New and atypical combinations: An assessment of novelty and interdisciplinarity

Magda Fontana*[a,b], Martina Iori[c,b], Fabio Montobbio[d,e,f,a], and Roberta Sinatra[g,h,i]

[a]Despina Big Data Lab, University of Turin, Lungo Dora Siena, 100A, Turin, 10153, Italy
[b]Dept. of Economics and Statistics, University of Turin, Lungo Dora Siena, 100A, Turin, 10153, Italy
[c]Institute of Economics & EMbeDS, Sant'Anna School of Advanced Studies, Piazza Martiri della Libertà, 33, 56127, Pisa, Italy
[d]Dept. of Economic Policy, Università Cattolica del Sacro Cuore, Largo Gemelli, 1, 20123 Milan, Italy
[e]BRICK, Collegio Carlo Alberto, Piazza Arbarello, 8, 10122, Turin, Italy
[f]ICRIOS, Bocconi University, Via Roberto Sarfatti, 25, 20100, Milan, Italy
[g]Department of Computer Science, IT University of Copenhagen, Rued Langgaards Vej, 7, 2300, Copenhagen, Denmark
[h]Complexity Science Hub, Josefstädter Strasse, 39, 1080, Vienna, Austria
[i]ISI Foundation, Via Chisola, 5, 10126, Turin, Italy

Abstract

Novelty indicators are increasingly important for science policy. This paper challenges the indicators of novelty as an atypical combination of knowledge (Uzzi et al., 2013) and as the first appearance of a knowledge combination (Wang et al., 2017). We exploit a sample of 230,854 articles (1985 - 2005), published on 8 journals of the American Physical Society (APS) and 2.4 million citations to test the indicators using (i) a Configuration Null Model, (ii) an external validation set of articles related to Nobel Prize winning researches and APS Milestones, (iii) a set of established interdisciplinarity indicators, and (iv) the relationship with the articles' impact. We find that novelty as the first appearance of a knowledge combination captures the key structural properties of the citation network and finds it difficult to tell novel and non-novel articles apart, while novelty as an atypical combination of knowledge overlaps with interdisciplinarity. We suggest that the policy evidence derived from these measures should be reassessed.

Keywords: Interdisciplinarity, Novelty, Impact, Indicator, Physics

JEL codes: I23; O31; O33; O38

*Corresponding author: magda.fontana@unito.it

1 Introduction

In the Schumpeterian tradition, novelty results from the recombination of existing bits of knowledge (March, 1991; Nelson and Winter, 1982; Schumpeter, 1939): discoveries do not appear out of thin air but are derived from what is already known (Arthur, 2009). The recent empirical literature (Uzzi et al., 2013; Wang et al., 2017) on novelty in science tends to give a narrow interpretation to this claim measuring the novelty of scientific discovery by unprecedented (Wang et al., 2017) and distant (Uzzi et al., 2013) combinations of disciplines provided by the backward citations of the article reporting the discovery. Recent empirical work has also used the disciplines associated with backward citations of a paper to quantify its interdisciplinarity (Porter and Rafols, 2009; Yegros-Yegros et al., 2015). Here, we provide evidence that the prevailing approaches to measure novelty are strongly related to how we quantify interdisciplinary research, and that the currently used operationalizations of the two concepts capture essentially the same property of a paper. Interdisciplinary research is believed to foster novelty (D'Este et al., 2019) and therefore some correlation is expected between the two measures. However, measures of novelty and interdisciplinarity should capture different properties of scientific discovery, since interdisciplinary research is not necessarily new, and new research is not necessarily interdisciplinary. For example, while Nobel Prize winning researches are often perceived as new, some result from monodisciplinary efforts and others from interdisciplinary approaches (Szell et al., 2018).

The Nobel Prize in physics awarded in 2007 for the discovery of giant magnetoresistance (Baibich et al., 1988; Binasch et al., 1989) was a fundamental breakthrough stemming from work in atomic physics. By contrast, the development of laser-based precision spectroscopy (Diddams et al., 2000; Reichert et al., 2000), awarded the Nobel Prize in 2005, involved a substantial interdisciplinary effort combining knowledge from material science, spectroscopy, chemistry, optoelectronics, and atomic physics. Interdisciplinarity and novelty are thus two distinct aspects of scientific discovery, and interdisciplinarity is not a requirement of producing novel research. The need for research policy to identify both novel and interdisciplinary research prompts us to investigate the currently used indicators.

The blurred boundaries between novelty and interdisciplinarity create interesting challenges to the science-policy debate and the related empirical work. Interdisciplinarity is often considered valuable in itself - i.e. without any particular emphasis on the resulting novelty - because it increases variety in research and counterbalances the increasing specialization in many fields (Cedrini and Fontana, 2017) and contributes to the search for adequate solutions to many social and scientific problems (Molas-Gallart and Salter, 2002; Rafols et al., 2012; Wang et al., 2017). Scientific institutions have systematically promoted interdisciplinary research suggesting that important ideas originate beyond the relatively narrow boundaries of a single discipline (European Union Research Advisory Board, 2004; Fortunato et al., 2018; The National Academies, 2005). Novelty, on the other hand, pushes the knowledge frontier and creates the conditions for innovation and productivity growth. Novel research raises interesting policy issues because it has a high potential for a large impact in the economy but, at the same time, it is subject to a high level of uncertainty in its outcomes (Mansfield, 1991).

