Is the panel fair? Evaluating panel compositions through network analysis. The case of research assessments in Italy

IS THE PANEL FAIR? EVALUATING PANEL COMPOSITIONS THROUGH NETWORK ANALYSIS. THE CASE OF RESEARCH ASSESSMENTS IN ITALY

Alberto Baccini
Dipartimento di Economia Politica e Statistica
Università degli Studi di Siena
Siena, Italy
alberto.baccini@unisi.it

Cristina Re
Dipartimento di Economia Politica e Statistica
Università degli Studi di Siena
Siena, Italy
cristina.re@unipr.it

ABSTRACT

In research evaluation, the fair representation of panels is usually defined in terms of observable characteristics of scholars such as gender or affiliations. An an empirical strategy is proposed for exploring hidden connections between panellists such that, despite the respect of formal requirements, the panel could be considered alike as unfair with respect to the representation of diversity of research approaches and methodologies. The case study regards the three panels selected to evaluate research in economics, statistics and business during the Italian research assessment exercises. The first two panels were appointed directly by the governmental agency responsible for the evaluation, while the third was randomly selected. Hence the third panel can be considered as a control for evaluating about the fairness of the others. The fair representation is explored by comparing the networks of panellists based on their co-authorship relations, the networks based on journals in which they published and the networks based on their affiliated institutions (universities, research centres and newspapers). The results show that the members of the first two panels had connections much higher than the members of the control group. Hence the composition of the first two panels should be considered as unfair, as the results of the research assessments.

Keywords Procedural Fairness · Research Assessment · Gatekeepers of Economics · Social Network Analysis

1 Introduction

Research evaluation has now become one of main instruments for university and research policies. It is no longer considered as just an essential step for distributing resources to competing research projects and for hiring and promoting researchers; it is also the core of the administrative processes labelled "performance based research funding" where ex-post evaluation is used for financing universities and other research institutions [Langfeldt, 2004, Hicks, 2012, Zacharewicz et al., 2018]. As a consequence, research evaluation has gained centrality in the scientific knowledge-making process, so much that it has taken "the function of gatekeeping, filtering, and legitimating" it [Lamont and Huutoniemi, 2011, p. 209].

Despite their procedural variety, evaluation is usually governed by panels of experts or peers that managed the processes of peer review [Whitley et al., 2010]. Non-scientific characteristics of these panels, like sex, age, or affiliation, may affect evaluation practices, with a concrete risk that specific groups can capture regulatory processes of scientific evaluation systems. This problem has been mainly analyzed in reference to the *ex-ante research evaluation,* specifically in reference to editorial evaluation adopted by journals and to evaluation of projects for funding [Cole, 1992, Cole and Cole, 1981, Cole et al., 1979, Lamont and Huutoniemi, 2011, Rahman et al., 2016]. Only a few studies have been done on the *ex-post research evaluation,* which is referred to national research assessments such us the ones applied in United Kingdom and Italy [Lee, 2007, Lee et al., 2013, Harley and Lee, 1997, Baccini, 2016c,b, Corsi et al., 2010, 2011].

The central notion in these discussions is *procedural fairness*, which should be "concerned with procedures used to arrive at [fair] outcomes" [Beersma and De Dreu, 2003, p. 220] and that includes the *fair representation* of all affected parties involved in the decision-making process [Leventhal, 1980]. The concrete applications of this procedural fairness is present in the European Peer Review Guide of the European Science Foundation [ESF, 2011] and it was adopted in the composition of panels of British Research Excellence Framework (REF) and also of the Italian research assessments. However, the problem is how to define who the affected parties are and what dimensions to take into account in defining their fair representation. The usual way is to define observable characteristics of scholars, such as research field, gender, affiliation and geography. But if a discipline is characterized by the coexistence of many schools of thought with different approaches, methodologies and policy recipes, a fair composition in terms of observable characteristics may not guarantee a fair representation of the different schools of thought. The unfair selection of panellists may contribute to the prevalence of one of competing paradigms or schools of thought, by reinforcing the normative standard, pre-existing journal rankings and more generally the hierarchy of a discipline.

Probably the most studied case is economics (see Corsi et al. [2019] for a review of the literature), also because the evaluation of economic theories affects economic policy [Ferguson and Johnson, 2018]. Pioneer of these studies was Frederic S. Lee, who analyzed and harshly criticised the Research Assessments Exercise (RAE) in the United Kingdom [Harley and Lee, 1997, Lee, 2007, Lee et al., 2013]. He explicitly individuated a precise role for the panel of RAE, composed by paradigmatic homogeneous experts "controlled by mainstream economists [who] have used [the RAE] to support particular neoclassical research over heterodox research and promote neoclassical departments over more pluralistic ones" [Lee, 2007, p.15]. A process aimed "to achieve a discipline-desired outcome that was (and is) compatible with the Government's pro-market ideological agenda" [Lee, 2007, p.14].

The Italian case also has been considered as an internationally relevant example on how it can disregard heterodox schools and historical methods in favour of mainstream approaches and quantitative methods in economics [Pasinetti, 2006, Corsi et al., 2010, 2011]. Re [2019] claimed that this process was connected to a cultural and political change from the Keynesian to the Ordoliberal ideology. The lack of fairness in the composition of the committees in charge of economic evaluation was also highlighted: their homogeneity

could probably have minimized the voices of disagreement with the evaluation methods and rules adopted [Pasinetti, 2006, Baccini, 2011, 2014, 2016c, Baccini and Ricciardi, 2012].

This paper proposes an empirical strategy for exploring if, despite formal adherence to fair panel composition in terms of easily observable characteristics, it is possible to detect connections among members such that panel composition can be considered as unfair. The case studies regards the composition of the panels appointed to evaluate research in economics, statistics and business during the Italian research assessment exercises, the so called VQR 2004-2010, VQR 2011-2014 and VQR 2015-2019. The Italian case is of particular interest since the three considered panels were selected by ANVUR, the governmental agency for the evaluation of university and research, by using different rules. Indeed, as it is detailed later on, while the first two panels were appointed directly by the governing board of ANVUR, the third was instead selected randomly by a lot. Hence the third panel can be considered as a sort of control group [Gillies, 2014]. For investigating the fairness of panel composition, the social and intellectual connections between panel members are investigated through network analysis techniques. Three kinds of connections, and hence three kinds of networks are considered and explored: (i) the coauthorship networks built by considering the publications of the members of the panels; (ii) the journal based networks built by considering the journals where the panellists publish their scholarly articles; and (iii) the "affinity networks" built by considering panellists' affiliations to universities and research centres, and their collaborations to newspapers or blogs.

The organization of the article is as follows: Section 2 reviews literature on procedural fairness; in Section 3, the Italian research assessments and the panel selection methods are described; in Section 4 the research design is presented; sections 5-7 report the analysis, respectively, of the co-authorship networks, of the journal based networks, and of the affinity networks. Section 8 discusses results and concludes.

2 The notion of fair representation in research evaluation

It has been shown that some of the perverse effects of peer review, such as cronyism, the pursuit of self-interest, and cognitive particularism, may be influenced by the way panels are set up [Lamont and Huutoniemi, 2011]. Both potential and observed risks in the peer review system have been discussed and documented. It is argued that the system is conservative and suppresses innovative research. Effects such as nepotism and old-boyism in peer review are seen to hinder pioneering research [Chubin and Hackett, 1990, Roy, 1985], while "cognitive particularism", "favoritism for the familiar" and "scholarly bias" support the school viewpoint or research topic the reviewers themselves are conducting (see, e.g., Porter and Rossini [1985], Travis and Collins [1991], Langfeldt [2004]). The relations between panellists and evaluated scholars is a channel of bias: Wennerås and Wold [1997] found that higher competence scores are given to applicants who are affiliated with a panellist than to applicants without such ties. Moreover, another possible group effect is groupthink, which refers to "a deterioration of mental efficiency, reality testing, and moral judgement that results from in-group pressures" [Janis, 1982, p.9]. Loyalty to the group "requires each member to avoid raising controversial issues, questioning weak arguments, or calling a halt to soft-headed thinking" [Janis, 1982, p.12]. Finally, according to Van den Brink [2010], in the Netherlands, more women in appointment committees led to higher numbers of women being appointed as full professor. This indicates preferences for same-sex candidates.

As a solution to these problems, some studies [Roy, 1985, Bell, 1992, GAO, 1994] focused on the fairness of the peer review process showing that fair process of evaluating knowledge can be put in place once particularistic considerations are eliminated. Lamont and Huutoniemi [2011] claimed instead that extracognitive factors do not corrupt the evaluation process but are intrinsic to it and that the "fairness of the process is not undermined by nonrational features but is created through intersubjective rules that evaluators follow to distinguish between legitimate and illegitimate behavior". For this reason, studies on evaluation and

peer reviews focused on the *procedural fairness*, which "is concerned with procedures used to arrive at those outcomes" [Beersma and De Dreu, 2003, p.220] and not solely on the allocative results. While the concept of *distributive fairness* restricts the analysis of perceived justice only to the final distribution of reward or punishment, in the *procedural fairness* people perceive fairness not solely in terms of the distribution of reward but also in terms of the social system which generates that distribution. "For example, fairness may be judged in terms of a procedure's consistency over time and across persons; its accuracy and prevention of personal bias; or its representativeness of the values, interests, and outlook of important subgroups in the population of persons affected by the allocative process" [Leventhal, 1980, p.54]. Thus, among these factors there is also the *fair representation* of all affected parties involved in the decision-making process. In the case of research assessment, this means to focus on the fairness of the composition of panels called to evaluate research.

In fact the role of panellists is very similar to that of the members of a popular jury in a trial [Baccini and Ricciardi, 2012]. In order to have a fair judgment by a panel of judges, it is necessary to designate a fair jury and therefore presumably less inclined to partiality.

Jury selection is the legal procedure in which randomization is most widely used [Duxbury, 2002]. Its use is justified as a way to give a fair chance to every qualified citizen to serve on a jury, and to provide defendants and litigants to be tried by a "representative cross-section of the population" [Duxbury, 2002, p. 75]. The rules of the United States *Jury Selection Service Act*, for example, state that the jury must be appointed by selecting "at random from a fair cross-section of the community". The fair composition of the popular jury aims to ensure fairness of the judgment. "Achieving representative cross-sections of the community in jury venires, and ensuring that our civil juries reflect the community as well, are essential components contributing to the fairness and legitimacy of our civil justice system [..] and the representativeness of juries is not merely an aspiration but a guarantee under state and federal constitutions and statutes" [Hans, 2021, p.1].

In particular, it has been shown that juries that reflect the full range of community perspectives are in a position to incorporate these diverse views into their fact finding. The best-known and best-documented examples concern the need to balance popular juries from the point of view of ethnic groups because it is believed that a jury composed mainly of members of the same ethnic group tends to be favourable towards a defendant of the same group, and unfavourable to a defendant from a different ethnic group. For example, Sommers and Ellsworth [2003] in a mock jury experiment, comparing the deliberations of all-white and racially mixed juries, discovered that diverse jury deliberations were more accurate, more expansive, and longer. It was not simply that the minority jurors contributed new and different information, the white jurors acted differently in all-white versus mixed-race juries: they made fewer factual mistakes, and raised more issues and evidence, during the deliberation. Moreover, representative juries are more likely to be seen as legitimate decision makers, which in turn contributes to public confidence in the justice system. For all these reasons, "courts should ensure that jury selection procedures serve the goal of maximizing the representativeness of jury pools and civil juries" [Hans, 2021, p.8].

Similar considerations can be found in documents related to the design of research assessments. For instance, in the document regarding the recruitment of panels for the British RAE, it is recommended a fair representation of "all affected parties" involved in the decision-making process. It was stated that the selection of panellist has to ensure that "the overall body of members reflects the diversity of the research community, including in terms of age, gender, ethnic origin, scope and focus of their home institution, and geographical location which represents the international reference on the subject" [REF, 2010]. The European Peer Review Guide of the European Science Foundation suggests that "the goal should be to ensure availability of diverse viewpoints, scientific perspectives and scholarly thinking" and that the criteria to be adopted for the selection of experts must also be the 'diversity' that is expressed in terms of "gender balance, scholarly thinking, background, geography, turnover" [ESF, 2011].

The basic tenet of this kind of reasoning is that, as for legal juries, a panel composed by a cross-section of members of a scientific community may reach not only procedurally fair decisions, but also substantively better decisions [Elster, 1989, p. 97].

3 The Italian research assessment exercises and the selection of panellists.

In the Italian case, the question of the fairness of panel composition was mainly considered in terms of easily observable characteristics of panellists. The National Agency for the Evaluation of the University and Research (ANVUR) was charged in 2011 of realizing a newly designed research assessment, called VQR 2004-2010 (DM 17 of 15 July 2011). Afterwards, two other research assessments were designed and realized: the VQR 2011-2014 (DM of 27 June 2015) and the VQR 2015-2019 (DM of 25 September 2020). The three VQRs were mandatory and organized around research areas as defined by the National University Council. The assessment of each area was managed and realized by a panel of scholars, the so called GEV (Group of Evaluation Experts). By and large, the areas were classified as "bibliometric" and "non-bibliometric". In both VQR 2004-2010 and VQR 2011-2014, the evaluation of bibliometric areas was conducted by a prevalent use of bibliometric algorithms, while for non-bibliometric areas it was realized by expert review. The VQR 2015-2019 adopted instead expert review, informed by bibliometrics, for all the research areas. The role of the panel was crucial in the three VQRs, since they defined for each area the specific rules for the evaluation. In particular panels decided bibliometric criteria, they defined the procedures for deciding which works should be evaluated with bibliometrics and which with peer review, they chose and coordinated the reviewers, they summarized the review reports, they evaluated in many cases directly the works submitted for evaluation [Baccini, 2016c,b, Baccini and De Nicolao, 2016].

Given their crucial role, the question of panel composition is central in the quality and credibility of the research assessment. The above cited Ministerial decrees defined also the procedures for the selection of panel members. These procedures are described in detail in the Supplementary materials A1. Here, it is sufficient to mention that the members of the committees for VQR 2004-2010 and VQR 2011-2014 were appointed directly by the ANVUR governing board mainly but not exclusively from among the Italian and foreign scholars who had applied in response to public calls to serve on the committees. Instead, the members of the panels for VQR 2015-2019 were selected exclusively by lot among those who applied as panellists. This institutional discontinuity introduced for the VQR 2015-2019 dramatically weakened the power of ANVUR governing board in the appointment of panels. It was the result of a political choice made by a government supported by a different majority than the previous ones.

Indeed, the VQR 2004-2010 procedure was designed and completed by minister Maria Stella Gelmini of a center-right government led by prime minister Silvio Berlusconi, and by Francesco Profumo, minister of the technical government of Mario Monti. The VQR 2011-2014 was designed and completed by Minister Stefania Giannini of a center-left government led by prime minister Matteo Renzi. The VQR 2015-2019 was originally designed by minister Lorenzo Fioramonti, of the Five-stars Movement (an 'anti-establishment' party) and center-left government led by prime minister Giuseppe Conte. Indeed, it can be conjectured that minister Fioramonti, in some occasions explicitly critical of previous research assessments, changed the rules for the appointment of panel members by following the public discussion developed during the previous years, which will be briefly illustrated below.

As anticipated, the literature on the procedural fairness suggests to compose panels in such a way that there is a fair representation of all affected parties involved in the decision-making process [Leventhal, 1980]. Actually, the rules of the three VQRs defined generic criteria or thresholds for panel composition in terms of the observable characteristics of scholars: gender and affiliation, and research fields. As for gender, the first two VQRs requires a fair gender distribution, and VQR 2010-2014 fixed a threshold of about 33.3%

for women in the panel. As for affiliation, VQR 2004-2010 formally required a presence of 20% of foreign scholars; VQR 2010-2014 a "significant percentage" without any explicit threshold; VQR 2015-2019 required 5%. For Italian scholars, a not better specified fair distribution of panellists among universities and institutions is generally required in all the three VQRs. In the final reports of the assessments, ANVUR claimed that gender and affiliation criteria were met, presenting some data on the whole set of panellists, but no evidence was provided for each area panel [ANVUR, 2013, 2017, 2022]. Hence, it may be that in some panels the criteria of fair composition in terms of gender and affiliations were not respected.

