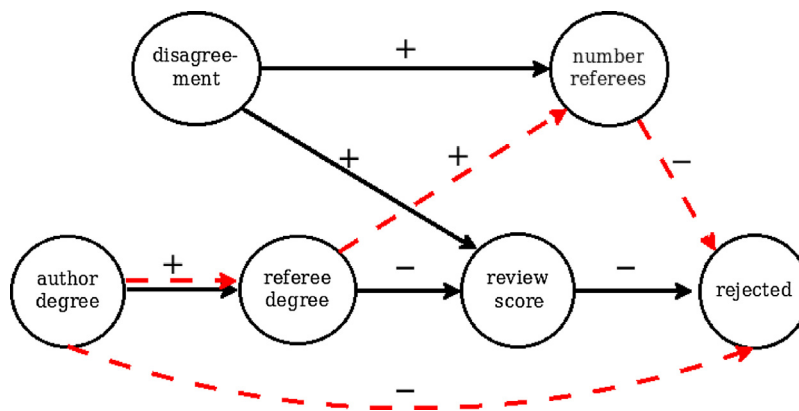
All parameters of the Bayesian network were learned on a training set consisting of a random sample of 80% of our data, while the remaining 20% were used for model validation. The resulting network successfully predicted 84% of the validation set cases when given all the information but the editorial decision. Table 9 shows the standardized coefficients for the network paths. Path coefficients express the effect that each upstream node has to the downstream nodes it is connected to. These coefficients were learned by the algorithm while ignoring the information on the journal, which implies that they can be interpreted as aggregated across the four outlets. Note that we omitted the coefficients linking the *journal* variables to the other nodes because, being *journal* a categorical variable, the analysis led to estimate one different coefficient per journal, with no straightforward way to summarize them in a single value.

To interpret the resulting network, it is important to follow all the different paths going from the authors' degree to the editorial decision. The one going through the review score (in green in Fig. 3) does not necessarily represent biases, as it

<sup>3</sup> For technical reasons, it is not possible to have a final node that has more than two levels, when parent nodes are continuous variables. However, the logistic and ordinal regression models in the previous section imply that results were not dramatically different when the editorial decision is included as a binary variable instead of considering all four levels.

<sup>4</sup> We tried both a Hill Climbing and different Incremental Association algorithms, all leading to the same structure.





**Fig. 3.** Structure of the Bayesian network resulting from a training set based on 80% of our data. Red (dashed) arrows mark the “biased” paths linking *author degree* to *rejection* without going through the *review score*. Plus and minus symbols indicate the direction of effect. Note that the categorical *journal* variable, which significantly affects all other nodes, is not included in the figure for better readability. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of the article.)

**Table 9**

Standardized path coefficients for the network in Fig. 3. Being *journal* categorical, no coefficients are shown for paths involving this variable.

Path	Standardized coefficient
Author degree → rejected	−0.041
Author degree → referee degree	0.066
Referee degree → review score	−0.129
Review score → rejected	−0.724
Referee degree → number referees	0.144
Disagreement → number referees	0.413
Disagreement → review score	0.263
Number referees → rejected	−0.047

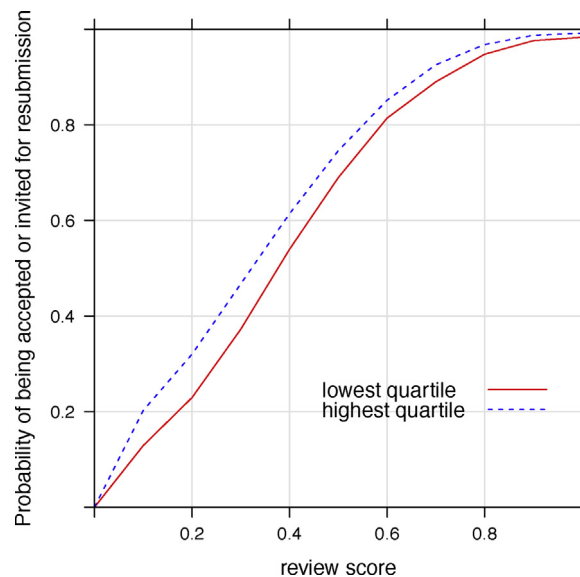
could simply reflect the fact that more experienced authors are more likely to produce high-quality papers. On the other hand, any path linking the authors' degree with the editorial decision and not going through the review score (red in Fig. 3) can be considered as a potential source of bias since it excludes the most central – and in principle only – element that should count in the review process.