The policy challenges associated with interdisciplinary research and novelty are not necessarily in contrast but are nonetheless different. For instance, Collaborative Interdisciplinary Team Science at NIH, the NSF's interdisciplinary research initiatives and the Synthesis Centers flourishing in the US, Europe, China, and Australia are all examples of policy efforts to bring interdisciplinary groups of specialists and experts together for extended periods of time. Here the policy focus is on the efficient design of funding schemes, peer review systems, and hiring policies that strike the right balance between interdisciplinarity and specialization. On the novelty side, the purpose of policies is to produce innovation through the right balance between funding of relatively safe projects and riskier projects (Boudreau et al., 2016; Stephan et al., 2017; Wang et al., 2018). We maintain that the key issue to tackling these challenges is the precise measurement of interdisciplinarity and novelty, for their blurred boundaries are also echoed in the construction of the indicators used in the recent empirical research. We will show that both novelty and interdisciplinarity indicators depend on the diversity and distance of the combined knowledge domains and therefore the two measurements are, by definition, tightly linked. Taken together, the contribution of this paper is to challenge the indicators underpinning the current literature on novelty (Criscuolo et al., 2017; Kim et al., 2016; Lee et al., 2015; Strumsky and Lobo, 2015; Verhoeven et al., 2016; Wang et al., 2018), namely those that identify novelty as an atypical combination of knowledge as in Uzzi et al. (2013) (hereafter, NoveltyU) and those that identify novelty as the first appearance of a new combination as in the indicator introduced by Wang et al. (2017) (hereafter, NoveltyW). First, we ask whether these indicators capture the variable they are intended to measure and, second, whether their relationship with performance indicators used to devise policies, namely the impact of articles, is statistically significant. Third, we aim to investigate whether the indicators can distinguish between novelty and interdisciplinarity.

This paper is the first attempt, to the best of our knowledge, to investigate the behavior of these indicators of novelty also in comparison with the measures of interdisciplinarity. Finally, the paper discusses the policy consequences of potential measurement errors in evaluating novelty and interdisciplinarity.

The empirical analysis exploits a sample of 230,854 focal articles, published in 8 journals of the American Physical Society (APS) between 1985 and 2005, 355,092 citing articles (1985-2015), and 2,439,359 citations. Fields are identified using the codes of the Physics and Astronomy Classification Scheme (PACS). The dataset has several advantages: the APS represents the variety of research in the field and its journals have published a fair share of articles whose authors have been awarded the Nobel Prize (among the Nobel Prizes appointed in the last 50 years, 31 related papers had been published in APS journals). In addition, we can rely on three lists of APS milestone articles (Physical Review E 25th Anniversary Milestones, Letters from the Past - A PRL Retrospective, Celebrating 125 years of The Physical Review) that considered groundbreaking by the experts to create a sample that we can use to corroborate our findings. PACS codes also provide a hierarchical and stable classification of knowledge and identify knowledge niches more precisely than do journals that publish articles from different fields and, thus, do not uniquely identify a specialty.

The analysis pivots on the comparisons between observed and randomized data obtained via a Configuration Null Model (CNM), commonly used in network science (Bollobás and Béla, 2001; Maslov and Sneppen, 2002). We create randomized data sets that are identical to the observed data in the number of papers, publication year, field, and the number of backward and forward citations of each paper but that differ in the recombination of backward citations. Secondly, we use different regression models to investigate the relationship between interdisciplinarity, novelty, and impact. We evaluate a measure testing whether a relationship with impact is significantly different from the one found using the indicators that adopt the random network (using the CNM). Finally, we adopt an external validation set for novelty based on the papers of Nobel Prize winners and on the lists of APS milestone articles.

We find that the indicators of interdisciplinarity capture the variable they intend to measure, whereas the indicators of novelty are hampered by several important issues. We show that NoveltyU fails to disentangle novelty from interdisciplinarity and that NoveltyW does not depend on the characteristics of the articles but is driven by the structure of the citations network. Moreover, NoveltyU is highly correlated with interdisciplinarity indicators and has the same effect on the impact of articles. In addition, we find that the two novelty indicators are inconsistent with one another since they return different sets of novel articles and assign a different score of novelty to the same article. By the same token, the two measures also fail at identifying as novel a number of articles that have led to discoveries awarded the Nobel Prize or that are considered groundbreaking by experts, or disagree on which of them should be considered novel.

We conclude that the identification of novelty as a recombination of knowledge is very sensitive to the choice of the indicator. Both the definitions of novelty, whether defining novelty as an atypical or new combination of knowledge, raise some concerns. We recommend that the policy evidence that is derived from these measures be carefully reassessed and that further effort be devoted to advancing the design of novelty indicators.

The article is structured as follows. The next section introduces our working definition of novelty and interdisciplinarity. Section 3 introduces the indicators of novelty and interdisciplinarity and clarifies the motivation of the paper. Section 4 explains the methodology and the research hypotheses. The data set is described in Section 5. Section 6 presents the results. The final section presents the conclusions, discusses some policy implications, and makes suggestions for further research.

2 Novelty and interdisciplinarity: definitions

The literature on novelty in science and technology explicitly follows or can easily be assimilated into the Schumpeterian tradition. The process that generates new knowledge is based on the recombination of existing knowledge and the process results in novelty when knowledge is combined "differently" or when "new combinations" are carried out (Schumpeter, 1934, pp. 65-66).