As for research fields, all three VQRs required the coverage of the research fields inside each Area, without any specific indication or thresholds. For the VQR 2004-2010, it was required that panels "cover all the cultural and research lines within the areas" [ANVUR, 2011]; in the VQR 2010-2014 and VQR 2015-2019 it was requested the "coverage of the scientific-disciplinary sectors (SSD)" [ANVUR, 2015]. No specific indication or thresholds were defined. The final reports of VQR 2004-2010 and VQR 2010-2014 did not present data about the respect of the coverage of the research fields inside each area [ANVUR, 2011, 2015]; for the VQR 2015-2019 uncommented raw data about panel composition in terms of scientific-disciplinary sectors were published (https://www.anvur.it/attivita/vqr/vqr-2015-2019/gev/).

Moreover, the respect of fairness criteria about observable characteristics of panellists such as gender, affiliations and research field did not guarantee that panels had a fair composition in terms of diverse viewpoints, scientific perspectives and scholarly thinking.

Indeed, the fairness of the composition of panels for the VQR 2004-2010 was questioned from its inception, by highlighting the lack of transparency in the members appointing procedures [Baccini, 2011]. The attention was especially focused on the panel for economics, statistics and business. This panel was composed not only by a small minority of women (16.7%), but its appointed members were closely linked to each other by co-authorship relationships. In the economics sub-panel, 9 panellists out of 20 (45%) were among the founders of a ultra-liberal party (Fare per fermare il declino) that participated in the political elections of 2013 by obtaining the 0.9% of the votes and no representatives in parliament [Baccini, 2018]. The same problem was also documented for the VQR 2011-2014 during a conference organised at the Italian Parliament: the panel of economics was again largely composed by scholars of the same ultra-liberal party [Baccini, 2016a]. In the public debate it was noted that the absence of fairness in the composition of the panel for economics, statistics and business had probably minimized the voices of dissent with respect to the evaluation methods and rules adopted by the panel [Baccini, 2011, Re, 2019]. It was noted also that there was a relevant precedent to consider: in the first experimental research assessment for the years 2001-2003, managed by CIVR (Steering Committee for Research Evaluation), Luigi Pasinetti, one of the panel member for economics, wrote a note of dissent documenting the absence of pluralism in evaluation [Pasinetti, 2006].

4 Research design

The main aim of the paper is to analyze whether the selection procedure of the panellist for the Italian research assessments gave rise to a composition that fairly represents the intellectual diversity of the research community of economics, statistics and business. The analysis of the composition of panels in terms of gender or affiliations is important, but it is not sufficient. Indeed, a fair composition in terms of gender and affiliations does not guarantee that it is also respected a fair intellectual composition of the panels in terms of diverse viewpoints, scientific perspectives and scholarly thinking heterogeneity.

Unfortunately, the analysis of the intellectual composition of a panel is a very complex task for at least two reasons. First of all, it is difficult to classify panel members according to their intellectual perspective. Even if this classification is finally accomplished, it is difficult to evaluate if the panel composition reflects the intellectual diversity of the scholarly community at large. Instead of adopting a qualitative analysis of

intellectual composition of panels and of scholarly communities of economists and statistics, this paper adopts a network analysis perspective. The basic idea is to explore the intellectual and social compositions of the panels by observing connections between members.

The interesting connections between scholars regard theoretical approaches, personal knowledge and economic-political visions. They are observed by building three different networks: a co-authorship network, a journal based network and an "affinity network".

A *co-authorship network* permits to observe collaborations among scholars: in it two nodes representing two scholars are linked by a weighted edge if they have authored at least a paper together; the weight of the edge is proportional to the number of co-authored papers. More precisely, we adopted a so called *ego*-coauthorship network approach (for a review see Arnaboldi et al. [2016]). We started from the lists of panelists; the co-authorship network of a given panelist is defined as the weighted network formed by her/him and all her/his co-authors; each edge is weighted with a measure of the strength of the collaboration, usually, again, the number of co-authored paper. This approach is adopted when one is interested, as we are, in studying direct collaborations between a set of scholars, since it allows one to identify only (i) direct collaborations between members of the starting set of authors, in our case panels, and (ii) ties generated by two members of the starting set collaborating with a common co-authors.

Collaboration in research is a complex social phenomenon that has been systematically studied since the 1960s and co-authorship is the most tangible and well documented forms of scientific collaboration (for a review see Kumar [2015]). Co-authorship is considered a reliable proxy of research collaborations because co-authors cannot write a paper together unless some degrees of personal acquaintance exists between them.

However, there are many scholars who know one another or who are intellectually similar to some degree but have never collaborated by writing a paper. For this reason, to detect hidden connections, it is possible to look also for similar specialization, similar training and similar affiliations. A proxy of intellectual affinity among scholars can be obtained by observing the network of journals where scholar published their papers.

In the *network based on journals* two scholars (nodes) are considered connected if they have published in the same journal (edge). The relation between social and intellectual community gathered around economics and statistical journals is largely documented [Baccini et al., 2020b, 2022]. Our starting hypothesis is therefore that journals represent different schools or methodology approaches and may cover different specializations. Hence, if two scholars have published in the same journals, it can be conjectured that there is some sort of intellectual or theoretical similarity between them.

Finally, the '*affinity network*' is here defined as a generalized affiliation network where two scholars (nodes) are connected if they are affiliated in the same university or in the same research centre, or if they studied at the same institution, or if they contributed to the same newspaper or blog (edge). This network is based on the hypothesis that, even if scholars have not published together or in the same journals, they can have a common set of relational patterns which can reflect theoretical or political similarities: not all economics departments and research centres carry out the same theoretical vision, and analogously magazines, newspapers and blogs have different editorial lines.

These three networks will be analyzed for exploring the connections between the panellists, by considering appropriate quantitative indicators. In every case, in absence of a reasonable benchmark for the considered indicators, it is difficult to conjecture about the strength of the connections between members that corresponds to an unfair composition of the panel.

In the absence of such a benchmark, a suitable control group representative of the research community of economics, statistics and business could be used for evaluating the fairness of the composition of panels. Theoretically, the best strategy would have been to build a control group for each of the three panels. By

comparing the existing connections among members of each panel with those in the corresponding control group, it would be possible to assess whether or not the heterogeneity present in the research community was fairly represented in the panel. Unfortunately, the building of such control groups is very difficult, both theoretically and practically.

Indeed, as saw above, the composition of panels was the result of different two-step procedures. For the three panels the first step consisted in the definition of a list of eligible candidates through a worldwide public call to serve on the panels. As a consequence, the eligible scholars were self-selected from the world population of scholars. Scholars demanding for being part of panels were scholars who self-evaluate themselves as "prominent scholars", and who accepted to work for a governmental agency in an administrative process. Moreover, foreign scholars were probably solicited to partecipate to the call by ANVUR or by other Italian scholars: as we documented in the Supplementary materials A1 the big majority of scholars with non-Italian affiliation have Italian first and last name.

The second step of the procedure of appointment was instead different for the three VQRs. As anticipated, ANVUR governing board directly selected panel members for VQR 2004-2010 and VQR 2011-2014; in both cases ANVUR governing board appointed members by choosing them from the list of candidates, but also off the list in a number of cases that have not been disclosed. For the VQR 2015-2019 members were selected by lottery exclusively among the list of candidates.

Hence we can affirm that in VQR 2004-2010 and VQR 2011-2014 the final composition of panels is due in part to self-selection of scholars and in part to ANVUR choices, while in 2015-2019 the final composition is due only to self-selection.

From a practical point of view, the building of control groups for each of the three panels would require (i) random selection from all scholars active worldwide in economics, statistics and business in the year of panel appointment; (ii) the respect of generic criteria – as described in Supplementary materials A1 –for the panel composition in terms of scientific production, gender, affiliations and geography. This strategy was practically unfeasible, given the huge amount of necessary information. If feasible, the adoption of this kind of control groups would give raise to different results for the three VQRs. Indeed, for VQR 2015-2019 an unfair composition would be due to self-selection mechanism. For VQR 2004-2010 and VQR 2011-2014 an unfair composition would be due to self-selection mechanism or to ANVUR choices or a combination of the two.

So, the strategy adopted in this paper consists in considering the panel for the VQR 2015-2019 as a sort of control group for the two preceding VQRs.

We started by hypothesizing that there are no reasons to think that self-selection of candidates in the three VQRs happened in different ways. A rather rigourous test of this hypotesis could have consisted in building random panels from those who had applied for being member of the panels VQR 2004-2010 and VQR 2011-2014. A comparison of these two random panels with the one of 2015-2019 would have permitted to verify if self-selection operated differently in the three VQRs. Unluckily, the lists of candidates VQR 2004-2010 and VQR 2011-2014 were not publicly available.

If self-selection of candidates in the three VQRs happened in similar ways, self-selection did not generated *per se* a different degree of unfairness in the composition of the three panels. Since in VQR 2015-2019 ANVUR governing board had no role in selecting panel members, a comparison of the composition of the panels of VQR 2004-2010 and VQR 2011-2014 with that of VQR 2015-2019 should allow us to isolate the effect of ANVUR governing board choices in the appointment procedures.

More explicitly, the panels for the VQR 2004-2010 and VQR 2011-2014 are considered as "treatment groups" and they are compared with the control group represented by the panel for the VQR 2015-2019.

Differences in the structural properties of the three networks (co-authorship, journal based and affinity) between treatment groups and the control group are considered as indicative of unknown systematic biases in the choice of panellists introduced by ANVUR governing board choices. These biases might result in a more unfair composition of the two "treated panels" that the third.

In any case, the three groups are selected in different periods of time and the results of previous research assessments had probably an effect in the last one. For example, a bad evaluation of certain schools of thought could have led to a reduction of their funds and of scholars of these schools, as documented by Lee et al. [2013] for the United Kingdom. Thus, it cannot be excluded that the level of heterogeneity in economics, statistics and business communities in 2011, when the panellists of VQR 2004-2010 were selected, was higher than the one in 2020, when the selection for the VQR 2015-2019 happened. This consideration would reinforce our research strategy, since the networks for the control group should be *naturally* more concentrated than the ones for the VQR 2004-2010 and VQR 2011-2014.

The network analysis was conducted by using Pajek version 5.14 [De Nooy et al., 2018]; visualizations are realized by VOSviewer version 1.6.15 [Van Eck and Waltman, 2020]. Data for replicating results are available here: 10.5281/zenodo.7244943.

5 The co-authorship networks

Three *ego* co-authorship networks are built by considering the set of publications of all the members of each panel. In each network nodes are panel members and scholars who co-authored at least a paper with a panel member. Ties between nodes indicate direct collaboration between a pair of panelists, or collaboration of panel members with external scholars. Each network therefore includes direct co-authorships between pairs of panelists and indirect collaborations realized by writing a paper with a common co-author. Ties are weighted with the number of co-authored papers. The comparison of the structures of the three networks may permit to conjecture about the presence of diverse viewpoints, scientific perspectives and finally of scholarly thinking heterogeneity.

For observing collaborations among members before they met on the panel, their publications were retrieved from Scopus for the 25 years preceding the starting date of the research assessment exercises.

The first observation is about the degree of overlapping of the three panels. 5 members of the panel 2004-2010 were appointed also as members of the panel 2011-2014 (Bartolucci F., Bertocchi G., Gambardella A., Ronchetti E., Schivardi F.). In particular the president of the panel 2011-2014 was also one of the member of the panel 2004-2010. In the co-authorship network 2011-2014 there are 10 other scholars that were also members of the panel 2004-2010 (Canova, Dardanoni, Dosi, Ellul, Frittelli, Jappelli, Peracchi, Rossi, Weber, Zamagni). In the panel 2015-2019 only one member was also a panelist in the panel 2011-2014 and one of the nodes of the co-authorship network of the panel 2004-2010 (Pagano).

Table 1 compares some structural statistics for the three coauthorship networks.

The networks appear as slightly dissimilar in terms of number of nodes (authors) and links (coauthorships). In particular, it appears that the sets of authors gathered around the first two panels, especially the first, are smaller than the one gathered around the third, also in terms of co-authors per panel member. The third panel has the greatest set of authors, but they produced only a lightly smaller number of papers than the other two panels. In the other two panels, especially the first, a smaller set of authors produced more papers: hence the degree of overlapping co-authorship is greater in the first two panels than in the third. Figure 1 reports the distribution of panelists according to the number of authors articles and number of coauthors.

In the first panel the degree centralization is also lower than the betweenness centralization since some nodes have zero betweenness centrality but they are still very close to the centre of the network. Thus, the

	Panel 2004-2010	**Panel 2011-2014**	**Panel 2015-2019**
N. of panel members	36	31	40 (37)*
Publication years	1987-2011	1991-2015	1996-2020
N. of co-authors of panel members	781	922	1,271
N. of co-authors per panel member	21.69	29.74	34.
N. of papers	1,190	1,188	1,079
N. of papers per panel member	33.06	38.32	29.2
Number of edges between authors	1,801	2,829	5,678
Minimum weight of an edge	1	1	1
Maximum weight of an edge	47	142	31
Number of edges with weight equal to 1	1,315	2,178	4,304
Number of edges with weight equal to 2	251	337	880
Number of edges with value greater than 2	235	314	494
Density	0.005	0.006	0.007
Average Degree	4.612	6.136	8.934
Number of components	12	17	25
Percentage of realized components over maximum	33.3	54.8	62.5

* Three panellists were dropped from the analysis because no information was retrieved from Scopus (Cori Enrico, De Vincentiis Paola and Pisoni Pietro Maria).

Table 1: Basic statistics of the co-authorship networks.

first coauthor network has more central actors than the others: few vertices are needed to create the network because they connect other vertices to each other. The second network is in an intermediate position while the third has nodes with similar values of centrality, by indicating the presence of localized and separated clusters of coauthors with no central vertices that contribute to connect the network.

Figure 2 and Tables A1, A2 and A3 report the degree distributions of the co-authorship networks, where the degree indicates the number of co-authors of each scholar. The three distributions are all rightly skewed, but they are statistically different, according to Kolmogorov-Smirnov test (first panel-second panel: $D = 0.125 (p = 3.48e - 06)$; first panel-third panel: $D = 0.31 (p = 0.00e + 00)$; second panel-third panel: $D = 0.213 (p = 0.000e + 00)$. *p-value adjusted for multiple comparison*). The coauthorship network for the third panel appears as radically different in terms of degree distribution: it has the lowest share of scholars with 1 or 2 couthorships (32% against 41% of the panel 2004-2010 and 54% of the panel 2011-2014) and the biggest share of scholars with more than 22 co-authors (11.64% against 1.41% of the panel 2004-2010 and 4.37% of the panel 2011-2014). The maximum degree in the three panels is reached respectively by Dosi with 78 couthorships, Alfò with 128 and Stingo with 259.