The most noticeable insight from the learned network was the direct path from the authors' degree to the editorial decision (the “reject” node) (Fig. 3), which can be interpreted as a direct bias on editorial decisions due to the experience/reputation of authors. However, the strength of this path is only about 6% of the strength of the path linking the review score to the reject variable. A second indirect path from the authors' degree to the editorial decision node went through the degree and number of referees. This suggested the presence a further bias due to the selection of different (in numbers and quality) referees for different authors who submitted to the same journal. Finally, we found that the journal node (not shown in the figure) affected all other variables, which was consistent with the intuition that different journals had, among others, different editors and rejection rates.

Using the trained network, we estimated how the probability of being accepted or invited for revisions (i.e., not to be rejected) varied depending on the review score and the author degree. Fig. 4 shows that it monotonically increased with the review score. Furthermore, authors belonging to the top quartile with regard to the network degree systematically had a 5–10% higher chance of being accepted when compared with the ones in the bottom quartile. It is worth noting that this probability was estimated after controlling for review scores, which suggested that these editorial decisions were not due to the quality of submissions – resulting in more positive reviews – but to the reputation of authors. This bias was more prominent when review scores were relatively low, which suggested that more important authors more easily escaped a presumably deserved rejection when they did not submit brilliant papers.

#### 4. Discussion and conclusions

In our work, we tried to use network data creatively to estimate editorial bias in four computer science journals, without recurring to in-depth information on manuscripts, authors or reviewers involved. It is worth noting that cross-journal, large-scale data on the editorial process of scholarly journals that include full information on manuscripts, authors and review scores are unfortunately only rarely available (Squazzoni et al., 2017). Need for preserving author anonymity, unavailability of ex-ante, objective measures of a submission's quality, lack of data to reconstruct network effects due to the prestige of scientists, are all factors that impede to quantitatively assess the rigour and impartiality of the editorial process in scholarly journals (Batagelj et al., 2017; Casnici et al., 2017). While data in principle exist, thanks to the full digitalisation of the editorial



**Fig. 4.** Relation between the review score of a paper and the probability of being accepted or invited for resubmission. The two lines show papers from authors belonging to the top and bottom quartile with regard to the authors' degree.

process in almost all journals, all quantitative indicators which have recently increased our understanding of research, citation and publication patterns and dynamics are rarely applied to measure peer review and editorial processes on a large scale (Mutz, Bornmann, & Daniel, 2017). In this respect, our study has the advantage of measuring certain characteristics of the editorial process across comparable journals without the need for in-depth, individual data, which are often unavailable.

The main limitation of the study resides in its reliance on a small sample of journals from a single discipline. Whether its conclusions could be generalized to other cases and scientific traditions hence remains an open question. Future research should try to apply and extend the proposed methodology to larger and more varied samples, taking also into account that several variables may interact to produce (or prevent) the reputation bias highlighted by the data analysed here.

This said, our findings clearly showed that editors could be biased towards more prestigious authors by giving them higher chances of publication even when they presumably did not deserve a special treatment. Although in our case the editorial bias was not so large to drastically alter the outcome of the review process, it was still significant, especially for low review scores. The resulting “implicit” premium to more prestigious authors can be due to several reasons. First, editors could rationally predict higher citations of articles from prestigious authors, who probably are even more prolific. As suggested by Petersen et al. (2014), while the effect of the prestige of authors on article citations may evaporate in the long run, this effect is effective especially in driving citation count early in the typical citation life cycle. Given that a journal's impact factor is calculated in a two years window, and that attention decay by the community seems more pronounced nowadays (Parolo et al., 2015), editors could be more indulgent with more prestigious authors as a means to maximise their journal's impact factor.