In what follows, we consider as novel the contributions that do not form part of the existing state of knowledge and are obtained from new ways of combining knowledge – i.e. the recombination of the same bits of knowledge can generate more that one novelty if the process of combination is new –, including new combinations of previously generated bits of knowledge (e.g. discoveries, papers, domains). We claim that novelty depends on what is recombined – bits of knowledge – and on how it is recombined – the process of recombination. For instance, the different manners (how is recombined) in which atoms (what is recombined) are bonded in carbon produce several allotropes: diamond, graphite, graphene, and fullerenes. Else, butane and isobutane, isomers with different biological and physical properties, are both combinations of four atoms of carbon and ten atoms of hydrogen (what is recombined) and differs only in the spatial structure of the molecule (how it is recombined). It follows that the approaches that identify novelty in the first appearance of a combination of knowledge bits (Wang et al., 2017) or in its rarity (Uzzi et al., 2013), at best, capture only a fraction of new contributions in that they look only at what is recombined (see Section 3) and neglect the differences in the recombination process. These differences are better appreciated ex-post, on the outcome of the recombination. In this case, the higher the distance from pre-existing outcomes the higher the novelty (Iori and Fontana, 2019; Kaplan and Vakili, 2015; Kelly et al., 2018).

Interdisciplinarity instead is strictly related to the characteristics of the recombined bits of knowledge. Interdisciplinarity is "a mode of research by teams or individuals that integrates information, data, techniques, tools, perspectives, concepts, and/or theories from two or more disciplines or bodies of specialized knowledge to advance fundamental understanding or to solve problems whose solutions are beyond the scope of a single discipline or area of research practice" (The National Academies, 2005). Interdisciplinarity, therefore, does not require that the process of recombination takes place differently and can be observed ex-ante. Thus, interdisciplinarity exists independently of novelty and it is not necessary nor sufficient to generate novelty. Iori and Fontana (2019) bring some evidence in this direction by showing that there is no correlation between the interdisciplinarity and novelty calculated on the outcome of recombination and that novelty can emerge from monodisciplinary research (see Section 3). This difference, to the best of our knowledge, is not explicitly discussed in the literature but it is already applied in funding policy. For instance, the evaluation guidelines of the Marie Sklodowska Curie Actions funded by the EU require as a necessary requisite the originality and innovative aspects of the research project. The interdisciplinary aspects of the action are evaluated separately, only if relevant.[1] In practice, these definitions imply that while interdisciplinarity is correctly reflected in the references of articles the same cannot be said for novelty. Novelty should be investigated in the content (e.g. text, topic) of papers.

This paper shows that NoveltyW and NoveltyU do not entirely cover the Schumpeterian definition of novelty and analyse the related measurement and conceptual drawbacks.

3 Background and Motivation

Both the novelty as the appearance of a new combination (Wang et al., 2017) and the novelty as atypical combination of knowledge (Uzzi et al., 2013) depend on the frequency of recombination and on the proximity of the involved knowledge but, in these two novelty measures, these elements are defined in rather different ways.

Wang et al. (2017) explore the relationship between novel research and impact for 661,643 papers published in the Web of Science in 2001 in all disciplines. They find that novel articles are more likely to rank in the top highly cited papers and to gain recognition in "foreign" fields. At the same time, they exhibit a longer citation lag, are less likely to be published in journals with a lower Impact Factor, and show a higher variance in the number of forward citations showing that novelty involves higher risks. For Wang et al. (2017), novel articles are those that make new combinations of referenced journals.

Novelty is computed for each paper as the sum of the distance of novel combinations that are found in backward citations:

$$Novelty_W = \sum_{i,j\ pairisnew} (1 - p_{ij}), \tag{1}$$

where i, j are the new pair of referenced journals and the term in parentheses is the distance (1 - proximity) between journals computed with the cosine similarity of co-citations – the number of co-citations is the frequency with which two journals are included in the backward citations of the same article.[2] According to equation 1, a paper is new when it makes at least one unprecedented

[1] Research Executive Agency, *MSCA-IF Evaluation step by step manual for evaluators 2018*. https://ec.europa.eu/info/sites/info/files/msca_if_2018_manual_for_evaluators.pdf

[2] The newness of pairs in the reference year – 2001 – is evaluated with respect to the previous 20 years, while the

combination of journals in its backward citations, and its novelty increases with the number of new combinations and the distance between the journals involved.

The definition of novelty as the first appearance of a combination raises some issues. On the one hand, it might underestimate novelty because, as argued in Section 2, novelty can also result from combinations of knowledge that have already appeared. On the other hand, it might overestimate novelty because it could capture extravagance in citation behavior. Moreover, from the operational viewpoint, the identification of novel articles is very sensitive to the unit of observation in backward citations (e.g. journals or classification codes). For instance, the first appearance in a paper's backward citations of an article in economics published in a multidisciplinary journal – say, *Science* – and of an article published in a more specialized journal – say, *Research Policy* – would signal that the paper making that combination is new. Any further combination between articles published in *Science*, no matter what discipline, and *Research Policy* would not be considered novel. The introduction of a new specialized journal instead would induce an increase of novelty even if the combinations are made with journals belonging to the same discipline. When using classification codes that uniquely identify disciplines and sub-disciplines, the results are more precise. However, the level of classification (e.g, digit) affects the chance of detecting new pairs, and thus of novel articles, which becomes more likely as the classification becomes more fine-grained.