The most central scholars can be identified by computing their betweenness centrality [Wasserman and Faust, 1994]. In this case, the distributions of betwenness centralities in the three panels are not statistically different, but their comparison gives some qualitative indications. First of all, the maximum value of betweenness is 0.25 for the first panel, it lowers to 0.099 for the second and goes down to 0.045 in third panel. The share of nodes with zero betweenness centrality, i.e. nodes that are coauthors of only one member of the panel, is similar in the three networks (respectively 615 nodes (78.75%); 707 nodes (76.68%) and 971 nodes (76.39%)), but the first and the second networks have higher share of nodes with relatively high value of betweenness. If we adopt a betweenness threshold of 0.002, in the third panel only 16 scholars (1.26%) have a betweenness higher than the threshold, against the 50 scholars (6.40%) of the first panel and the 31 scholars (3.36%) of the second panel. The list of these most central scholars are reported in the supplementary

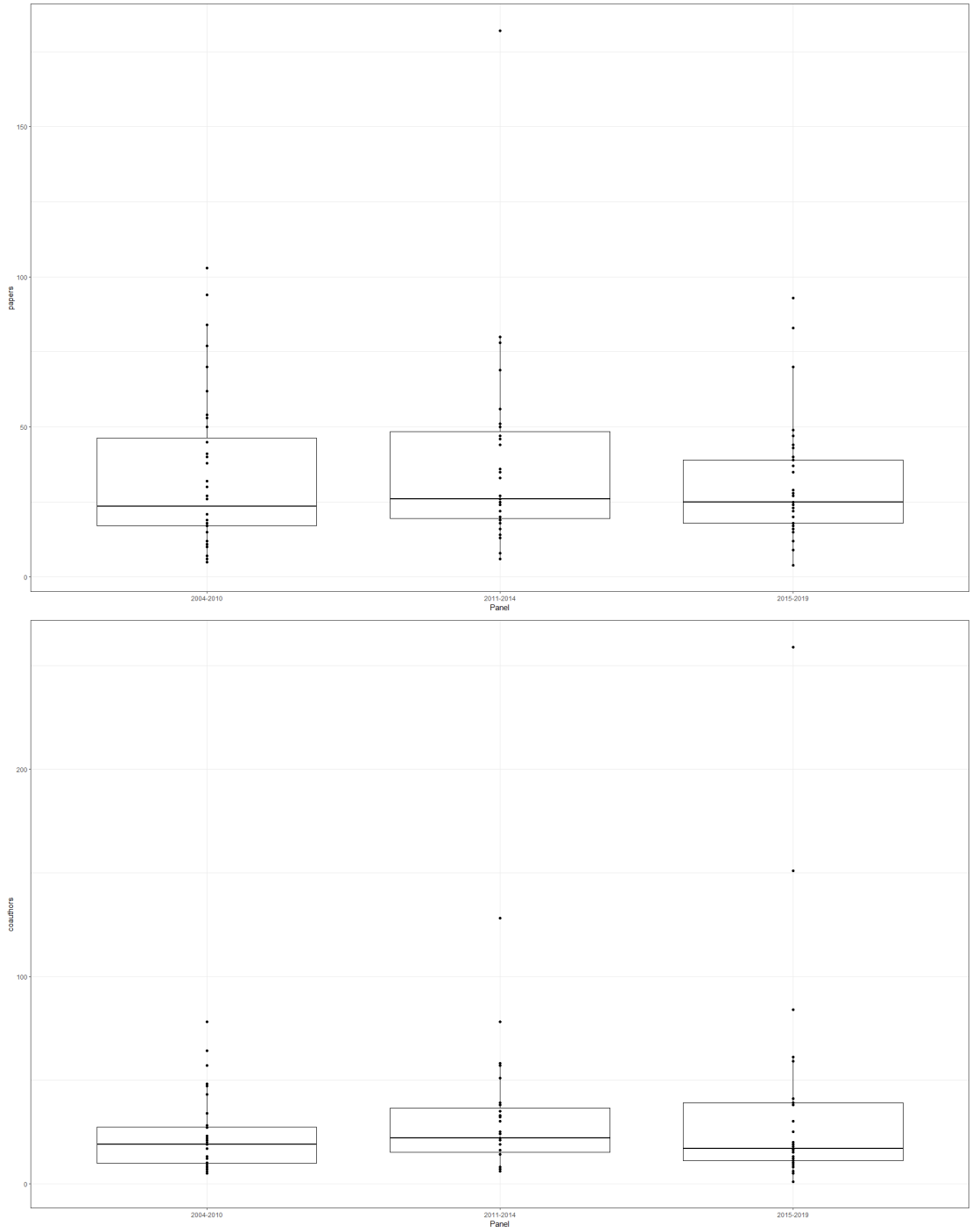


Figure 1: Distribution of panelists according to the number of authored articles and coauthors.

materials Table A4, A5 and A6. Again, 10 scholars appears in both the lists of the most central scholars of the first two panels (Bartolucci, Bertocchi, Dosi, Gambardella, Guiso, Jappelli, Lippi, Pagano, Peracchi, Schivardi). Only one of these scholars is also among the most important scholars of the panel 2015-2019 (Pagano).

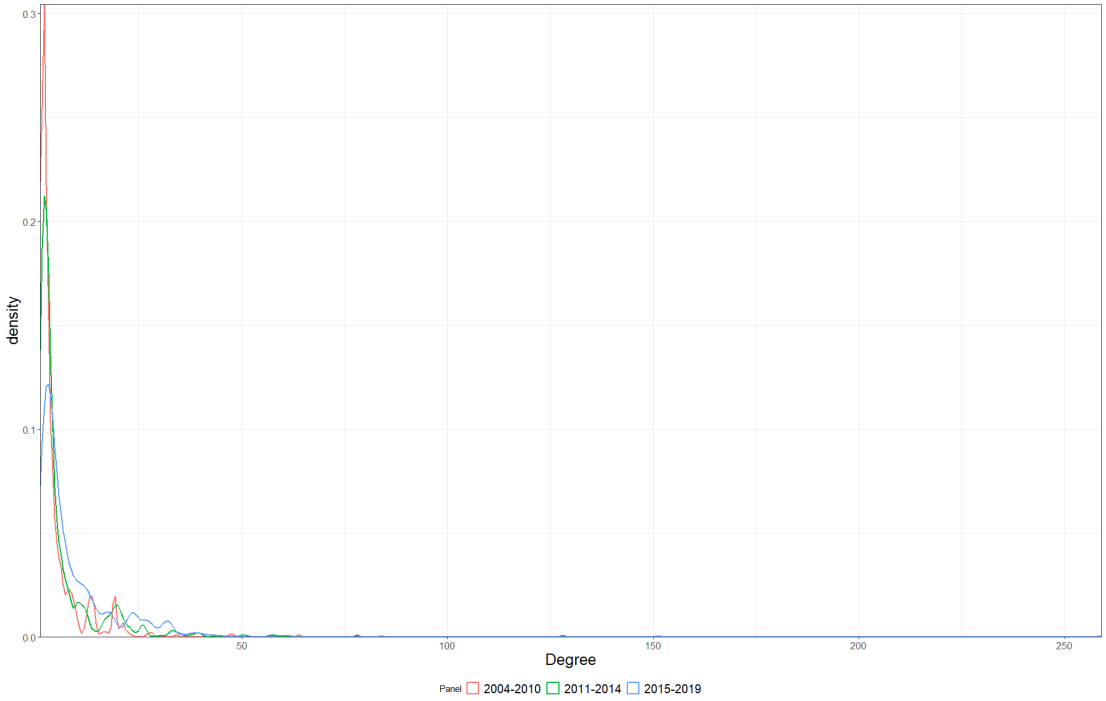


Figure 2: The distributions of degrees in the three coauthorship networks.

Figure 3 shows that each of the three networks is disconnected, i.e. not all nodes can be reached by a finite path from every other node [Newman, 2018]. The three disconnected networks are composed by components, i.e. connected subnetworks where there is a path between each pair of nodes. Obviously no edge exist between nodes of two different components [Newman, 2018]. It is therefore straightforward to interpret the components of each network as communities with relatively strong co-authorship relations. Please note that the size of networks and the fact that they are naturally partitioned in components, makes it unnecessary to resort to community detection algorithms [Newman, 2018].

In the three *ego* co-authorship networks, the maximum level of fragmentation is reached when the number of components is equal to the number n of panellists. In this configuration each component of the network includes a member of the panel and all her/his co-authors; there is no co-authorship relationship between scholars from different components of the network. The ratio between the actual numbers of components of a network and the maximum can be considered as an indicator of the fragmentation of the panel, with values in the range $[1/n, 1]$. If the two co-authorship networks for the panels of VQR 2004-2010 and of VQR 2011-2014 have less or larger clusters than the third one, this can be considered as a clue of an unfair composition.

The last two rows of Table 1 indicates that the first panel has the lowest number of components (12), while the third panel has the highest (25); moreover, the panel 2015-2019 is the most fragmented with a ratio of actual components over the maximum of 0.625 against 0.333 and 0.548 of, respectively, the first and the second panel.

The analysis of the frequency distribution of the components in the panel 2004-2010, reported in the appendix A7 shows the presence of a big component containing 512 nodes, i.e. more than 65% of the nodes of the network, and 24 panel members out of 36 (66%). 10 other components gather each one panel member and the set of her/his coauthors; only one component contains two panel members and their coauthors. The graph of the whole co-authorship network is reported in Fig.3a, while in Fig.A1 in the supplementary materials the

biggest component is represented. These figures show clearly the central role that some scholars, who are not panel members, play in the construction of the network.

As anticipated, the 2011-2014 co-authorship network is a bit more fragmented than the first one. Table A8 reports the frequency distribution of the components. Also in this case, the network is characterized by only one big component with 403 nodes, i.e. more than 43% of the scholars in the network. In this big component, there are 13 panel members representing the 42% of the panellists. 17 components gather each one panel member and the set of her/his coauthors; only one component contains three panel members and their coauthors. The graph of the whole co-authorship network is reported in Fig.3b, while Fig.A2 in the supplementary materials draws the biggest component. Even in this case, the central role of non-panellists in the structure of the network is clearly visible.

As we have seen, the 2015-2019 co-authorship network is the most fragmented. Table A9 reports the frequency distribution of the components. In this case, the largest big components contains only 324 nodes, i.e. the 25% of the network, and only 5 panellists out of 40 (13%). In the network there is another big component that contains 260 nodes (20,4%), but only 1 panel member (Stingo F.), who has an exceptional number of co-authorship connections (259). There are also 6 components gathered around 2 or 3 panel members, and with variable size. The graph of the co-authorship network reported in Fig.3c shows a relatively high fragmented network.

	Panel 2004-2010	**Panel 2011-2014**	**Panel 2015-2019**
N. of panel members	22 (66.6%)	13 (41.9%)	5 (13.5%)
N. of co-authors of panel members	512 (65.5%)	403 (43.7%)	324 (25.4%)
Number of edges between authors	1257	1452	1631
Minimum weight of edges	1	1	1
Maximum weight of edges	47	19	23
Number of edges with weight equal to 1	963	1221	1156
Number of edges with weight equal to 2	169	140	404
Number of edges with weight greater than 2	125	91	71
Density	0.009	0.017	0.031
Average Degree	4.910	7.205	10.067

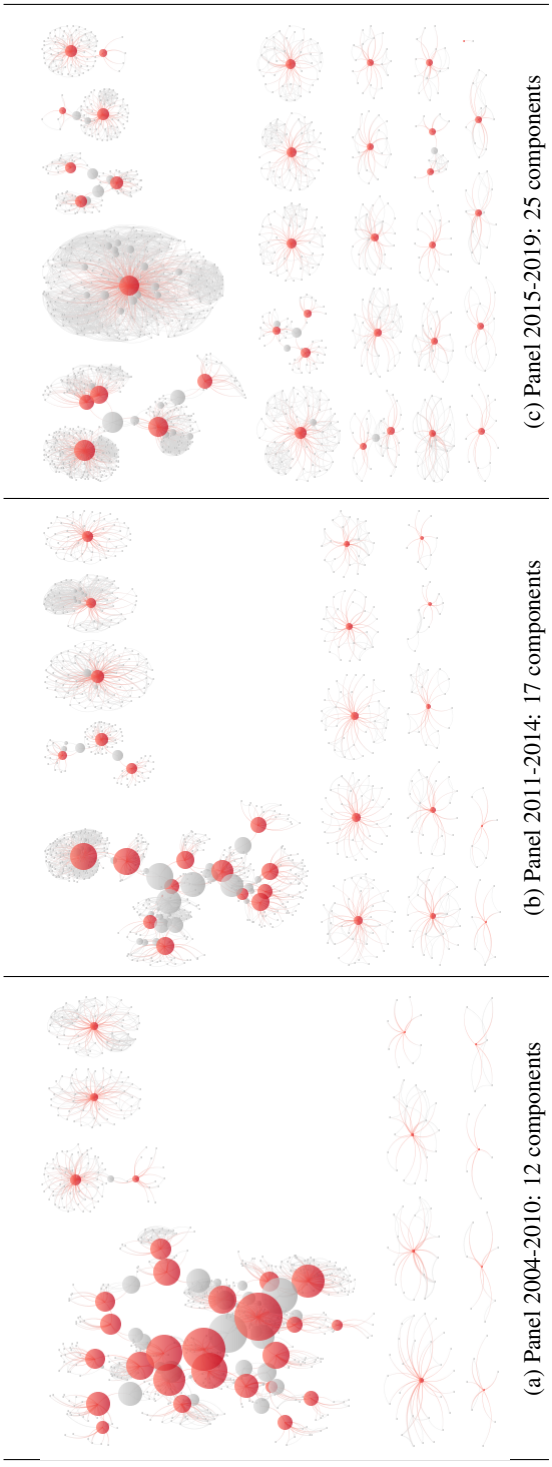
Table 2: Basic statistics of the largest component of each co-authorship networks

Table 2 compares some characteristics of the biggest components of the three panels.

In sum, the comparison of the three co-authorship networks shows that the first two panels are structurally different from the third. Especially in the first one, few central actors are mutually connected in one single large cluster that contains more than half of the whole network and of panellists. The second network has similar characteristic and shares with the first one a same group of central nodes. Instead, the third panel has few members in common with the others and no nodes show particular high centrality values; it is more fragmented in smaller components, more centralized than the ones found in the others.

6 The journal based networks

The analysis of the journal based networks aims to investigate whether the panel members have published in the same or in different journals. If two authors have published articles in the same set of journals, then they are in some sense similar in terms of topics or theoretical approaches or methodologies. Also in this case the



(a) Panel 2004-2010: 12 components

(b) Panel 2011-2014: 17 components

(c) Panel 2015-2019: 25 components

Figure 3: Components in co-authorship networks. Size of vertices is proportional to betweenness centrality and red nodes are panel members. An edge between two nodes indicates that the two scholars wrote together at least one paper. Thickness of edges is proportional to the number of common papers.

analysis consists in comparing the networks of the panels 2004-2010 and 2011-2014 with the control group represented by the panel 2015-2019. If the sets of journals of the first two panels are narrower than the set of the control group, this may be considered a clue indicating that ANVUR's selection introduced unfairness into the panel composition.

The dataset for journal based networks is the same used for the co-authorship networks. For each panel, the starting point is a bipartite network where panel members are linked to the journals where they have published at least a paper. For each panel, the one mode projection is then derived, where nodes are the panel members and the edge between two panel members is weighted according to the number of journals where they both published at least a paper.

Table 3 reports the basic statistics of the three bipartite networks. Despite the three networks are similar in terms of panel members, the sets of journal is growing from the first panel (360 journals), to the second (467), to the third (566). In particular the first panel has the lowest number of journals per panellist. The members of the third panel on average published their work in a greater sets of publishing outlets than the first two.

	Panel 2004-2010	**Panel 2011-2014**	**Panel 2015-2019**
N. of journals	360	467	566
N. of panel members	36	31	37
Number of edges	511	675	721
Number of journals per panel member	10	15.1	15.3
Density [2-Mode]	0.040	0.046	0.034

Table 3: Basic statistics of the journal based networks

As a consequence, in the one-mode projection networks of scholars density and average degree tend to be lower in the third panel than in the first two, as it is reported in Table 4.

	Panel 2004-2010	**Panel 2011-2014**	**Panel 2015-2019**
N. of panel members	36	31	37
Number of edges	148	137	137
Lowest weight of an edge	1	1	1
Highest weight of an edge	8	13	14
Number of edges with weight equal to 1	86	59	86
Number of edges with weight equal to 2	32	18	35
Number of edges with weight greater than 2	30	60	16
Density	0.228	0.285	0.200
Average Degree	8.222	8.838	7.405
Islands	3	2	3
Number of off-island panellists	9	11	22
Percentage of off-island panellists	25.0%	35.5%	59,5%

Table 4: Journal-based networks: basic statistics of the one-mode projection networks of scholars.

Tables A10, A11 and A12 report the degree distributions of the one-mode networks of scholars for each panel, where the degree of a panel member is the number of other panellists who published at least a paper in at least a journal where he or she also published. The maximum degree in the first panel was reached by T. Jappelli (the president of the panel) and A. Bisin linked to other 17 members for having written in same

journals; by L. Sarno in the second panel linked to 19 panellists (while the president of the panel G. Bertocchi has the second highest degree (17)); by M. Piva in the third, with a degree of 22.

The search for clusters of panellists in the one-mode projection networks is conducted by using a simple edge-cut technique based on edge weights. The algorithm used permit to individuate "islands" i.e. maximal subnetworks of nodes connected directly or indirectly by edges with a value greater than the edges to nodes outside the subnetwork [De Nooy et al., 2018]. For interpreting the results, it is useful to consider that a panel with the maximum level of diversity is a panel where none of the panel members belongs to an island, i.e. panel members tend to publish their articles in different journals and the publication of articles in a same journal is not systematic or a rare event. A simple indicator of intellectual diversity of the panel is therefore the share of panel members who are not clustered in an island; this share tends to 100% for a panel with a maximum diversity.