Secondly, prestigious authors could react more strategically against a manuscript rejection by refusing to review or avoiding to submit further work, so punishing editors and their journals afterwards (Squazzoni & Gandelli, 2013). On the other hand, given the complexity of any editorial decision and perhaps the unavailability of fully informative reviews, editors could intentionally trust more the confidence of prestigious authors than the opinion of imprecise reviewers at least to avoid false negatives (Lee et al., 2013). This would suggest that although the editorial management of a submission should be in principle an “one-shot game”, the past performance of authors or the predictive impact of their research could implicitly influence the process.

However, there is also another interpretation of our findings, which does not convey any negative message on the editorial process of these four journals, whose estimated level of bias was in any case only minimal. On the one hand, while reviewers and editors are increasing under the spotlight for a variety of explicit or even implicit bias (Lee et al., 2013; Pinholster, 2016), it must be said that they operate in a situation of hyper-competition, time trade-offs and pressures due to the increasing number of submissions, journals and publications (Edwards & Roy, 2017; Kovanis, Porcher, Ravaud, & Trinquart, 2016). On the other hand, this criticism implicitly suggests that editorial decisions have an optimal reference point, with reviewers and editors being eventually responsible for deviations and mistakes. These perspectives tend to deny the fact that peer review and editorial processes are intrinsically complex and imperfect because they cannot but reflect what science itself essentially is (Cowley, 2015). While it is important to explore measures to assess the editorial process and improve its transparency and efficacy, the idea that an optimal world would exist, if only scientists were not exposed to external influences and distortive incentives or had not malicious intentions, is naive.

## Author contributions

Giangiaco Bravo: Conceived and designed the analysis; Performed the analysis; Wrote the paper.  
 Mike Farjam: Contributed data or analysis tools; Performed the analysis; Wrote the paper.  
 Francisco Grimaldo Moreno: Collected the data; Contributed data or analysis tools; Wrote the paper.  
 Aliaksandr Birukou: Contributed data or analysis tools.  
 Flaminio Squazzoni: Conceived and designed the analysis; Wrote the paper; Other contribution.

## Acknowledgements

The authors gratefully acknowledge Annette Hinze, Alfred Hofmann, Katarina Kreissig and Tamara Welschot for their help with the data extraction and the anonymisation procedure, and Ralf Gerstner for both helping with the data extraction and his useful comments on an earlier version of the paper. A preliminary version of this article was presented at a WG2 PEERE meeting at UvA in Amsterdam on February 2017 and benefited from comments by Ana Marušić, Bahar Mehmani, Virginia Dignum and Michael Willis. This work was partially supported by the COST Action TD1306 “New frontiers of peer review” ([www.peere.org](http://www.peere.org)) and by the Spanish Ministry of Science and Innovation Project TIN2015-66972-C5-5-R. Last but not least, the authors gratefully acknowledge the comments made by two anonymous referees, which nicely helped to improve the quality of the paper.