Uzzi et al. (2013) analyze 17.9 million articles in all the disciplines included in the Web of Science (1950-2000) and find that novelty has a positive effect on impact (computed as the number of forward citations) only if it is counterbalanced by more traditional combinations of knowledge (i.e. Conventionality). Uzzi et al. (2013) consider the cumulative distribution of the proximity between pairs of journals in backward citations and heuristically define the degree of novelty for each article, NoveltyU, as the tenth percentile of such distribution to express the idea that novelty resides in more distant combinations of existing knowledge. Proximity is computed as the number of co-citations among journals normalized with a Configuration Null Model, whose details are explained in Section 4. The more frequent the occurrence of a pair of journals with respect to all the other occurrences in the sample, the higher the proximity of the recombined disciplines. Uzzi et al. (2013) also define an indicator of article's conventionality as the median of the proximity among journals in the article backward citations (see Figure 1).

proximity is computed with respect to the 3 preceding years.

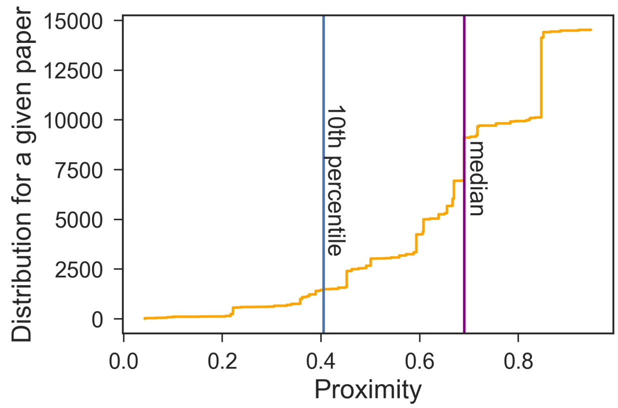


Figure 1: Illustration of NoveltyU calculated on our data (see Section 5). Novelty is the 10th percentile of the distribution of the proximity between pairs, computed as the number of co-citations normalized with the Configuration Null Model, in article backward citations. Conventionality is the median of such distribution.

Since this measure targets rare – and not only unprecedented – combinations, proximity is computed with respect to the pair frequency in the entire sample thereby solving the problems related to false negatives and positives that afflict NoveltyW. Moreover, in NoveltyU, novelty is based on the frequency of proximity between pairs and not on the frequency of the pairs themselves. This also mitigates the problem of sensitivity to the unit of observation since journals and disciplines that have similar co-citation patterns will also have a similar distance.

Both NoveltyW and NoveltyU are grounded on the integration of knowledge from different disciplines and therefore are tightly connected to the measurement of interdisciplinarity. This issue is evident also in literature on interdisciplinarity where NoveltyU is sometimes used as or listed within interdisciplinarity indicators (D'Este et al., 2019; Yegros-Yegros et al., 2015).

In this work, interdisciplinarity is measured with a set of well-established indicators that highlight diverse aspects of the integration of knowledge in articles: Variety, Balance, Disparity, and Integration Score (Porter and Rafols, 2009; Yegros-Yegros et al., 2015).[3]

Variety counts the number of different disciplines involved in knowledge production:

$$v = \sum_{i \in C} 1, \tag{2}$$

where C is the set of fields i in backward citations of the paper.

[3]For a survey on the origin and applications of these indicators see Stirling (2007) and Zeng et al. (2017, paragraph 6.1.1).

Balance refers to the evenness of their distribution and it is operationalized as the normalized Shannon entropy, returning a value between 0 and 1:[4]

$$b = \frac{1}{\log v} \sum_{i \in C} f_i \log f_i, \tag{3}$$

where f_i is the frequency of discipline i in backward citations.

Disparity measures the degree to which the involved disciplines are similar or different by introducing the notion of proximity to account for dissimilarity in integrated knowledge. The literature (Rafols et al., 2012; Wagner et al., 2011; Yegros-Yegros et al., 2015) measures Disparity as:

$$d = \frac{1}{v(v-1)} \sum_{\substack{i,j \in C \\ i \neq j}} (1 - p_{ij}), \tag{4}$$

where p_{ij} is the proximity between disciplines i and j measured as the cosine similarity of the number of co-citations. Disparity is defined for values between 0 and 1 and is independent of Variety and Balance.

The Integration Score, also known as the Rao-Stirling diversity index (Stirling, 2007), is defined as:

$$IS = \sum_{\substack{i,j \in C \\ i \neq j}}^{N} (1 - p_{ij}) f_i f_j. \tag{5}$$

It aggregates the previous dimensions of interdisciplinarity and returns values between 0 and 1. For all the measures higher values correspond to more interdisciplinary research.

As in the case with novelty, interdisciplinarity is often studied in relationship with impact intended as the number of forward citations. Results vary widely but the most recurrent pattern is an inverted U-shaped relationship between interdisciplinarity intensity and impact (Adams et al., 2007; Yegros-Yegros et al., 2015).[5]

[4] Notice that with this normalization the measure is independent of Variety.