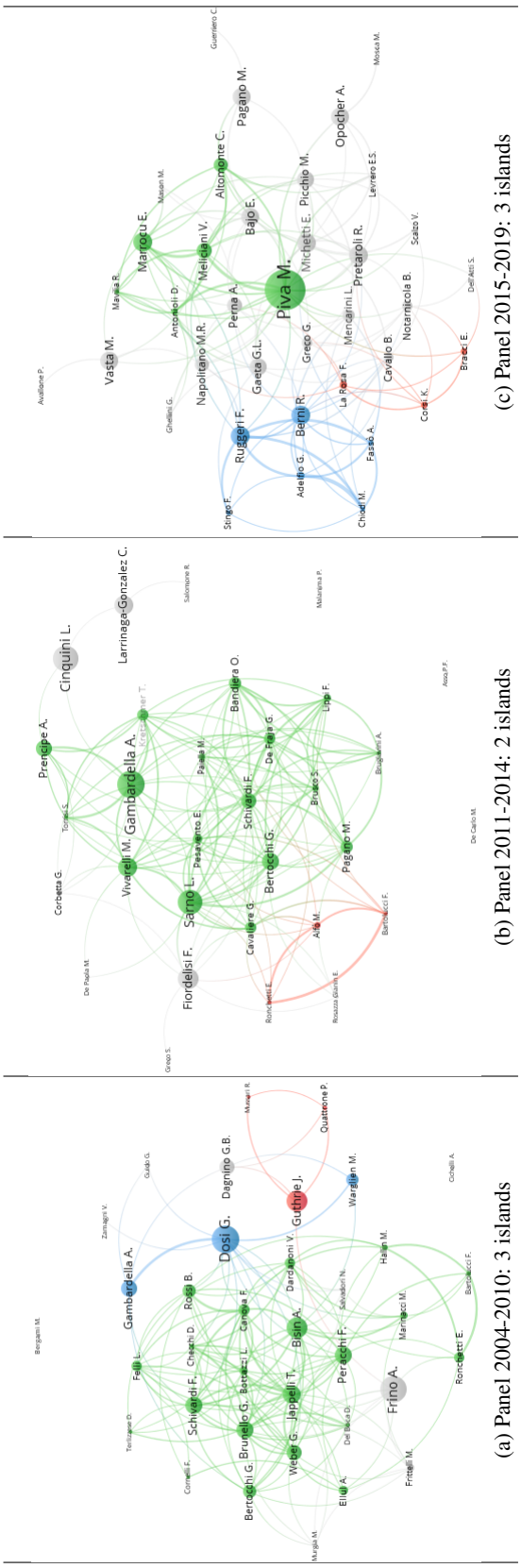
The search for islands (of minimum size 1 and maximum size of 3/5 of the number of panellists, respectively 21, 18 and 22) in the three networks individuated a similar numbers of clusters (3, 2 and 3), but with different configuration. Indeed, in the third panel about 60% of members are not part of any island, against about 42% of the first panel and 45% of the second. But the main difference emerge when the largest island of each network is analyzed, as reported in Table 5. In the third panel the largest island gathers only 6 panellists, i.e. a 16.2% of the total members, while in the first and second panel the biggest islands gather respectively 21 members, representing the 58.3% of the panel, and 17 members for a 54.8%. Moreover, the largest island of the first panel is less centralized with respect to the other, by showing that in this first panel all the members have similar roles in structuring the island, while, especially in the third, there is a prevailing role of a very small set of nodes. In Tables A13, A14 and A15 are reported the name of the scholars, the belonging island and the betweenness values with the connected ranking for each panel.

	Panel 2004-2010	**Panel 2011-2014**	**Panel 2015-2019**
N. of panel members	21	17	6
Percentage of totale panel members	58.3%	54.8%	16.2%
Number of edges	58	25	8
Lowest weight of an edge	2	5	3
Highest weight of an edge	7	10	14
Number of edges with weight equal to 1	0	0	0
Number of edges with weight equal to 2	30	0	0
Number of edges with weight greater than 2	28	25	8
Density1 [loops allowed]	0.263	0.173	0.444
Average Degree	5.523	2.941	2.666

Table 5: Basic statistics of the largest island in the one-mode projection networks of scholars.

As visualized in Figure 4(a), only 2 members of the 2004-2010 panel are isolated (Cicchelli and Bergami), i.e. they wrote in journals where none of the other panellists wrote. The biggest green island is composed by nodes sharing a relatively high betweenness centrality, by confirming that the big part of panellists contribute similarly to the network structure. Bisin has the highest betweenness centrality followed by the president of the panel Jappelli; only three scholars of the green island have betweenness centrality lower than 0.002 (Cornelli, Del Boca and Terlizzese), by occupying a peripheral role in the network.

The blue island is composed by only three nodes with relatively high betweenness centrality. The red island also hosts only three scholar, but only one of them has a relatively high betweenness centrality. Among the scholars that are not part of an island, 6 have zero or almost zero betweenness centrality; Frino has the



(a) Panel 2004-2010: 3 islands

(b) Panel 2011-2014: 2 islands

(c) Panel 2015-2019: 3 islands

Figure 4: The clusters of scholars in the islands of the journal based networks. Each node is a panel member. Size of nodes is proportional to their betweenness centrality. An edge between two nodes indicates that the two members wrote papers in at least a same journal. Thickness of the links is proportional to the number of common journals. The color of nodes indicates different islands. Grey nodes are not part of any island.

second largest value, i.e. he wrote in many different journals in which other panellists wrote, but the number of journals shared with others is not high enough to attract him to an island.

The focus on the connection between panel members through journals unveiled a different centrality ranking and different clusters of people with respect to the co-authorship network. In particular, Ronchetti, Bisin, Bottazzi, Rossi, who appeared isolated in the co-authorship network, are now part of the green island that gathers most of the panel. This indicates, for example, that Bisin, while not collaborating directly with the other group members, wrote in the same journals as most of the other participants, working on similar topics or adopting similar theoretical approaches or methodologies. In contrast, the panellists of the red island are completely integrated in the biggest component of the co-authorship network. This may indicate that, for example, Dosi did not share many journals with other members of the panel, but it is linked by direct (co-authorship) or indirect (common co-authors) collaborations with most of the panel.

In the 2011-2014 panel, reported in Figure 4(b), 3 panellists appear as isolated (Asso, De Carlo, and Malanima). Also in this network, the biggest green island is composed by nodes with similar and relatively high betweenness centrality, by indicating that the role in structuring the network is similar for most of them. The green island gathers 17 scholars, i.e. about 55% of the members. Gambardella is the scholar with the highest betweenness centrality; other two scholars have values of betweennes centrality higher than the president of the panel (Bertocchi). The red island gathers three panellists with relatively low betweenness centrality and relatively peripheral position in the network. Among the 11 scholars that are not part of an island, 7 have a zero or almost zero betwenness centrality, while Cinquini has the second highest value of the network. Also in this second panel, the focus on the connection between panel members through journals reveals clustering structure only slightly different with respect to the co-authorship network. In particular, with the only exceptions of Alfò and Bartolucci, all the panellists belonging to biggest component of the co-authorship network are now part of the green island. Six panellists (Sarno, Vivarelli, Cavaliere, De Fraja, Kretschmer, Pesavento) who were not in the biggest component of the coauthorship network are grouped together in the green island of journal-based network. Inversely, two of the three members of red island were part of the biggest component in the co-authorship network.

The main features of the 2015-2019 panel visualized in Figure 4(c) are the absence of a big island, and a majority of nodes (22 out of 37) that are no part of any islands. The analysis of the connections through journals shows few differences with respect to the co-authorship network. The green island includes six panellists who were grouped in two small components in the co-authorship network. The blue island gathers six scholars too: four were grouped in the biggest component of the co-authorship network and the other two were isolated (Berni and Stingo). Two of the three nodes of the red island (La Rosa and Corsi) were grouped together also in a component of the co-authorship network. 14 isolated scholars in the co-authorship network do not belong to any island.

In sum, the analysis of the journal-based networks shows that the first two panels are different from the third one. In the first two panels, a majority of members published on a relatively small set of journals, and only a small minority of them appears to have no systematic connections with others members. In contrast, the vast majority of members in the third panel did not wrote systematically in the same journals. The third panel appears as characterized by greater diversity in terms of publishing outlets than the first two panels.

7 The affinity networks

The analysis of the affinity networks aims to investigate whether panel members have studied or are affiliated with the same research centres and universities, whether they have published non-specialized articles in the same newspapers, magazines or blogs. If panel members have studied or are affiliated with the same research centres and universities, or published in the same magazines, newspapers and blogs, then they probably have

personal ties or theoretical or political affinities. A panel characterized by intellectual diversity has members that studied in different universities, with different affiliations, and writing in different magazines, newspapers and blogs. Also for the affinity networks, the analysis consists in comparing the networks of the three panels, by considering the third one as a control group.

The dataset is built by considering panel members and the most central nodes in each co-authorship network, i.e. the coauthors of panellists with a betweenness centrality value larger than 0.002. They are reported in Table A4, Table A5, and Table A6. This choice permits to pay attention to the ties between members and their coauthors. The basic idea is that a person, connected to people who in turn are not directly connected, can mediate with each other and profit from mediation [De Nooy et al., 2018]. In our case, people acting as bridges in connecting panellists in the co-authorship network, could be an additional element in reducing panel diversity: if these bridges come from the same universities or research centres or publish in the same newspapers, they could spread the same theoretical vision. Therefore, studying scholars acting as bridges is useful to detect further connections and possibilities for less representation of diferrent views in the panel.

The *curriculum vitae* of the panellists and of their more central co-authors were collected online from 17/02/2021 to 09/04/2021 and were manually processed to derive the following information: institutions where they graduated (maximum 2); institutions where they did MSc/MA and PhD (maximum 2); universities where they declared affiliations (maximum 2); declared affiliation to research centres (maximum 5); magazines, newspapers and blogs in which they wrote (maximum 5). Hereinafter all these entities are refereed to as "affiliated institutions".

The affinity networks are bipartite networks where panel members and their more central co-authors are linked to affiliated institutions. From each of these bipartite networks is possible to build two projections. In the projection network of scholars, two scholars are connected if they have at least a common affiliation, and the weight of their link is proportional to the number of their common affiliations. In the projection network of institutions, two institution are linked if they both have an affiliation by at least one same scholar, and the weight of their link is proportional to the number of common scholars.

Table 6 reports basic descriptive statistics of the three bipartite affinity networks. The panel 2004-2010 has the highest number of scholars, since it includes the biggest number of central coauthors of panel members. The basic statistics of the one-mode projection networks of scholars reported in Table 7 and of affiliated institutions reported in Table 8 reveal that the third panel is structurally different from the others.

	Panel 2004-2010	**Panel 2011-2014**	**Panel 2015-2019**
N. of scholars	58	40	44
N. of coauthors of panellists	22	9	4
N. of affiliated institutions	191	147	171
Number of edges	426	306	282
Number of scholars per institution	0.30	0.27	0.25
Number of institutions per scholar	3.29	3.67	3.88
Density [2-Mode]	0.038	0.052	0.037

Table 6: Basic statistics of the affinity networks

In particular for the projection networks of scholars (Figure 7), the third panel has lower average degree and density than the other. Moreover, the third panel has also systematically lower edge weights, i.e. the number of common affiliated institutions between pairs of scholars is systematically lower in the third panel than in the others.

Analogously, for the projection networks of affiliated institutions (Figure 8), the third panel appears as less connected: the number of edges is lower and very few institutions are linked by more than two affiliated scholars. As a consequence, in the third panel the average degree and density are lower than in the other two panels.

	Panel 2004-2010	**Panel 2011-2014**	**Panel 2015-2019**
Number of scholars	58	40	44
Number of edges	551	345	185
Lowest weight of edges	1	1	1
Highest weight of edges	7	7	5
Number of edges with weight equal to 1	272	172	152
Number of edges with weight equal to 2	129	95	24
Number of edges with weight greater than 2	150	78	9
Average Degree	19.000	17.250	8.409
Density	0.327	0.431	0.191
Number of islands	4	2	7
Largest number (and %) of important vertices in the same island	15 (88.2%)	14 (82.3%)	4 (23.5%)
Number of off-island scholars	36	23	22
Number of off-island important scholars	2	3	7

Table 7: Basic statistics of the network of scholars

	Panel 2004-2010	**Panel 2011-2014**	**Panel 2015-2019**
Number of affiliated institution	191	147	171
Number of edges	1277	944	889
Lowest weight of edges	1	1	1
Highest weight of edges	15	12	4
Number of edges with weight equal to 1	1083	798	845
Number of edges with weight equal to 2	124	97	37
Number of edges with weight greater than 2	70	49	7
Average Degree	13.371	12.843	10.397
Density	0.070	0.087	0.061
Number of islands	8	2	8
Largest number (and %) of important vertices in the same island	15 (88.2%)	8 (47.1%)	11 (64.7%)
Number of off-island institutions	153	131	131
Number of off-island important institutions	2	9	0

Table 8: Basic statistics of the network of affiliated institutions

A better understanding of the features of the bipartite affinity networks can be achieved by adopting a two-step analysis. In a first step, the sets of important scholars and affiliated institutions are searched by an algorithm based on eigenvector centrality [Newman, 2018, p. 159]: a scholar or an institution is important if it is linked to other important scholars or affiliated institutions in the network. The lists of the most important scholars and institutions in the three networks are reported in Tables A16, A17, A18. The second step consists in looking for clusters, separately, in the networks of scholars and in the networks of affiliated institutions through the island algorithm, by fixing the minimum size to 1 and the maximum size to 17. It is then possible

to observe how many of the 17 most important scholars/institutions are part of a same island. As for the networks based on journals, the maximum degree of diversity of the panel is when all scholars and all the affiliated institutions are disconnected. A similar test can be applied also to the most important scholars and institutions: the maximum degree of diversity is reached when all the most important scholars/institutions belong to different islands.

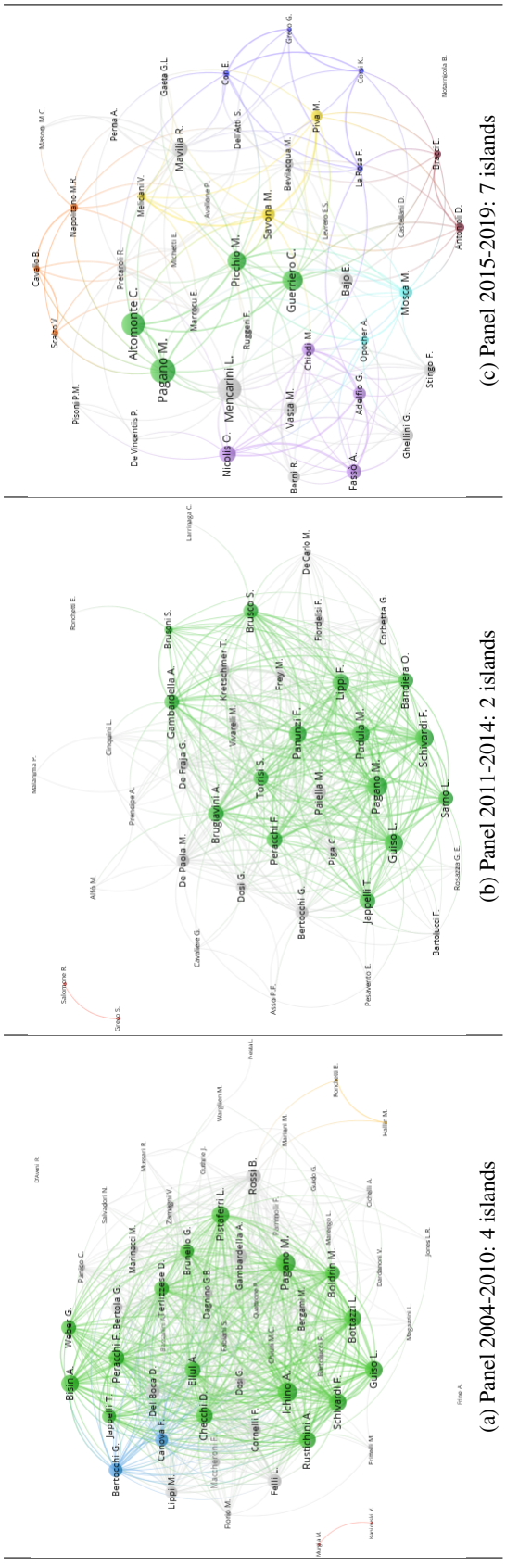
The basic statistics of the scholar networks are in Table 7; the number and dimensions of the clusters are reported in Tables A19, A20 and A21. The third panel appears as very different from the others. In the third panel, a big part of scholars (50%) and important scholars (41.2%) are not part of any island; moreover, the other are dispersed in 7 islands with a maximum of 4 scholars clustered together in a same island. The other two panels are characterized instead for having not only a smaller number of islands, but mainly for having a big central island gathering the big part of the important scholars. Moreover, in the first two networks co-authors of panelists play a very central role in structuring the affinity network, while in the third they have mainly a peripheral position.