## References

- Batagelj, V., Ferligoj, A., & Squazzoni, F. (2017). The emergence of a field: A network analysis of research on peer review. *Scientometrics*, 113(1), 503–532. <http://dx.doi.org/10.1007/s11192-017-2522-8>
- Birukou, A., Wakeling, J. R., Bartolini, C., Casati, F., Marchese, M., Mirylenka, K., Osman, N., Ragone, A., Sierra, C., & Wassef, A. (2011). Alternatives to peer review: Novel approaches for research evaluation. *Frontiers in Computational Neuroscience*, 5. <http://dx.doi.org/10.3389/fncom.2011.00056>
- Bornmann, L. (2011). Scientific peer review. *Annual Review of Information Science and Technology*, 45(1), 197–245. <http://dx.doi.org/10.1002/aris.2011.1440450112>
- Bornmann, L., & Daniel, H.-D. (2009). Extent of type I and type II errors in editorial decisions: A case study on *Angewandte Chemie International Edition*. *Journal of Informetrics*, 3(4), 348–352. pii:S1751157709000406.
- Bornmann, L., & Williams, R. (2017). Can the journal impact factor be used as a criterion for the selection of junior researchers? A large-scale empirical study based on ResearcherID data. *Journal of Informetrics*, 11(3), 788–799. pii:S1751157717300378.
- Casnici, N., Grimaldo, F., Gilbert, N., Dondio, P., & Squazzoni, F. (2017). Assessing peer review by gauging the fate of rejected manuscripts: The case of the journal of artificial societies and social simulation. *Scientometrics*, <http://dx.doi.org/10.1007/s11192-017-2241-1>
- Casnici, N., Grimaldo, F., Gilbert, N., & Squazzoni, F. (2016). Attitudes of referees in a multidisciplinary journal: An empirical analysis. *Journal of the Association for Information Science and Technology*, 68(7), 1763–1771. <http://dx.doi.org/10.1002/asi.23665>
- Cowley, S. J. (2015). How peer-review constrains cognition: On the frontline in the knowledge sector. *Frontiers in Psychology*, 6, 1706. <http://dx.doi.org/10.3389/fpsyg.2015.01706>
- Edwards, M. A., & Roy, S. (2017). Academic research in the 21st century: Maintaining scientific integrity in a climate of perverse incentives and hypercompetition. *Environmental Engineering Science*, 34(1), 51–61. <http://dx.doi.org/10.1089/ees.2016.0223>
- Faust, K. (1997). Centrality in affiliation networks. *Social Networks*, 19(2), 157–191. [http://dx.doi.org/10.1016/s0378-8733\(96\)00300-0](http://dx.doi.org/10.1016/s0378-8733(96)00300-0)
- Friedman, N., Geiger, D., & Goldszmidt, M. (1997). *Bayesian network classifiers*. *Machine Learning*, 29(2–3), 131–163.
- Fyfe, A. (2015). Journals, learned societies and money: Philosophical transactions, ca. 1750–1900. *Notes and Records of the Royal Society*, 69(3), 277–299. <http://rsnr.royalsocietypublishing.org/content/69/3/277>
- García, J. A., Rodríguez-Sánchez, R., & Fdez-valdivia, J. (2015). The author–editor game. *Scientometrics*, 104(1), 361–380. <http://dx.doi.org/10.1007/s11192-015-1566-x>
- Gelman, A. (2008). Scaling regression inputs by dividing by two standard deviations. *Statistics in Medicine*, 27(15), 2865–2873.
- Grimaldo, F., & Paolucci, M. (2013). A simulation of disagreement for control of rational cheating in peer review. *Advances in Complex Systems*, 16(07), 1350004. <http://dx.doi.org/10.1142/S0219525913500045>
- Hsiehchen, D., & Espinoza, M. (2016). Detecting editorial bias in medical publishing. *Scientometrics*, 106(1), 453–456. <http://dx.doi.org/10.1007/s11192-015-1753-9>
- Kovanis, M., Porcher, R., Ravaud, P., & Trinquart, L. (2016). The global burden of journal peer review in the biomedical literature: Strong imbalance in the collective enterprise. *PLoS ONE*, 11(11), 1–14. <http://dx.doi.org/10.1371/journal.pone.0166387>
- Lee, C. J., Sugimoto, C. R., Zhang, G., & Cronin, B. (2013). Bias in peer review. *Journal of the American Society for Information Science and Technology*, 64(1), 2–17. <http://dx.doi.org/10.1002/asi.22784>
- Lin, Z., Hou, S., & Wu, J. (2016). The correlation between editorial delay and the ratio of highly cited papers in nature, science and physical review letters. *Scientometrics*, 107(3), 1457–1464. <http://dx.doi.org/10.1007/s11192-016-1936-z>
- Moustafa, K. (2015). Is there bias in editorial choice? Yes. *Scientometrics*, 105(3), 2249–2251. <http://dx.doi.org/10.1007/s11192-015-1617-3>
- Mutz, R., Bornmann, L., & Daniel, H.-D. (2017). Are there any frontiers of research performance? Efficiency measurement of funded research projects with the Bayesian stochastic frontier analysis for count data. *Journal of Informetrics*, 11(3), 613–628. pii:S1751157716303741.
- Parolo, P. D. B., Pan, R. K., Ghosh, R., Huberman, B. A., Kaski, K., & Fortunato, S. (2015). Attention decay in science. *Journal of Informetrics*, 9(4), 734–745. pii:S1751157715200442.
- Petersen, A. M., Fortunato, S., Pan, R. K., Kaski, K., Penner, O., Rungi, A., Riccaboni, M., Stanley, H. E., & Pammolli, F. (2014). Reputation and impact in academic careers. *Proceedings of the National Academy of Sciences*, 111(43), 15316–15321. <http://dx.doi.org/10.1073/pnas.1323111111>
- Petersen, J., Hattke, F., & Vogel, R. (2017). Editorial governance and journal impact: A study of management and business journals. *Scientometrics*, 112(3), 1593–1614. <http://dx.doi.org/10.1007/s11192-017-2434-7>
- Pinholster, G. (2016). Journals and funders confront implicit bias in peer review. *Science*, 352(6289), 1067–1068. <http://mfkp.org/INRMM/article/14042738>
- Resnik, D. B., & Elmore, S. A. (2016). Ensuring the quality, fairness, and integrity of journal peer review: A possible role of editors. *Science and Engineering Ethics*, 22(1), 169–188. <http://dx.doi.org/10.1007/s11948-015-9625-5>
- Righi, S., & Takács, K. (2017). The miracle of peer review and development in science: An agent-based model. *Scientometrics*, 113(1), 587–607. <http://dx.doi.org/10.1007/s11192-017-2244-y>
- Sarigöl, E., García, D., Scholtes, I., & Schweitzer, F. (2017). Quantifying the effect of editor–author relations on manuscript handling times. *Scientometrics*, <http://dx.doi.org/10.1007/s11192-017-2309-y>
- Schneier, B. (1996). *Applied cryptography: Protocols, algorithms, and source code in C* (2nd ed.). New York: Wiley.