[5] Steele and Stier (2000) compute Variety and Balance on 750 articles in forestry, finding a positive effect of interdisciplinarity on impact measured as average annual citation rate. Rinia et al. (2001) calculates Balance at the journal level for the publications of physicists in The Netherlands finding no effect of interdisciplinarity on impact. Levitt and Thelwall (2008) examine all science and social science articles indexed in the Web of Science (WoS) and Scopus computing Variety at the journal level. Their results posit a negative relationship between interdisciplinarity and impact in some disciplines. Larivière et al. (2015) measures interdisciplinarity through the Disparity of co-citations for all the papers published in WoS in the period 2000-2012 finding mainly a positive effect on impact. Wang et al. (2015) use Variety, Balance, and Disparity at the article level for all the papers published in WoS in 2001. They find that Variety and Disparity have a positive effect on impact while Balance has a negative sign.

Both interdisciplinarity and novelty indicators increase with the number and the distance of the combined disciplines. Their similarity is particularly evident in the indicator proposed by Uzzi et al. (2013). Intuitively, Conventionality, as the median of the distribution of proximity between fields in backward citations, is highly correlated with the Integration Score, defined as the weighted mean of the distance between the same fields. In parallel, NoveltyU, which measures the dispersion of the proximity (or distance) distribution, increases with the Disparity between fields since the latter measures their average distance.

To sum up, our analysis of the novelty measures leaves us with the following questions:

- How good are these indicators at identifying novelty as recombination of knowledge?
- Are novelty indicators able to distinguish between novelty and interdisciplinarity?

We exemplify these issues by analysing a set of articles associated with some key discoveries (see footnote 6) that have been awarded the Nobel Prize in physics (details on the data can be found in Section 5). The motivations for the awards explicitly refer to discoveries and new methods or tools thus we can reasonably assume that the content of those articles is new to our definition.[6]

We ask whether NoveltyU and NoveltyW identify Nobel articles as novel in the first place and, secondly, whether they return values that are higher than the average of the other articles published in the same year. Secondly, if novelty is an intrinsic property of the articles, NoveltyU and NoveltyW should lead to a similar evaluation. Finally, we should also observe some variation across papers in terms of interdisciplinarity. Figure 2, which reports NoveltyU, NoveltyW and Integration Score, as a compound indicator of interdisciplinarity, shows that for these 19 articles associated with 8 discoveries (see footnote 6 for abbreviations) none of the expectations is met and results clash with expert evaluations and the relevance in terms of technological developments of these discoveries. For instance, both novelty indicators fail at identifying as top new contributions the articles by Albert Fert and Peter Grünberg that were awarded the Nobel Prize in 2007 for the independent

[6] These Nobel Prizes have been awarded with the following motivations: (1997)"for development of methods to cool and trap atoms with laser light (CTA)", (2001) "for the achievement of Bose-Einstein condensation in dilute gases of alkali atoms, and for early fundamental studies of the properties of the condensates (BEC)", (2005) "for the contributions to the development of laser-based precision spectroscopy, including the optical frequency comb technique (LPS)", (2007) "for the discovery of giant magnetoresistance (GMR)", (2012) "for ground-breaking experimental methods that enable measuring and manipulation of individual quantum systems (IQS)", (2015) "for the discovery of neutrino oscillations, which shows that neutrinos have mass (NO)", (2016) "for theoretical discoveries of topological phase transitions and topological phases of matter (TP)", (2018) "for groundbreaking inventions in the field of laser physics, in particular, the optical tweezers and their application to biological systems (OT)".

discovery of the Giant Magnetoresistance effect (GMR). The GMR system makes it possible to retrieve densely packed digital data from hard disks and, having been translated into industrial-scale technology, has led to the development of smaller and thinner hard disks with increasing storage capacities. Revolutionizing the data storage industry, it has also found application in space science and engineering biology (Tian and Yan, 2013). Similarly, the articles by Steven Chu, Claude Cohen-Tannoudji, and William D. Phillips were awarded the Nobel Prize in 1997 for their methods to cool and trap atoms with laser light which have proven invaluable for advancing experiments in quantum physics. This latter case is particularly interesting because this discovery exhibits a low level of interdisciplinarity. This suggests that the recombination knowledge that generates novelty need not necessarily derive from distant disciplines.

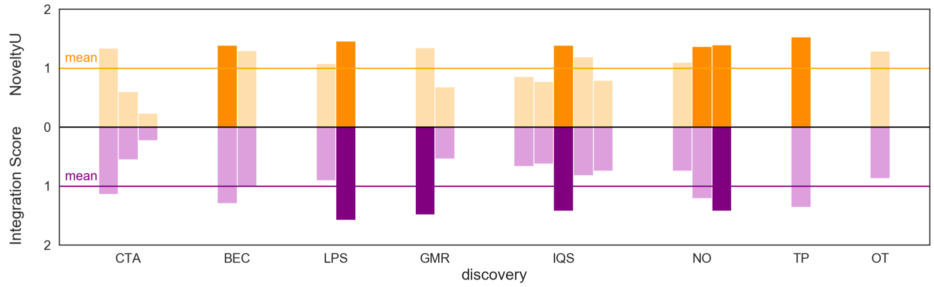
The figure also shows that NoveltyU and interdisciplinarity (Integration Score) have a very similar pattern, which suggests that the issue concerning the disentanglement of interdisciplinarity and novelty deserves further investigation.

This preliminary evidence corroborates our concerns and motivates an in-depth analysis of the functioning of novelty indicators.