In particular, in the 2004-2010 affinity network, the network is composed by 58 scholars: 36 panel members and 22 co-authors. Among the 17 most important scholars 6 are coauthors of panel members; 4 of these coauthors are among the top five most important scholars. In the network of scholars drawn in Figure 5(a), there are 36 scholars (62%) that are not part of an island; but only 2 important scholars out of 18 are not part of an island. The other scholars belong to 4 islands: three gather each only two scholars and none important node, while the biggest island gathers 16 scholars. Specifically, 15 out of 17 most important scholars are part of this biggest island, by showing that near all the most important scholars have a high degree of common affiliations.

The 2011-2014 network of scholars, drawn in Figure 5(b), is composed by 31 panel members and 9 co-authors. Among the 17 most important scholars, 5 are coauthors of panel members and 3 are in the top five. There are 23 scholars (57.5%), three of which important, that are not part of an island. The others belong to a pair of island. The smallest with only two not-important scholar; the biggest collecting 15 scholars, 14 of which important. Also in this second panel, a strong majority (82.3%) of the most important scholars have a high degree of common affiliations.

Finally, the network of scholars for the third panel is composed by 40 panel members and 4 coauthors. It is represented in Figure 5(c). Among the 17 most important scholars only 2 are co-authors of panel members and none of these is in the top five. 22 scholars (41%) are not part of any island; 7 important scholars out of 17 are not part of any island. The remaining 22 scholars are dispersed in 7 islands of maximum size 4 scholars. Three of these islands have zero important scholars, three have only one important scholar and two island of size 4 are populated by important scholars only. Moreover, among the important scholars, reported respectively in Tables A16 A17, A18, it appears that 6 scholars are among the important vertices of both the 2004-2010 and the 2011-2014 affinity networks, and among them the president of the first panel (Gambardella, Guiso, Jappelli, Pagano, Peracchi, Schivardi). Among these scholars, Guiso is the only one who is present in both but has never been a panel member. Only one panelist of the 2015-2019 panel (Pagano) appears also among the important vertices of scholars of the previous panels.

The projection network of affiliated institutions also appears structurally different for the third panel than in the others. The only common trait is that Bocconi University in Milan is the university with the highest eigenvector centrality. But while in the first two panels Bocconi University is the center of the island that concentrates most of the most important connections and scholars, in the third panel Bocconi's island is small in size and flanked by other islands of not much different size even in terms of important vertices, as it is shown in Tables A19 A20, A21. In particular, the 2004-2011 network of affiliated institutions drawn in Figure 6(a) shows that that 15 of the 17 most important institutions are clustered in the same island. As



(a) Panel 2004-2010: 4 islands

(b) Panel 2011-2014: 2 islands

(c) Panel 2015-2019: 7 islands

Figure 5: The clusters of scholars in the islands of the affinity networks. Each island has a different colour, and the dimension of vertices is proportional to eigenvector centrality in the affinity network. An edge between two nodes indicates that the two scholars have at least one common affiliation. Thickness of edges is proportional to the number of common affiliations.

reported in Tables A16, the other important vertices of the largest island gathers American economic policy think tank (CEPR and NBER), universities research centres (EIEF, strictly linked to the Bank of Italy), other non-academic institutions (Bank of Italy), newspaper (IlSole24ore, Il Foglio) and blogs (lavoce.info).

The 2011-2014 network of affiliated institutions, drawn in Figure 6b, is very similar to the previous one. As reported in Table A17, the green island collects only important institutions. Essentially, the green island is a concentrated replica of the biggest island detected for the 2004-2010 panel, bringing together only one Italian university: Bocconi and its IGIER research centre, the Bank of Italy and the research center EIEF it founded, the blog lavoce.info and the American economic policy think tanks CEPR and NBER. The red island collects only not important vertices.

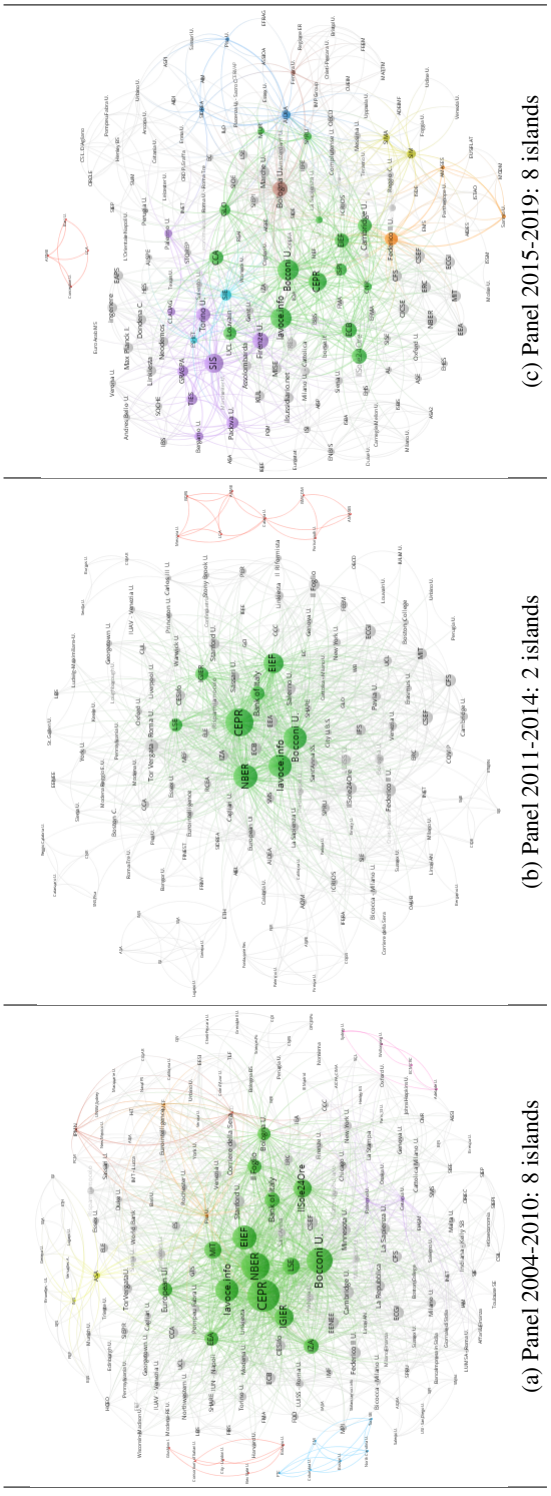
The control group is drawn in Figure 6(c). It is much more fragmented since the institutions form 8 islands: six of these islands, with a limited size of 2 to 4 nodes, have zero or only one important vertex. The biggest green island gathers 15 institutions, 11 of which are important institutions. The second biggest blue island contains 9 institutions, 4 of which are important. The list of important vertices of the biggest island (Table A18) is the same of the biggest islands of the two previous panel. However, only 4 important scholars are connected to this island. The second biggest island gathers three Italian universities and a scholarly society (the Italian statistical society).

In sum, if we consider the affinity network of the third panel as the benchmark for judging about the degree of diversity, the networks of the panel 2004-2010 and of the panel 2011-2014 appear as very far from the benchmark. Indeed, they are characterized by the prominent presence of a bulk of important scholars affiliated to a bulk of important affiliated institutions. These scholars and affiliated institutions are largely the same in the two networks. This is the result of the direct overlapping of scholars serving as panellists in both research assessment, and mainly of the presence in the first panel of members coauthoring papers with members of the second and viceversa.

8 Discussion and conclusions

Research evaluation and especially administrative massive research evaluation such as the British REF have gained a central role in university and research policies. In this respect research evaluation has taken "the function of gatekeeping, filtering, and legitimating knowledge" [Lamont and Huutoniemi, 2011]. Despite their procedural variety, evaluation is usually governed by panels of experts that managed the processes of peer review [Whitley et al., 2010]. Considering its importance, a great focus has been given to the fair representation in the composition of the panels of all affected parties involved in the decision-making process. This recommendation is present in the European Peer Review Guide of the European Science Foundation and it was formally adopted by the British REF and by the Italian research assessments. However, it is not easy to define who the affected parties are and what dimensions to take into account in defining their fair representation. The observable characteristics of scholars, such as research field, gender, affiliation and geography, seems not to be sufficient, in particular if a discipline is characterized by the coexistence of many schools of thought with different approaches, methodologies and policy recipes. This is precisely the issue pointed out for economics in the case of the research assessments in the United Kingdom and in Italy.

This paper proposes an empirical strategy for exploring if, despite the formal respect of the fair composition of panels in terms of easily observable characteristics such as gender and current university affiliation, it is possible to individuate hidden connections between members such that the panel composition could be considered as unfair. Three main type of connections are defined. The first is the direct or indirect collaboration of panellists in writing articles: a disproportionate diffusion of co-authorships among the members of a panel may indicate that its composition is restricted to people who have personal ties and possibly theoretical similarities. The second type of connections can be revealed by observing the set of journals where panellists



(a) Panel 2004-2010: 8 islands (b) Panel 2011-2014: 2 islands (c) Panel 2015-2019: 8 islands

Figure 6: The clusters or islands of affiliated institutions for the three panels. Each island has a different colour, grey nodes do not belong to any island and the size of vertices is proportional to eigenvector centrality in the affinity network. An edge between two nodes indicates that the two affiliations are connected at least by one scholars. Thickness of edges is proportional to the number of common scholars.

published: a concentration of panel publication in a relatively narrow set of journals can be considered as an indicator of the lack of intellectual diversity in the panel. The third type of connections can be revealed by observing if members have studied in a narrow set of institutions, if they had in the past or currently have affiliations in a narrow groups of institutions, if they contributed to a narrow set of magazines, newspapers and blogs. Also in this case the existence of hidden connections can be considered as an indicator of limited intellectual and social diversity in the panel.

Explorative network analysis is the natural tool for investigating these connections. The main difficult with this kind of approach is to define critical thresholds for collaborations, concentration of publications and narrowness of affiliated institutions.

The case studies presented in this paper regards the composition of the panels appointed to evaluate research in economics, statistics and business during three Italian research assessment exercises referring to the years 2004-2010, 2011-2014, 2015-2019. Three panels are then considered. In all the three research assessments, candidates to be appointed as panelists responded to a public call. The first two panels were appointed directly by ANVUR board members; the panel members were chosen, in undisclosed proportion, from among the candidates who responded to the public invitation and from outside this group. The third panel, on the other hand, was selected by lot and exclusively from among those who responded to the public call. As a consequence, the composition of the first two panels is the results of auto-selection of candidates responding to the call and of the deliberate choice of panellists made by ANVUR governing board; in the appointment of the third panel ANVUR governing board did not play a role. It is therefore possible to consider the third panel as a control group: if the first two panels appear structurally different from the third in terms of collaborations, concentration of publications and narrowness of affiliated institutions, it can be conjectured that all that is due to ANVUR governing board intervention. Indeed, this research strategy permits to overcome the problem of defining critical thresholds.

The results show that the members of the first two panels had connections in terms of co-authorship, common journals and affiliations much higher than the members of the control group. The first two panels appears filled with a prominent group of scholars linked by hidden connections. They appear as a single community of scholars with a shared intellectual ground, common social relations and also shared policy directions. As for policy orientation, it is not difficult to recognize in the affiliation networks people and institutions that, according to Helgadóttir [2016], shape European policy response to the 2008 recession. All the evidence obtained suggests that the first two panels are structurally different from the third panel, thus the composition of the first two panels do not appear to have a fair representation in terms of diverse viewpoints, scientific perspectives and scholarly thinking heterogeneity.

Three possible objections to this conclusion should be considered. The first one argues that in both the first and the second panel there were still a few non-mainstream scholars, such as Giovanni Dosi. It is possible to counter-argue that the presence of a small minority of members in a quasi-monolithic panel is a form of tokenism, i.e. a symbolic effort to make the panel appear as inclusive and pluralist.

The second objection was argued in an official document [ANVUR, 2012] where ANVUR replied to early critics to the 2004-2010 panel composition. The objection is that the panel "is a group of scholars with a high scientific profile and diversified in terms of skills and geographical origin, in full compliance with the criteria for the selection of panels published on the ANVUR website" [ANVUR, 2012]. This objection is valid if the analysis is limited to the *actual*, i.e. at the moment of panel formation, diversities in terms of geography and affiliations. This paper definitively document that when multiple affiliations, instead of the first one only, and the panellists academic careers are considered, the diversification disappears.

The third objection is also argued in the official document cited above. It takes seriously the question of the strong co-authorship links among the members of the panel 2004-2010, and it is based on the comparison

of the co-authorship network generated by the panel with a control group built by considering the top-20 or top 50 Italian economists in a well known database (www.repec.org). The arguments sounds as follows: the coauthorship network linking the top 50 Italian economists and the president of the panel is similar to the one linking panel members. In Italy, in economics, all the scholars of "high scientific profile" collaborate. It is possible to counter-argue that the choice of the top-20 or top-50 italian economists in Repec is not really a control group, but an *ad hoc* choice. First of all: the call for panellists was open worldwide, hence it is not correct to limit the analysis to Italian economists. Moreover, Repec does not even represents a complete sample of Italian economists since it "is based on a limited sample of the research output in Economics and Finance. Only material catalogued in RePEc is considered. [..] Thus, this list is by no means based on a complete sample" [RePEc, 2017]. Finally, Repec ranking is based on bibliometrics and it is therefore biased towards gender, multidisciplinary methods, and any research orientations pursued by a minority of researchers in their respective disciplines [Corsi et al., 2019].

Once documented that the fairness of composition of panels was not guaranteed, it follows that also the results of the research assessments should be considered as *unfair*. In other words, procedural unfairness in panel composition determines the general unfairness of the research assessment, and unfair results. The notion of "unfair" results should be accurately distinguished from "biased" results. If we take seriously the central tenet that a fair panel composition determines not only a just procedure, but also substantively better decisions, some interesting questions arise: to what extent did procedural unfairness result in biased evaluations? How did unfair panel composition result in distortions in the research evaluation results? What were the mechanisms that translated the unfair composition into bias in the evaluations? What were the effect of excluding from the panels varieties of research experiences?

In the Italian case, some works has documented anomalies in procedures and results of the first two research assessment exercises for economics, statistics and business [Baccini and De Nicolao, 2016, Baccini et al., 2020a, Baccini and De Nicolao, 2021, Corsi et al., 2019]. It can be conjectured that these are the outcome of mechanism such as group thinking that prevent the possibility for panellists to have access to different perspectives and ideas. The presence of systematic biases in the final results of the whole research assessment exercises should be accurately investigated. Insights in this direction can arrive, comparatively, also from a systematic analysis of the fairness of the composition of panels of other research fields in the Italian research assessment exercises.

More generally, our work shows the importance of a careful evaluation of panel composition in administrative research evaluation procedures. An unfair composition may be the result of a careless design: for example, a procedure involving a public call for panellists and then a random draw may not ensure a fair cross-sectional representation of the communities of scholars to be evaluated. But an unfair composition may also be the result of the capture of the regulator-evaluator by some segments of the scholarly community. At the other extreme, an unfair composition may be the result of the strategy of the regulator, who is able to predetermine the results of the evaluation by a suitable choice of panellists.

The results of our work show that it is possible to develop a research strategy that can verify the degree of fairness in panel composition by gathering information to represent their hidden connections. This search strategy can be appropriately adapted to different contexts for verifying the fairness of panel compositions in evaluation procedures.

References

Anonymous. La Ricerca Degli Atenei Italiani All'esame Dell'Anvur, 2011. URL https://www.unipi.it/index.php/amministrazione/item/447-la-ricerca-degli-atenei-italiani-all%E2%80%99esame-dell%E2%80%99anvur?tmpl=component&print=1.

ANVUR. Messaggio di accompagnamento liste GEV – 15 Dicembre 2011, 2011. URL https://web.archive.org/web/20210614022508/https://www.anvur.it/news/messaggio-di-accompagnamento-liste-gev-15-dicembre-2011/.

ANVUR. La bibliometria della VQR, 2012. URL https://www.roars.it/online/wp-content/uploads/2012/03/la_bibliometria_della_vqr.pdf.

ANVUR. Rapporto finale. valutazione della qualità della ricerca 2004-2010 (vqr 2004-2010), 2013. URL https://www.anvur.it/rapporto/files/Area03/VQR2004-2010_Area03_RapportoFinale.pdf.

ANVUR. La selezione dei componenti GEV. Valutazione della Qualità della Ricerca 2011-2014, 2015. URL https://www.anvur.it/wp-content/uploads/2015/09/Doc%20per%20selezione%20membri%20~.pdf.