- Ganguly, N., & Mukherjee, A., 2017. Influence of reviewer interaction network on long-term citations: A case study of the scientific peer-review system of the Journal of High Energy Physics. <http://adsabs.harvard.edu/abs/2017arXiv170501089S>.
- Siler, K., Lee, K., & Beroc, L. (2015). Measuring the effectiveness of scientific gatekeeping. *Proceedings of the National Academy of Sciences*, 112(2), 360–365.
- Squazzoni, F., & Gandelli, C. (2012). Saint Matthew strikes again: An agent-based model of peer review and the scientific community structure. *Journal of Informetrics*, 6(2), 265–275, pii:S1751157711001179
- Squazzoni, F., & Gandelli, C. (2013). Opening the black-box of peer review: An agent-based model of scientist behaviour. *Journal of Artificial Societies and Social Simulation*, 16(2), 3. <http://jasss.soc.surrey.ac.uk/16/2/3.html>
- Squazzoni, F., Grimaldo, F., & Marusic, A. (2017). Publishing: Journals could share peer-review data. *Nature*, 546, 352. <http://dx.doi.org/10.1038/546352a>
- Teele, D. L., & Thelen, K. (2017). Gender in the journals: Publication patterns in political science. *PS: Political Science & Politics*, 50(2), 433–447.

**Giorgio Bravo** is professor at the Department of Social Studies and coordinates the computational social science group at the Centre for Data Intensive Sciences and Applications, Linnaeus University, Sweden. His main research interests cover collective action problems and environmental sustainability.

**Mike Farjam** is a post-doc at the Department of Social Studies at Linnaeus University, Sweden. His main interest is the perception of risk involved when interacting with others/machines and cooperation in general.

**Francisco Grimaldo Moreno** received his Ph.D. in computer science from the University of Valencia in 2008. He is associate professor at the University of Valencia and research fellow at the Italian National Research Council. His research interests are agent-based modelling and simulation, machine learning and data analysis and visualization.

**Aliaksandr Birukou** received his Ph.D. in computer science from the University of Trento in 2009. He is now Executive Editor for Computer Science at Springer Nature.

**Flaminio Squazzoni** is associate professor of economic sociology at the Department of Economics and Management, University of Brescia, Italy, where he leads the GECS-Research Group on Experimental and Computational Sociology.