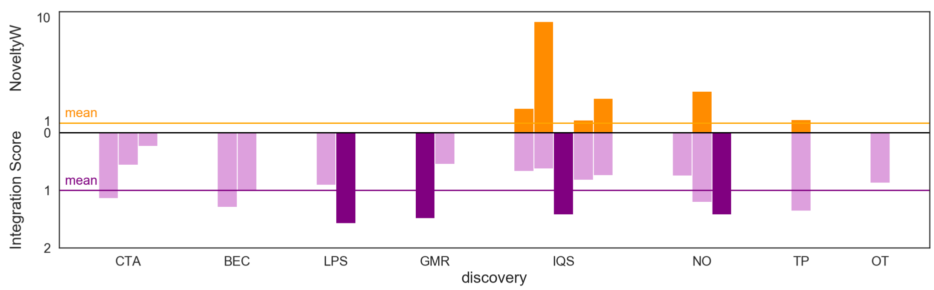
4 Hypothesis testing: Methodology

The analysis starts from the construction of a Configuration Null Model (CNM) (Bollobás and Béla, 2001; Maslov and Sneppen, 2002), a randomization method used in network science. The CNM is identical to the observed data in some key structural properties but is otherwise random. Namely, the randomized data is a citation network identical to the observed one in the number of papers, publication year, field, and the number of backward and forward citations for each paper, but differs in the recombination of backward citations.

The randomization of backward citations simulates a scenario in which researchers reference articles without any topic preference, resulting in backward citations that have no connection to the specific content of the articles. Since the indicators introduced in the previous section are based on backward citations, the randomized procedure assigns a random value of novelty and interdisciplinarity to articles, while keeping the degree of nodes and the years of citing and cited papers as observed in the data. Figure 3 illustrates the procedure for the generation of the



(a) Novelty as an atypical combination (NoveltyU) and interdisciplinarity (Integration Score).



(b) Novelty as first appearance of a combination (NoveltyW) and interdisciplinarity (Integration Score).

Figure 2: Novelty and interdisciplinarity of the 19 articles associated with 8 discoveries (see footnote 6 for abbreviations) awarded with the Nobel Prize in physics. Values have been normalized by the average value of the indicator in the publication year of the article. Darker bars indicate papers in the top 20% of the novelty and interdisciplinarity distribution of their publication year - a very broad notion of top indicator scores. Novelty indicators values are not systematically above the average value for the whole sample. According to NoveltyU most of the discoveries have a novelty value that is close to the mean of their year. NoveltyU identifies as top new contributions (darker bars) 6 articles and picks 5 different discoveries. According to NoveltyW, 13 articles have a novelty value of 0. NoveltyW identifies as top new contributions just 6 articles and picks 3 different discoveries. The indicators only agree on 2 discoveries. If we consider papers in the top 10% of novelty distributions to identify top new contributions, only 2 articles (2 discoveries) according to NoveltyU and only 3 articles (2 discoveries) according to NoveltyW are top new contributions. None of these articles lies in the top 1% of NoveltyU and NoveltyW distributions.

Configuration Null Model for the citation network among articles. Starting from the observed citation network (left), for all papers a, links are randomly redistributed while preserving for each node the number of incoming and outgoing links, the years, and the disciplines of the citing and cited papers c (Figure 3). This procedure, in contrast to a random selection from a distribution, preserves the structural properties of each element of the network and their time dependence, since the publication year of each article and the years of citations are the same as in the observed data.

We test whether the novelty and interdisciplinarity indicators calculated on the observed data behave differently from the ones obtained from the randomized data. The rationale behind this

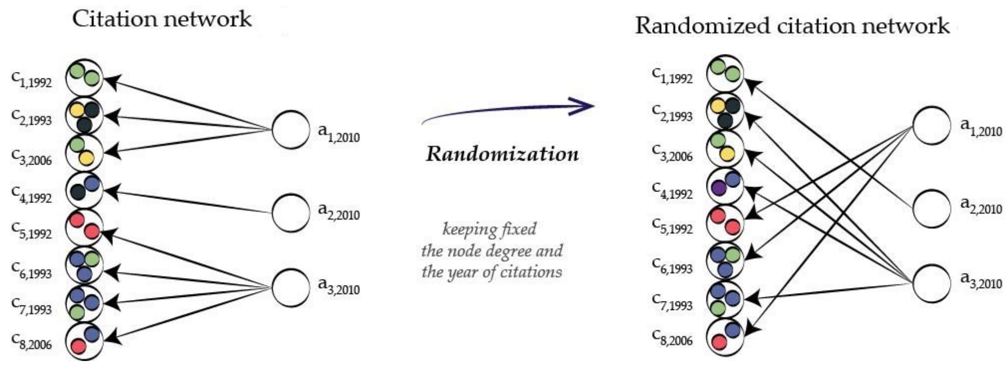


Figure 3: Configuration Null Model. The randomized citation network (right-hand side) is obtained by applying a CNM to the observed citation network (left-hand side). Citation randomization is performed fixing the years of citing (a) and cited (c) articles. Colored circles represent the disciplines of the papers in backward citations.

hypothesis is that a measure should capture novelty in the observed data. If the indicators return the same values in the observed and randomized data, then the measures are spurious: they depend on structural properties of the network (number of nodes and links, year and field of the papers, and number of backward citations for each paper), that have no relationship with the novelty and interdisciplinarity of the articles. Combining these insights we put forward the following hypotheses:

H1a: If an indicator captures novelty, then its distribution calculated on the set of observed backward citations is different from its distribution calculated over the same set of articles with randomized backward citations and preserved key structural network properties.