ANVUR. Valutazione della qualità della ricerca 2011-2014. rapporto finale, 2017. URL http://www.anvur.it/rapporto-2016/files/VQR2011-2014_RapportoFinale_2016.pdf.

ANVUR. Avvisi per la formazione dei GEV della VQR 2015-2019, 2020a. URL https://www.anvur.it/news/avvisi-per-la-formazione-dei-gev-della-vqr-2015-2019/.

ANVUR. Pubblicati gli esiti del sorteggio dei GEV disciplinari VQR 2015-2019, 2020b. URL https://www.anvur.it/news/pubblicati-gli-esiti-del-sorteggio-dei-gev-discipliari-vqr-2015-2019/.

ANVUR. Sorteggio GEV Disciplinari VQR 2015-2019, 2020c. URL https://www.anvur.it/news/sorteggio-gev-disciplinari-vqr-2015-2019/.

ANVUR. VQR 2015-2019 – Avviso 1/2020 – Elenco dei candidati ammessi al sorteggio per la composizione dei GEV disciplinari, 2020d. URL https://www.anvur.it/news/vqr-2015-2019-avviso-1-2020-elenco-dei-candidati-ammessi-al-sorteggio-per-la-composizione-dei

ANVUR. Valutazione della qualità della ricerca 2015-2019. rapporto finale, 2022. URL https://www.anvur.it/wp-content/uploads/2022/07/VQR-2015-2019_Rapporto_Finale_EC_21luglio2022.pdf.

V. Arnaboldi, R. I. M. Dunbar, A. Passarella, and M. Conti. Analysis of co-authorship ego networks. In A. Wierzbicki, U. Brandes, F. Schweitzer, and D. Pedreschi, editors, *Advances in Network Science*, pages 82–96. Springer International Publishing, 2016.

A. Baccini. Gli Esperti Di Valutazione All'italiana, 2011. URL http://www.roars.it/online/gli-esperti-di-valutazione-allitaliana/.

A. Baccini. La VQR di Area 13: una riflessione di sintesi. *Statistica & Società*, 3(3):32–37, 2014. URL https://usiena-air.unisi.it/retrieve/handle/11365/983557/53041/baccini_stasoc-3_2014.pdf.

A. Baccini. Anomalie, costi ed effetti collaterali della valutazione all'italiana, 2016a. URL https://www.roars.it/online/baccini-su-anvur-anomalie-costi-ed-effetti-collaterali-della-valutazione-allitaliana/.

A. Baccini. Napoléon et l'évaluation bibliométrique de la recherche. considérations sur la réforme de l'université et sur l'action de l'agence national d'évaluation en italie. *Canadian Journal of Information and*

Library Science-Revue Canadienne des Sciences de l'Information et de Bibliotheconomie, 40(1):37–57, 2016b.

A. Baccini. La VQR ed oltre. Riflessioni sulla possibilità di valutare credibilmente la ricerca in Italia. *Evoluzione e valutazione della ricerca giuridica*, pages 95–117, 2016c.

A. Baccini. Performance-based incentives, research evaluation systems and the trickle-down of bad science, May, 27-29 2018.

A. Baccini and G. De Nicolao. Do they agree? bibliometric evaluation versus informed peer review in the italian research assessment exercise. *Scientometrics*, 108(3):1651–1671, 2016. ISSN 1588-2861. doi:10.1007/s11192-016-1929-y.

A. Baccini and G. De Nicolao. Just an artifact? the concordance between peer review and bibliometrics in economics and statistics in the italian research assessment exercise. *Quantitative Science Studies*, pages 1–18, 2021. ISSN 2641-3337. doi:10.1162/qss_a_00172.

A. Baccini and M. Ricciardi. VQR, la composizione dei GEV ed una questione di fairness, 2012. URL `https://www.roars.it/online/vqr-la-composizione-dei-gev-ed-una-questione-di-fairness/`.

A. Baccini, L. Barabesi, and G. De Nicolao. On the agreement between bibliometrics and peer review: Evidence from the italian research assessment exercises. *PLOS ONE*, 15(11):e0242520, 2020a. doi:10.1371/journal.pone.0242520.

A. Baccini, L. Barabesi, M. Khelfaoui, and Y. Gingras. Intellectual and social similarity among scholarly journals: An exploratory comparison of the networks of editors, authors and co-citations. *Quantitative Science Studies*, 1(1):277–289, 2020b. doi:10.1162/qss_a_00006.

F. Baccini, L. Barabesi, A. Baccini, M. Khelfaoui, and Y. Gingras. Similarity network fusion for scholarly journals. *Journal of Informetrics*, 16(1):101226, 2022. ISSN 1751-1577. doi:https://doi.org/10.1016/j.joi.2021.101226.

B. Beersma and C. K. De Dreu. Social motives in integrative negotiation: The mediating influence of procedural fairness. *Social Justice Research*, 16(3):217–239, 2003.

R. Bell. *Impure science: Fraud, compromise, and political influence in scientific research*. New York: John Wiley, 1992.

S. Benedetto. La valutazione della qualità della ricerca (VQR 2004-2010), 2012. URL `https://www.unipd.it/sites/unipd.it/files/presentazione_vqr.pdf`.

D. Chubin and E. Hackett. *Peerless science: Peer review and US science policy*. Albany: State University of New York Press, 1990.

J. Cole and S. Cole. *Peer Review in the National Science Foundation: Phase Two of a Study*. Washington, DC: National Academy Press, 1981.

J. Cole, S. Cole, and L. Rubin. *Peer Review in the National Science Foundation: Phase One of a Study*. Washington, DC: National Academy Press, 1979.

S. Cole. *Making Science: Between Nature and Society*. Cambridge, MA: Harvard University Press, 1992.

M. Corsi, C. D'Ippoliti, and F. Lucidi. Pluralism at risk? Heterodox economic approaches and the evaluation of economic research in Italy. *American Journal of Economics and Sociology*, 69(5):1495–1529, 2010. doi:https://doi.org/10.1111/j.1536-7150.2010.00754.x.

M. Corsi, C. D'Ippoliti, and F. Lucidi. On the evaluation of economic research: the case of Italy. *Economia politica*, 28(3):369–402, 2011. doi:https://doi.org/10.1428/35915.

M. Corsi, C. D'Ippoliti, and G. Zacchia. Diversity of backgrounds and ideas: The case of research evaluation in economics. *Research Policy*, 48(9):103820, 2019. ISSN 0048–7333. doi:https://doi.org/10.1016/j.respol.2019.103820.

W. De Nooy, A. Mrvar, and V. Batagelj. *Exploratory social network analysis with Pajek: Revised and expanded edition for updated software*. Cambridge university press, third edition, 2018.

N. Duxbury. *Random justice: On lotteries and legal decision-making*. Oxford University Press, USA, 2002.

J. Elster. *Solomonic judgements: Studies in the limitation of rationality*. Cambridge University Press, 1989.

ESF. European Peer Review Guide. Integrating Policies and Practices into Coherent Procedures, 2011. URL `https://it.scribd.com/document/63421310/European-Peer-Review-Guide`.

T. Ferguson and R. Johnson. Research evaluation in economic theory and policy: identifying and overcoming institutional dysfunctions. *Policy Brief to the G20 Argentina.*, 2018. URL `https://www.ineteconomics.org/perspectives/blog/inet-memo-to-g20-the-trouble-with-economic-research-evaluation`.

GAO. Peer Review Reforms Needed to Ensure Fairness in Federal Agency Grant Selection. *Report to the Chairman, Committee on Governmental Activities, U.S. Senate. Washington, DC: General Accounting Office*, 1994.

D. Gillies. Selecting applications for funding: why random choice is better than peer review. *RT. A Journal on research policy and evaluation*, 2(1), 2014. doi:https://doi.org/10.13130/2282-5398/3834.

V. P. Hans. Challenges to Achieving Fairness in Civil Jury Selection. *Cornell Legal Studies Research Paper*, (21–23), 2021. doi:http://dx.doi.org/10.2139/ssrn.3899957.

S. Harley and F. S. Lee. Research selectivity, managerialism, and the academic labor process: The future of nonmainstream economics in UK universities. *Human relations*, 50(11):1427–1460, 1997. doi:https://doi.org/10.1177/001872679705001105.

O. Helgadóttir. The bocconi boys go to brussels: Italian economic ideas, professional networks and european austerity. *Journal of European Public Policy*, 23(3):392–409, 2016. ISSN 1350-1763. doi:10.1080/13501763.2015.1106573.

D. Hicks. Performance-based university research funding systems. *Research Policy*, 41(2):251–261, 2012. ISSN 0048-7333. doi:10.1016/j.respol.2011.09.007.

I. L. Janis. *Groupthink: Psychological studies of policy decisions and fiascoes*. Houghton Mifflin School, 1982.

S. Kumar. Co-authorship networks: a review of the literature. *Aslib Journal of Information Management*, 2015. doi:https://doi.org/10.1108/AJIM-09-2014-0116.

M. Lamont and K. Huutoniemi. Comparing customary rules of fairness. In M. Lamont, C. Camic, and N. Gross, editors, *Social knowledge in the making*, chapter 6, pages 209–232. University of Chicago Press, 2011.

L. Langfeldt. Expert panels evaluating research: decision-making and sources of bias. *Research Evaluation*, 13(1):51–62, 2004. doi:https://doi.org/10.3152/147154404781776536.

F. S. Lee. The Research Assessment Exercise, the state and the dominance of mainstream economics in British universities. *Cambridge Journal of Economics*, 31(2):309–325, 2007. doi:https://doi.org/10.1093/cje/bel021.

F. S. Lee, X. Pham, and G. Gu. The UK Research Assessment Exercise and the narrowing of UK economics. *Cambridge Journal of Economics*, 37(4):693–717, 2013. doi:https://doi.org/10.1093/cje/bet031.

G. S. Leventhal. What should be done with equity theory? In K. J. Gergen, M. S. Greenberg, and R. H. Willis, editors, *Social Exchange: Advances in Theory and Research*, pages 27–55. Springer US, 1980. doi:https://doi.org/10.1007/978-1-4613-3087-5_2.

M. E. J. Newman. *Networks*. Oxford University Press, Oxford, 2018.

L. Pasinetti. A note on points of dissent. In CIVR, editor, *Relazione finale di area, Area 13*. 2006. URL `https://www.roars.it/online/wp-content/uploads/2011/11/panel_13.pdf`.

A. L. Porter and F. A. Rossini. Peer review of interdisciplinary research proposals science, technology & human values, 10 (1985) 33-38. *Science, Technology, and Human Values*, 10:33—-38, 1985.

A. I. M. J. Rahman, R. Guns, L. Leydesdorff, and T. C. E. Engels. Measuring the match between evaluators and evaluees: cognitive distances between panel members and research groups at the journal level. *Scientometrics*, 109(3):1639–1663, 2016. ISSN 1588-2861. doi:10.1007/s11192-016-2132-x.

C. Re. Teaching Economics in the Time of Ordoliberalist Hegemony. *History of economic ideas*, 27(3): 115–128, 2019. doi:https://doi.org/10.19272/201906103006.

REF. Units of assessment and recruitment of expert panels, 2010. URL `https://www.ref.ac.uk/2014/media/ref/content/pub/unitsofassessmentandrecruitmentofexpertpanels/01_10.pdf`.

RePEc. Top 25% Institutions and Economists in Italy, as of May 2017, 2017. URL `https://ideas.repec.org/top/old/1705/top.italy.html`.

R. Roy. Funding science: The real defects of peer review and an alternative to it. *Science, Technology, & Human Values*, 10(3):73–81, 1985. doi:https://doi.org/10.1177/016224398501000309.

S. R. Sommers and P. C. Ellsworth. How much do we really know about race and juries-a review of social science theory and research. *Chicago-Kent Law Review*, 78(3):997, 2003.

G. D. L. Travis and H. M. Collins. New light on old boys: Cognitive and institutional particularism in the peer review system. *Science, Technology, & Human Values*, 16(3):322–341, 1991. doi:https://doi.org/10.1177/016224399101600303.

M. Van den Brink. *Behind the scenes of science: Gender practices in the recruitment and selection of professors in the Netherlands*. Amsterdam University Press, 2010.

N. J. Van Eck and L. Waltman. *VOSviewer Manual: Manual for VOSviewer version 1.6. 15*. Leiden: Centre for Science and Technology Studies (CWTS) of Leiden University, 2020.

S. Wasserman and K. Faust. *Social Network Analysis: Method and Application*. Cambridge University Press, Cambridge, 1994.

C. Wennerås and A. Wold. Nepotism and sexism in peer-review. *Nature*, 387:341–343, 1997. doi:https://doi.org/10.1038/387341a0.

R. Whitley, J. Gläser, and L. Engwall. *Reconfiguring knowledge production: Changing authority relationships in the sciences and their consequences for intellectual innovation*. OUP Oxford, 2010.

T. Zacharewicz, B. Lepori, E. Reale, and K. Jonkers. Performance-based research funding in eu member states—a comparative assessment. *Science and Public Policy*, 46(1):105–115, 2018. ISSN 0302-3427. doi:10.1093/scipol/scy041.

Supplementary materials

A1 The panellists selection process

2004-2010 panel. For the VQR 2004-2010, the Ministerial decree established that ANVUR governing board should appoint 450 members divided in 14 area panels, and contemporaneously also the 14 panel presidents. However, the ANVUR governing board acted in disagreement with the Ministerial decree: they appointed firstly the presidents of the panels (list published on 10 October 2011), and after almost two months the panel members (list published on 12 December 2011). (In official documents, ANVUR reported that the appointment of presidents and panel members took place at the same time [Benedetto, 2012].) The presidents of the panels were consulted during the drafting of the operating rules of the VQR [Anonymous, 2011]; and it is therefore likely that they had a say in the choice of the other panel members [Baccini, 2016c].

The members of the panels were chosen largely from a list of scholars realized by the CIVR for the never realized VQR 2004-2008 [ANVUR, 2013]. The list was compiled after a public call for experts the deadline of which was 30 June 2010 (https://web.archive.org/web/20100514160659/http://civr.miur.it/modulo.html). ANVUR did not disclose any data on call partecipants. ANVUR described the choice of panel members as a two steps procedure. The first step consisted in defining a set of scholars taking into consideration their qualifications and continuity of scientific production, as well as the evaluation experience. In the second step, ANVUR had to select panel members by covering all the cultural and research lines within the areas, by assuring a 20% of foreign scholars, by having a fair distribution of affiliations and geography, and by paying attention to gender distribution. "In a limited number of cases", but the data were never disclosed, ANVUR chose outside this list. In particular, it selected non-listed names for members of foreign universities [ANVUR, 2013].

ANVUR presented in the final report data about the distribution of the total set of panellists in terms of gender (23.6% women), affiliation (Italian or foreign (20%) affiliation) and geography (Italian scholars are divided according to North, South or Center of Italy) (Table 2.12 in ANVUR [2013]). According to ANVUR, the data showed that the criteria of fairness defined were respected for the whole set of panellists. No evidence about the respect of the criteria for each single panel was presented.

For economics, statistics and business area, the panel was composed by 36 members. 6 members were women (16.7%); 10 members (27.8%) were affiliated with foreign institutions, but only two did not have Italian first and last name. Tullio Jappelli was selected as the president of the panel.

2011-2014 panel. For the VQR 2011-2014, the selection of panellists started with a public call for experts and ended with the formal approval of the composition of the panels and their presidents by the ANVUR governing board (3 September 2015) [ANVUR, 2015]. ANVUR finally selected 400 panellists, divided into 16 areas. Also in this case, the selection process proceeded in two steps. The first step of the selection process of panellists was primarily based on "quality" "measured, where possible, by h-index, total number of citations, any awards of scientific merit, analysis of the elements of curriculum vitae in the expression of interest, etc." [ANVUR, 2015, translation by the authors]. Criteria or thresholds, if any, adopted for this evaluation were not disclosed. In the second step the selection was made by trying to fulfil the following conditions for each panel: coverage of the scientific-discipinary subfield (settore scientifico-disciplinare) inside each area with a number of panellists proportional to the number of expected products to be evaluated; significant percentage of members with foreign affiliation; balanced gender distribution; for Italian candidates, fair distribution of affiliations where possible. If these criteria could not be met by using the list of candidates who responded to the call, ANVUR could have appointed directly other non-listed scholars. The call for panellists was published on the ANVUR website on 5 May 2015, with a deadline of 5 June 2015, then extended to 15 June 2015. 2,149 candidates (30.4% women) responded to the call, 171 of whom for Area

13. Area 13 applicants were predominantly from affiliations with Italian institutions (76.4%)); 30.4% were women).

The selection process took about two months. The appointed panellists received the official invitation to participate in August 2015. The positive responses were close to 99%. Subsequently, ANVUR replaced those who had not accepted the invitation, reaching the final lists.