H1b: If an indicator captures interdisciplinarity, then its distribution calculated on the set of observed backward citations is different from its distribution calculated over the same set of articles with randomized backward citations and preserved key structural network properties.

We operationalize the test of H1a and H1b by comparing the distributions of the novelty and interdisciplinarity measures computed on the observed and on the randomized networks using the Kolmogorov-Smirnov and the Hellinger distance, and with the Kullback-Leibler divergence.

We extend the same reasoning to the relationship between the measurement of novelty or interdisciplinarity and impact. We assess the impact of a specific publication with the number of forward citations, its generality (G), and the number of different fields in forward citations (number of citing fields). Generality is calculated as:

$$G = 1 - \sum_{i \in C_f} f_i^2, \tag{6}$$

where C_f and f_i are respectively the set of disciplines and the frequency of discipline i in forward citations. Generality is widely used with patents to assess the characteristics of innovations (Hall et al., 2001).

The observed impact of each article is related to its observed and simulated novelty and interdisciplinarity indicators. If the relationship between the indicators and impact is the same in the simulated and observed scenarios, we conclude that, once again, the relationship with impact depends upon the invariant properties of the data and not on the precise measurement of novelty and interdisciplinarity.

We, therefore, test the following hypotheses:

H2a: Assuming that the article's novelty affects its impact, if an indicator captures novelty, the relationship between novelty and impact should be different (relatively to the one calculated on real data) when novelty is measured with a randomized set of backward citations that preserve the key structural properties of the network.

H2b: Assuming that the article's interdisciplinarity affects its impact, if an indicator captures interdisciplinarity, the relationship between interdisciplinarity and impact should be different (relatively to the one calculated on real data) when interdisciplinarity is measured with a randomized set of backward citations that preserve the key structural properties of the network.

To test H2a and H2b, we exploit a pooled sample that includes both observed and randomized data. In this setting, there are two observations for each article. The use of observed and randomized data refers only to the calculation of the indicators of novelty and interdisciplinarity that are used as covariates, while we maintain as a dependent variable, for each article, the impact observed in real data. We add a dummy variable that has a value equal to one for the observations that contain the observed data and has a value equal to zero for the observations that contain the randomized data. We run a regression analysis on the entire sample looking at the relationship between novelty or interdisciplinarity and impact. So, to test H2a and H2b, we look at the interaction term between the dummy variable and the measure of interdisciplinarity or novelty: if it is significantly different from zero this means that novelty and/or interdisciplinary affects the article's impact differently when we consider the observed data relative to the situation in which novelty and interdisciplinarity are measured with randomized data. Consequently, the adopted indicators provide a measurement of the underlying degree of novelty and interdisciplinarity and not just the structural properties of the network.

Our third hypothesis is the substantiation of our previous points through an independently established test set.[7] In particular, we focus on a sample of articles that have been unequivocally considered novel by the scientific community. We consider a set of articles related to researches that have been awarded the Nobel Prize and the APS milestone articles (see below Section 5 for details).

H3. If an indicator correctly captures novelty, it assigns a higher value to a set of articles unequivocally considered novel by the scientific community (external test set).

In order to test H3, we compare the mean values of the novelty indicators for the articles related to Nobel Prize winning research and for the APS milestone articles with the mean value of the novelty indicator for the entire sample. We run multivariate regression analyses using as dependent variables NoveltyU and NoveltyW and including dummies to identify papers that are related to researches that have been awarded the Nobel Prize and APS milestones. If the indicators correctly capture novelty, the coefficient associated with these dummy variables should be positive and significantly different from zero.

Our last point concerns the ability of the indicators to disentangle novelty from interdisciplinarity. Novelty and interdisciplinarity are distinct phenomena, so indicators should provide a distinct measure of these two key features of scientific research. So it is important to explore the size and nature of the correlation between the measures of novelty and interdisciplinarity and to discuss some implications. In order to understand the correlation patterns we proceed in three steps:

- We analyse the correlation between the indicators in real data and in the randomized network. So we ask whether the correlation structure depends upon the properties of the measures or the specific set of data.

- We ask which components of interdisciplinarity (Variety, Balance, Disparity, and Integration

[7] We are grateful to one of the referees for suggesting this extension.

Score) are correlated with NoveltyU and Conventionality. We clarify correlation patterns through a dimensionality reduction and perform a principal component analysis (PCA) to understand the structure of these relationships.

- We explore the articles related to Nobel Prize winning researches and the articles selected as APS milestones. In case of a strong correlation between novelty and interdisciplinarity, there are two possibilities: (a) if the novelty indicator really captures the novelty of the papers, all novel papers are also interdisciplinary. As discussed in the initial example, this would contradict our assumptions and the conventional wisdom in the field; (b) if the novelty indicator fails to fully capture the novelty of these papers, as suggested by Figure 2, this would mean that the novelty indicators measure interdisciplinarity and not novelty as recombination of knowledge.

Finally, it is also important to discuss some implications on previous research results and, in particular, on the main characteristics of high impact scientific articles. Both Wang et al. (2017) and Uzzi et al. (2013) use their novelty indicators to predict some forms of articles' impact (e.g. the number of citations received, generality, and the impact on different fields). So, we replicate their regression analysis and control whether we obtain the same results using interdisciplinarity indicators rather than novelty.