For economics, statistics and business area, ANVUR selected 31 members and everybody accepted to participate. One undisclosed member was chosen out of those who responded to the call. 9 members were women (29%); 7 members (22.5%) were affiliated with foreign institutions, but only two did not have Italian first and last name. Gabriella Bertocchi was appointed as president.

2015-2019 panel. The VQR was organized into 17 scientific areas and 1 interdisciplinary area for the evaluation of the activities of the 'Third Mission' (the set of activities, beyond teaching and research, with which universities have direct interaction with society). The area of economics, statistics and business was splitted in two different panels: (13a) "economics, statistics and business"; and (13b) "Economics and business sciences".

For the VQR 2015-2019 the panel selection procedure was different from the previous ones. It started, as the previous ones, with a call for experts. Then the 600 panellists were not chosen by ANVUR governing board, but randomly selected from from those who had applied and met the requirements of high qualification and international experience in research and its evaluation at the time. It should be noted also that, differently from the previous procedures, all the panellists were selected from among the scholars who responded to the call. In particular, applications for being panellist were open from 5 February to 2 March 2020 [ANVUR, 2020a]. On 11 September 2020, ANVUR published the lists of candidates admitted to the draw [ANVUR, 2020d]. The list included 4.137 scholars, 350 of whom were candidates for Area 13 panels (174 candidates for panel 13a and 176 for panel 13b). Among these candidates 37.7% were women; no candidates did have a foreign affiliation. The draw, transmitted also in streaming, took place on 17 September 2020 [ANVUR, 2020c]. The results of the draw were made available on the same day (https://web.archive.org/web/20220308022222/https://www.anvur.it/attivita/vqr/vqr-2015-2019/gev/esito-del-sorteggio-dei-gev-disciplinari/). The composition of the panels resulted from the draw was later slightly modified by ANVUR governing board, by adding 4 members and by replacing a few of resigning members; these modification were completely disclosed [ANVUR, 2020b]. To completely exclude ANVUR's intervention in the selection of panel members, we considered the results of the draw for Area "13a economics, statistics and business" and Area "13b Economics and business sciences" as our "control group". In the paper we refer to this group of 40 scholars as the Panel of the 2015-2019.

Where possible, each panel was formed in compliance with a complex list of requirements: at least 25% of the members had to be full professors (Professore ordinario); at least 20%, respectively, had to be associate professor (professore associato) or research fellow (ricercatore) in Italian universities; up to a maximum of 30% can be researchers structured at Public Research Bodies (EPR); at least 5% had to be researchers in foreign universities or research bodies. Each panel had to have at least one member for each different recruitment field (settore concorsuale) and for each disciplinary sub-field (settore scientifico-disciplinare) of the area with at least 50 members. The remaining members had to be distributed in proportion to the size of the different fields in the area. Moreover: each gender had to be represented for at least one third; no more than 20% of the members may have belonged to the panel 2011-2014. Once members of the panels were appointed, the ANVUR governing board identified, choosing among them, the 18 panel coordinators.

Result of the draw for "13a economics, statistics and business" was a list of 22 members; the panel in its final composition was composed by 23 members. Results of the draw for "13b Economics and business sciences" was a list of 18 members; the panel in its final composition was composed by 21 members. Overall

the two panel included 17 women (38,6%); no members were affiliated with foreign institutions. The appointed coordinators were respectively Emanuela Marrocu and Maria Rosaria Napolitano.

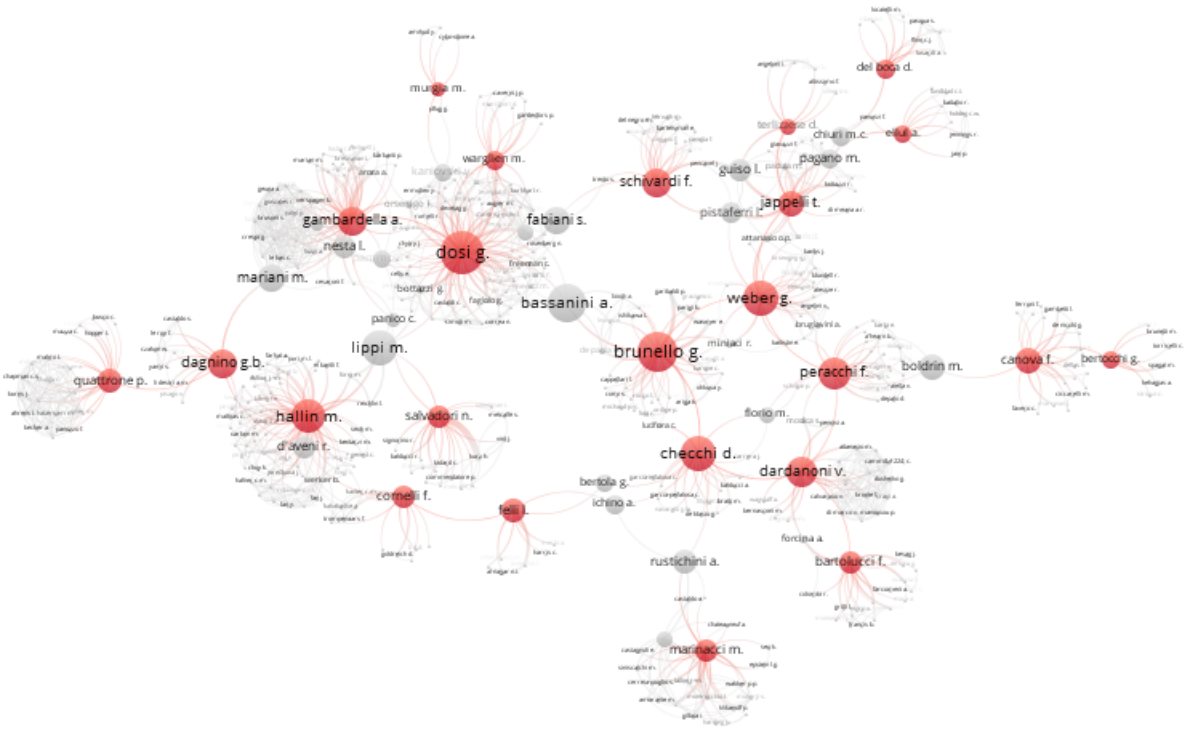


Figure A1: The largest component in the co-authorship network for the panel 2004-2010. Size of nodes is proportional to betweenness centrality. Red nodes are panel members.

Degree	Frequency	Frequency (%)	Degree	Frequency	Frequency (%)	Degree	Frequency	Frequency (%)
1	169	21.64	12	5	0.64	23	1	0.13
2	249	31.88	13	17	2.18	27	1	0.13
3	109	13.96	14	14	1.79	28	2	0.26
4	59	7.55	16	2	0.26	34	1	0.13
5	33	4.23	17	2	0.26	43	1	0.13
6	27	3.46	18	1	0.13	47	1	0.13
7	14	1.79	19	18	2.30	48	1	0.13
8	19	2.43	20	2	0.26	57	1	0.13
9	15	1.92	21	6	0.77	64	1	0.13
10	7	0.90	22	2	0.26	78	1	0.13

Table A1: Degree frequency distribution of the co-authorship network for the panel 2004-2010

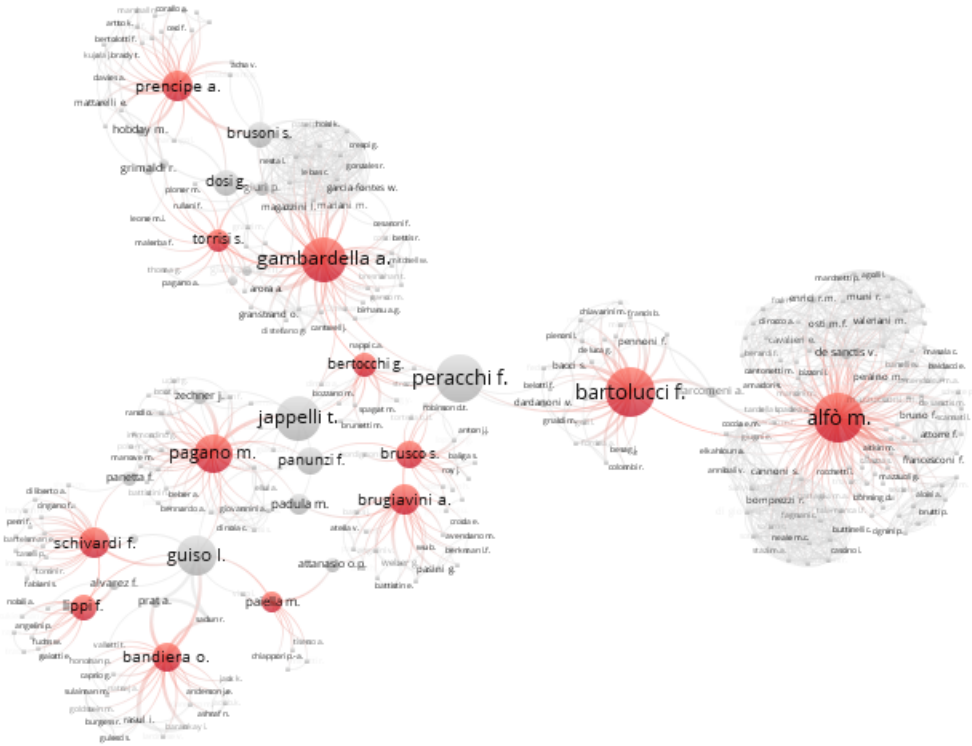


Figure A2: The largest component of the co-authorship network for the panel 2011-2014. Size of vertices is proportional to betweenness centrality. Red nodes are panel members.

Degree	Frequency	Frequency (%)	Degree	Frequency	Frequency (%)	Degree	Frequency	Frequency (%)
1	136	14.75	16	6	0.65	33	4	0.43
2	242	26.25	17	10	1.08	34	2	0.22
3	168	18.22	18	8	0.87	35	2	0.22
4	79	8.57	19	15	1.63	37	1	0.11
5	44	4.77	20	16	1.74	38	1	0.11
6	38	4.12	21	8	0.87	39	2	0.22
7	22	2.39	22	5	0.54	40	2	0.22
8	24	2.60	23	4	0.43	44	1	0.11
9	6	0.65	24	1	0.11	50	1	0.11
10	20	2.17	25	2	0.22	51	1	0.11
11	12	1.30	26	9	0.98	57	1	0.11
12	16	1.74	27	1	0.11	58	1	0.11
13	3	0.33	30	1	0.11	78	1	0.11
14	4	0.43	32	1	0.11	128	1	0.11

Table A2: Degree frequency distribution of the co-authorship network for the panel 2011-2014

Degree	Frequency	Frequency (%)	Degree	Frequency	Frequency (%)	Degree	Frequency	Frequency (%)
1	67	5.27	17	7	0.55	34	1	0.08
2	219	17.23	18	36	2.83	35	1	0.08
3	196	15.42	19	2	0.16	36	1	0.08
4	136	10.70	20	3	0.24	37	1	0.08
5	80	6.29	21	3	0.24	38	3	0.24
6	69	5.43	22	2	0.16	39	5	0.39
7	61	4.80	23	27	2.12	41	2	0.16
8	34	2.68	24	22	1.73	42	2	0.16
9	44	3.46	25	3	0.24	44	2	0.16
10	15	1.18	27	29	2.28	46	1	0.08
11	55	4.33	28	1	0.08	49	1	0.08
12	14	1.10	29	4	0.31	59	2	0.16
13	37	2.91	30	3	0.24	61	2	0.16
14	23	1.81	31	3	0.24	84	1	0.08
15	5	0.39	32	28	2.20	151	1	0.08
16	15	1.18	33	1	0.08	259	1	0.08

Table A3: Degree frequency distribution of the co-authorship network for the panel 2015-2019

Name	Betweenness centrality	Rank Betweenness	panel member	Name	Betweenness centrality	Rank Betweenness	panel member
Dosi G.	0.246	1	Yes	D'Aveni R.	0.023	26	No
Brunello G.	0.183	2	Yes	Marinacci M.	0.023	27	Yes
Bassanini A.	0.147	3	No	Pistaferri L.	0.019	28	No
Weber G.	0.116	4	Yes	Pagano M.	0.017	29	No
Checchi D.	0.116	5	Yes	Ichino A.	0.017	30	No
Lippi M.	0.115	6	No	Warglien M.	0.017	31	Yes
Hallin M.	0.096	7	Yes	Chiuri M.C.	0.016	32	No
Peracchi F.	0.088	8	Yes	Del Boca D.	0.015	33	Yes
Dardanoni V.	0.066	9	Yes	Ellul A.	0.013	34	Yes
Gambardella A.	0.064	10	Yes	Bertola G.	0.013	35	No
Dagnino G.B.	0.058	11	Yes	Pammolli F.	0.010	36	No
Schivardi F.	0.057	12	Yes	Bertocchi G.	0.010	37	Yes
Mariani M.	0.045	13	No	Kaniovski Y.	0.010	38	No
Fabiani S.	0.045	14	No	Terlizzese D.	0.008	39	Yes
Jappelli T.	0.044	15	Yes	Maccheroni F.	0.008	40	No
Boldrin M.	0.043	16	No	Marengo L.	0.008	41	No
Canova F.	0.042	17	Yes	Florio M.	0.008	42	No
Rustichini A.	0.035	18	No	Guthrie J.	0.007	43	Yes
Quattrone P.	0.033	19	Yes	Murgia M.	0.007	44	Yes
Felli L.	0.031	20	Yes	Panico C.	0.005	45	No
Salvadori N.	0.028	21	Yes	Magazzini L.	0.004	46	No
Bartolucci F.	0.027	22	Yes	Frino A.	0.003	47	Yes
Cornelli F.	0.027	23	Yes	Cichelli A.	0.003	48	Yes
Nesta L.	0.026	24	No	Jones L.R.	0.002	49	No
Guiso L.	0.024	25	No	Mussari R.	0.002	50	Yes

Table A4: Betweenness centrality and rank betweenness of the co-authorship network for the panel 2004-2010

Name	Betweenness centrality	Rank Betweenness	panel member	Name	Betweenness centrality	Rank Betweenness	panel member
Bartolucci F.	0.099	1	Yes	Brusoni S.	0.012	14	No
Alfò M.	0.097	2	Yes	Dosi G.	0.011	15	No
Peracchi F.	0.096	3	No	Lippi F.	0.010	16	Yes
Jappelli T.	0.073	4	No	Bertocchi G.	0.009	17	Yes
Gambardella A.	0.071	5	Yes	Torrisi S.	0.007	18	Yes
Guiso L.	0.051	6	No	Vivarelli M.	0.006	19	Yes
Pagano M.	0.041	7	Yes	Greco S.	0.006	20	Yes
Brugiavini A.	0.019	8	Yes	Padula M.	0.005	21	No
Schivardi F.	0.019	9	Yes	Paiella M.	0.005	22	Yes
Prencipe A.	0.018	10	Yes	De Fraja G.	0.004	23	Yes
Bandiera O.	0.017	11	Yes	Piga C.	0.003	24	No
Panunzi F.	0.015	12	No	Sarno L.	0.003	25	Yes
Brusco S.	0.014	13	Yes	Frey M.	0.002	26	No

Table A5: Betweenness centrality and rank betweenness of the co-authorship network for the panel 2011-2014

Name	Betweenness centrality	Rank Betweenness	panel member	Name	Betweenness centrality	Rank Betweenness	panel member
Ruggeri F.	0.045	1	Yes	Meliciani V.	0.003	9	Yes
Nicolis O.	0.041	2	No	Mencarini L.	0.002	10	Yes
Stingo F.	0.036	3	Yes	Pagano M.	0.002	11	Yes
Fassò A.	0.031	4	Yes	Castellani D.	0.002	12	No
Chiodi M.	0.016	5	Yes	Piva M.	0.002	13	Yes
Adelfio G.	0.008	6	Yes	Notarnicola B.	0.002	14	Yes
Bevilacqua M.	0.007	7	No	Savona M.	0.002	15	No
Perna A.	0.006	8	Yes	Antonioli D.	0.002	16	Yes

Table A6: Betweenness centrality and rank betweenness of the the co-authorship network for the panel 2015-2019

Cluster	Freq	Freq%	Members Freq	Members Freq%
1	512	65.5	24	66.6
2	6	0.7	1	2.7
3	71	9.0	2	5.5
4	18	2.3	1	2.7
5	49	6.2	1	2.7
6	18	2.3	1	2.7
7	48	6.1	1	2.7
8	6	0.7	1	2.7
9	8	1.0	1	2.7
10	28	3.5	1	2.7
11	8	1.0	1	2.7
12	9	1.1	1	2.7
Sum	781	100	36	100

Table A7: Frequency distribution among components of the co-authorship network for the panel 2004-2010

Cluster	Freq	Freq%	Members Freq	Members Freq%
1	79	8.5	1	3.2
2	52	5.6	1	3.2
3	403	43.7	13	41.9
4	34	3.6	1	3.2
5	7	0.7	1	3.2
6	58	6.2	1	3.2
7	25	2.7	1	3.2
8	20	2.1	1	3.2
9	22	2.3	1	3.2
10	17	1.8	1	3.2
11	22	2.3	1	3.2
12	79	8.5	3	9.6
13	36	3.9	1	3.2
14	12	1.3	1	3.2
15	39	4.2	1	3.2
16	9	0.9	1	3.2
17	8	0.8	1	3.2
Sum	922	100	31	100

Table A8: Frequency distribution among components of the co-authorship network for the panel 2011-2014

Cluster	Freq	Freq%	Members Freq	Members Freq%
1	68	5.3	2	5.4
2	324	25.4	5	13.5
3	64	5.0	2	5.4
4	42	3.3	1	2.7
5	18	1.4	1	2.7
6	260	20.4	1	2.7
7	16	1.2	1	2.7
8	11	0.8	1	2.7
9	44	3.4	3	8.1
10	39	3.0	1	2.7
11	88	6.9	3	8.1
12	60	4.7	1	2.7
13	27	2.1	2	5.4
14	13	1.0	1	2.7
15	2	0.1	1	2.7
16	18	1.4	1	2.7
17	42	3.3	1	2.7
18	26	2.0	1	2.7
19	18	1.4	1	2.7
20	14	1.1	1	2.7
21	17	1.3	1	2.7
22	15	1.1	2	5.4
23	12	0.9	1	2.7
24	21	1.6	1	2.7
25	12	0.9	1	2.7
Sum	1271	100	37	100

Table A9: Frequency distribution among components of the co-authorship network for the panel 2015-2019

Degree	Frequency	Frequency (%)	Degree	Frequency	Frequency (%)	Degree	Frequency	Frequency (%)
0	2	5.56	6	4	11.11	14	1	2.78
1	1	2.78	7	2	5.56	15	1	2.78
2	2	5.56	9	2	5.56	16	4	11.11
3	3	8.33	11	2	5.56	17	2	5.56
4	2	5.56	12	3	8.33			
5	4	11.11	13	1	2.78			

Table A10: Degree frequency distribution of the journal based network for the panel 2004-2010

Degree	Frequency	Frequency (%)	Degree	Frequency	Frequency (%)	Degree	Frequency	Frequency (%)
0	3	9.68	6	2	6.45	14	5	16.13
1	2	6.45	7	1	3.23	15	1	3.23
2	2	6.45	8	2	6.45	17	2	6.45
3	1	3.23	10	2	6.45	19	1	3.23
4	1	3.23	12	2	6.45			
5	1	3.23	13	3	9.68			

Table A11: Degree frequency distribution of the journal based network for the panel 2011-2014

Degree	Frequency	Frequency (%)	Degree	Frequency	Frequency (%)
1	3	8.11	9	1	2.70
3	3	8.11	10	3	8.11
4	2	5.41	11	2	5.41
5	5	13.51	12	1	2.70
6	3	8.11	13	2	5.41
7	4	10.81	14	1	2.70
8	6	16.22	22	1	2.70

Table A12: Degree frequency distribution of the journal based network for the panel 2015-2019

Name	Betweenness centrality	Rank Betweenness	Island Color	Name	Betweenness centrality	Rank Betweenness	Island Color
Dosi G.	0.138	1	Blue	Ellul A.	0.018	19	Green
Frino A.	0.125	2	Grey	Bottazzi L.	0.015	20	Green
Guthrie J.	0.081	3	Red	Frittelli M.	0.010	21	Grey
Bisin A.	0.079	4	Green	Marinacci M.	0.010	22	Green
Jappelli T.	0.058	5	Green	Hallin M.	0.006	23	Green
Schivardi F.	0.051	6	Green	Checchi D.	0.006	24	Green
Brunello G.	0.050	7	Green	Quattrone P.	0.005	25	Red
Peracchi F.	0.050	8	Green	Bartolucci F.	0.004	26	Green
Gambardella A.	0.048	9	Blue	Salvadori N.	0.001	27	Grey
Weber G.	0.042	10	Green	Cornelli F.	0.001	28	Green
Dagnino G.B.	0.041	11	Grey	Del Boca D.	0.001	29	Green
Bertocchi G.	0.031	12	Green	Bergami M.	0.000	30	Grey
Rossi B.	0.031	13	Green	Cichelli A.	0.000	31	Grey
Warglien M.	0.028	14	Blue	Guido G.	0.000	32	Grey
Canova F.	0.025	15	Green	Murgia M.	0.000	33	Grey
Dardanoni V.	0.022	16	Green	Mussari R.	0.000	34	Red
Felli L.	0.020	17	Green	Terlizzese D.	0.000	36	Green
Ronchetti E.	0.020	18	Green	Zamagni V.	0.000	35	Grey

Table A13: Betweenness centrality, rank betweenness and island of the journal based network of scholars of the panel 2004-2010

Name	Betweenness centrality	Rank Betweenness	Island Color	Name	Betweenness centrality	Rank Betweenness	Island Color
Gambardella A.	0.130	1	Green	Brusco S.	0.009	17	Green
Cinquini L.	0.115	2	Grey	Alfò M.	0.009	18	Red
Sarno L.	0.100	3	Green	Paiella M.	0.008	19	Green
Fiordelisi F.	0.076	4	Grey	Corbetta G.	0.008	20	Grey
Vivarelli M.	0.060	5	Green	Brugiavini A.	0.006	21	Green
Larrinaga-G. C.	0.060	6	Grey	Ronchetti E.	0.003	22	Red
Bertocchi G.	0.052	7	Green	Bartolucci F.	0.003	23	Red
Prencipe A.	0.043	8	Green	Torrisi S.	0.002	24	Green
Schivardi F.	0.040	9	Green	Asso P.F.	0.002	25	Grey
Bandiera O.	0.027	10	Green	De Carlo M.	0.000	26	Grey
Pagano M.	0.025	11	Green	De Paola M.	0.000	27	Grey
Cavaliere G.	0.023	12	Green	Greco S.	0.000	28	Grey
De Fraja G.	0.023	13	Green	Malanima P.	0.000	29	Grey
Kretschmer T.	0.021	14	Green	Rosazza G. E.	0.000	30	Grey
Pesavento E.	0.016	15	Green	Salomone R.	0.000	31	Grey
Lippi F.	0.011	16	Green				

Table A14: Betweenness centrality, rank betweenness and island of the journal based network of scholars of the panel 2011-2014

Name	Betweenness centrality	Rank Betweenness	Island Color	Name	Betweenness centrality	Rank Betweenness	Island Color
Piva M.	0.272	1	Green	Cavallo B.	0.016	20	Grey
Michetti E.	0.063	2	Grey	Adelfio G.	0.010	21	Blue
Marrocu E.	0.061	3	Green	La Rosa F.	0.009	22	Red
Pretaroli R.	0.059	4	Grey	Bracci E.	0.008	23	Red
Berni R.	0.059	5	Blue	Corsi K.	0.008	24	Red
Ruggeri F.	0.059	6	Blue	Antonioli D.	0.007	25	Green
Pagano M.	0.056	7	Grey	Fassò A.	0.006	26	Blue
Opocher A.	0.056	8	Grey	Mavilia R.	0.004	27	Green
Vasta M.	0.056	9	Grey	Stingo F.	0.004	28	Blue
Bajo E.	0.054	10	Grey	Chiodi M.	0.004	29	Blue
Picchio M.	0.047	11	Grey	Levrero E.S.	0.004	30	Grey
Gaeta G.L.	0.046	12	Grey	Scalzo V.	0.003	31	Grey
Meliciani V.	0.041	13	Green	Avallone P.	0.000	32	Grey
Perna A.	0.038	14	Grey	Dell'Atti S.	0.000	33	Grey
Napolitano M.R.	0.034	15	Grey	Ghellini G.	0.000	34	Grey
Altomonte C.	0.034	16	Green	Guerriero C.	0.000	35	Grey
Mencarini L.	0.027	17	Grey	Mason M.	0.000	36	Grey
Greco G.	0.020	18	Grey	Mosca M.	0.000	37	Grey
Notarnicola B.	0.019	19	Grey				

Table A15: Betweenness centrality, rank betweenness and island of the journal based network of scholars of the panel 2015-2019

Scholars					Affiliated institutions			
Name	**Eigenvector centrality**	**Rank Eigenvector**	**Island Color**	**Panel member**	**Name**	**Eigenvector centrality**	**Rank Eigenvector**	**Island Color**
Pagano M.	0.307	1	Green	No	CEPR	0.514	1	Green
Ichino A.	0.286	2	Green	No	NBER	0.394	2	Green
Schivardi F.	0.255	3	Green	Yes	Bocconi U.	0.345	3	Green
Guiso L.	0.252	4	Green	No	EIEF	0.289	4	Green
Pistaferri L.	0.239	5	Green	No	lavoce.info	0.275	5	Green
Bisin A.	0.234	6	Green	Yes	IGIER	0.196	6	Green
Bottazzi L.	0.227	7	Green	Yes	IlSole24Ore	0.176	7	Green
Rustichini A.	0.224	8	Green	No	LSE	0.153	8	Green
Checchi D.	0.214	9	Green	Yes	MIT	0.148	9	Green
Peracchi F.	0.203	10	Green	Yes	IZA	0.126	10	Green
Ellul A.	0.189	11	Green	Yes	Bank of Italy	0.114	11	Green
Rossi B.	0.183	12	Grey	Yes	EEA	0.110	12	Green
Boldrin M.	0.183	13	Green	No	European UI	0.105	13	Green
Terlizzese D.	0.173	14	Green	Yes	Il Foglio	0.101	14	Green
Weber G.	0.168	15	Green	Yes	Bologna U.	0.094	15	Green
Gambardella A.	0.166	16	Grey	Yes	Cambridge U.	0.085	16	Grey
Jappelli T.	0.157	17	Green	Yes	CSEF	0.076	17	Grey

Table A16: The important vertices for the affinity network of the panel 2004-2010. The number of important vertices is fixed to 17, about a half of the size of the panel.

Network of scholars					Network of affiliation			
Name	**Eigenvector centrality**	**Rank Eigenvector**	**Island Color**	**panel member**	**Name**	**Eigenvector centrality**	**Rank Eigenvector**	**Island Color**
Pagano M.	0.350	1	Green	Yes	CEPR	0.519	1	Green
Schivardi F.	0.294	2	Green	Yes	NBER	0.383	2	Green
Guiso L.	0.287	3	Green	No	lavoce.info	0.348	3	Green
Padula M.	0.285	4	Green	No	Bocconi U.	0.341	4	Green
Panunzi F.	0.282	5	Green	No	EIEF	0.272	5	Green
Lippi F.	0.239	6	Green	Yes	Bank of Italy	0.136	6	Green
Brugiavini A.	0.216	7	Green	Yes	IGIER	0.119	7	Green
Peracchi F.	0.215	8	Green	No	LSE	0.116	8	Green
Jappelli T.	0.214	9	Green	No	Roma Tor Vergata U.	0.112	9	Grey
Bandiera O.	0.199	10	Green	Yes	Napoli Federico II U.	0.110	10	Grey
Brusco S.	0.196	11	Green	Yes	CSEF	0.108	11	Grey
Torrisi S.	0.196	12	Green	Yes	CFS	0.108	12	Grey
Gambardella A.	0.195	13	Green	Yes	ECGI	0.108	13	Grey
Sarno L.	0.193	14	Green	Yes	ECB	0.100	14	Grey
Paiella M.	0.166	15	Grey	Yes	EEA	0.094	15	Grey
De Paola M.	0.148	16	Grey	Yes	IlSole24Ore	0.093	16	Grey
Bertocchi G.	0.144	17	Grey	Yes	Il Foglio	0.091	17	Grey

Table A17: The important vertices for the affinity network of the panel 2011-2014. The number of important vertices is fixed to 17, about a half of the size of the panel.

Network of scholars					Network of affiliation			
Name	**Eigenvector centrality**	**Rank Eigenvector**	**Island Color**	**Panel member**	**Name**	**Eigenvector centrality**	**Rank Eigenvector**	**Island Color**
Pagano M.	0.471	1	Green	Yes	Bocconi U.	0.364	1	Green
Mencarini L.	0.415	2	Grey	Yes	lavoce.info	0.341	2	Green
Altomonte C.	0.385	3	Green	Yes	SIS	0.283	3	Violet
Guerriero C.	0.312	4	Green	Yes	CEPR	0.252	4	Green
Picchio M.	0.247	5	Green	Yes	Firenze U.	0.216	5	Violet
Nicolis O.	0.203	6	Violet	No	ECB	0.176	6	Green
Mavilia R.	0.185	7	Grey	Yes	IlSole24Ore	0.176	7	Green
Fassò A.	0.158	8	Violet	Yes	Bologna U.	0.175	8	Brown
Vasta M.	0.144	9	Grey	Yes	Cambridge U.	0.161	9	Green
Bajo E.	0.143	10	Grey	Yes	EIEF	0.161	10	Green
Savona M.	0.137	11	Yellow	No	Torino U.	0.153	11	Violet
Mosca M.	0.133	12	Light Blue	Yes	CCA	0.150	12	Green
Chiodi M.	0.125	13	Violet	Yes	Napoli Federico II U.	0.135	13	Orange
Ghellini G.	0.124	14	Grey	Yes	UCL - Louvain	0.130	14	Green
Adelfio G.	0.121	15	Violet	Yes	ISPI	0.117	15	Green
Berni R.	0.119	16	Grey	Yes	Padova U.	0.111	16	Violet
Stingo F.	0.113	17	Grey	Yes	GLO	0.105	17	Green

Table A18: The important vertices for the affinity network of the panel 2015-2019. The number of important vertices is fixed to 17, about a half of the size of the panel.

Network of scholars				Network of affiliation			
Island	**Scholars Number**	**Important vertices Freq**	**Important vertices Freq%**	**Island**	**Affiliations Number**	**Important vertices Freq**	**Important vertices Freq%**
Grey	36	2	11.8	Grey	153	2	11.8
Red	2	0	0	Red	5	0	0
Yellow	2	0	0	Yellow	2	0	0
Blue	2	0	0	Light Blue	6	0	0
Green	16	15	88.2	Green	15	15	88.2
				Violet	2	0	0
				Pink	4	0	0
				Orange	2	0	0
				Brown	2	0	0
Sum	58	17	100	**Sum**	191	17	100

Table A19: Frequency distribution of island values of panel 2004-2010 affinity network

Network of scholars				Network of affiliation			
Island	**Scholars Number**	**Important vertices Freq**	**Important vertices Freq%**	**Island**	**Affiliations Number**	**Important vertices Freq**	**Important vertices Freq%**
Grey	23	3	17.7	Grey	131	9	52.9
Red	2	0	0	Red	8	0	0
Green	15	14	82.3	Green	8	8	47.1
Sum	40	17	100	**Sum**	147	17	100

Table A20: Frequency distribution of island values of panel 2011-2014 affinity network

Network of scholars				Network of affiliation			
Island Color	**Scholars Number**	**Important vertices Freq**	**Important vertices Freq%**	**Island Color**	**Affiliations Number**	**Important vertices Freq**	**Important vertices Freq%**
Grey	22	7	41.2	Grey	131	0	0
Blue	4	0	0	Blue	3	0	0
Brown	2	0	0	Brown	2	1	5.9
Orange	3	0	0	Orange	3	1	5.9
Light Blue	2	1	5.9	Light Blue	2	0	0
Violet	4	4	23.5	Violet	9	4	23.5
Green	4	4	23.5	Green	15	11	64.7
Yellow	3	1	5.9	Yellow	2	0	0
				Red	4	0	0
Sum	44	17	100	**Sum**	171	17	100

Table A21: Frequency distribution of island values of panel 2015-2019 affinity network