5 Data and preliminary evidence

We exploit a sample of articles in physics from the American Physical Society (APS) database.[8] The original database includes 577,870 articles published in the 13 APS journals from 1893 to 2015,

[8]Physics - like all other scientific fields (Jones, 2009) - is experiencing a progressive specialization and fragmentation that make communication between sub-fields as hard as the one between different disciplines. In the words of the President of the American Physical Society, John Hopfield, in 2007: "As physics began to be larger than the span of single individuals, and as the number of physicists working in an area became so large that no one personally knew all the major contributors even to his/her own field, let alone to physics as a whole" and "the symptom that the field is maturing and fragmenting is seen in the rather specific nature of the prizes and awards, which are carefully allocated to sub-fields of physics." (Hopfield, 2017). Physics also exhibits an increase in the age of first invention (Jones, 2009) and in teamwork (Wuchty et al., 2007). For the purpose of this study, we can picture physics as a cut down version of science where the ongoing processes are the same as the ones observed at a larger scale. Within-discipline interdisciplinarity is known and analyzed in literature as narrow interdisciplinarity. The interaction between sub-fields of a discipline is, in principle, less challenging than the interaction among disciplines at least in epistemological terms since the concepts, theories and/or methods are relatively similar in their presuppositions. However, even within a discipline, the sub-fields still refer to specialized and distinct domains (Huutoniemi et al., 2010; Kelly, 1996; Klein, 2005).

their citations in APS journals, and a 6-digit hierarchical classification (self-attributed) known as PACS. The PACS[9] is the subject classification system of the American Institute of Physics (AIP) for categorizing publications in physics and astronomy. It consists of 10 top-level categories that represent broad fields. Its hierarchical structure makes it possible to progressively identify more specific research areas as well as sub-fields up to five levels of successive specifications where each article can have a maximum of five PACS codes.[10]

The analysis is performed on a subset of the APS database. Since PACS are available from 1970 and widespread from 1985 and because we analyze a 10-year window of citations for each article, we focus on focal articles published between 1985 and 2005 with at least one reference and one PACS code in backward citations (more than 90% of articles in the database have a PACS code, see Figure A.1 in Appendix A.1).

The resulting database is composed of 230,854 focal articles (1985–2005), 203,910 referenced articles (1970–2005), 355,092 citing articles in the first 10 years from the publication of the focal articles (1985–2015), and 8 journals. The observed and the simulated citation networks have 246,935 nodes, 2,046,055 links, and an average in-degree of 8.4.

The set of Nobel Prizes and milestone articles contains 19 articles associated to Nobel winning discoveries and 56 articles that have made long-lasting contributions to physics (of which 14 are Nobel Prizes), either by introducing significant discoveries such as the discovery of high-temperature superconductivity (Wu et al., 1987) or quantum cryptography (Bennett et al., 1992; Ekert, 1991) and teleporting (Bennett et al., 1993), or by initiating new areas of research such as network science (Newman et al., 2001). These articles also have a high impact: they have an average number of forward citations equal to 198 (in the entire sample it is 11), an average number of citing fields - at the 4-digit level - equal to 48 (versus 9 in the entire sample), and an average generality of 0.85 (vs 0.68).[11]

We consider PACS at the 1-digit level (10 PACS), a 2-digit level (64 PACS), and at the 4-digit

[9] Further details about PACS are given in Appendix A.1.

[10] PACS are better proxies for sub-fields than other indicators. Referenced journals are quite imprecise since journals do not uniquely identify disciplines. Keywords assigned by authors have the advantage of a closer connection with the content of articles (Carayol et al., 2019). However, keywords are not standardized and the use of different terms to identify the same topic artificially expands the combination space. The use of codes (e.g. JEL in economics or PACS in physics) is possibly a more appropriate choice if their structure is stable over time.

[11] The articles are available at the following websites: Celebrating 125 years of The Physical Review, The 25th anniversary of Physical Review E, and Letters from the Past - A PRL Retrospective.

Table 1: Descriptive statistics.

(a) Descriptive statistics of articles per 1-digit PACS code.

1-digit PACS code	**Number of articles**	**Average number of backward citations**	**Average number of forward citations**
0 - General Physics (General)	54534	8.37	11.91
1 - The Physics of Elementary Particles and Fields (HEP)	31194	8.11	10.86
2 - Nuclear Physics (Nuclear)	20516	7.05	9.07
3 - Atomic and Molecular Physics (AMO)	25462	7.51	10.00
4 - Electromagnetism, Optics, Acoustics, Heat Transfer, Classical Mechanics, and Fluid Dynamics (Classical)	29358	7.97	10.60
5 - Physics of Gases, Plasmas, and Electric Discharges (Plasma)	6875	5.13	6.24
6 - Condensed Matter: Structural, Mechanical and Thermal Properties (CondMat I)	55014	7.69	9.01
7 - Condensed Matter: Electronic Structure, Electrical, Magnetic, and Optical Properties (CondMat II)	92933	9.41	11.58
8 - Interdisciplinarity Physics and Related Areas of Science and Technology (Interdisc)	23340	7.15	10.10
9 - GeoPhysics, Astronomy and AstroPhysics (Astro)	9885	8.97	13.74

(b) Descriptive statistics of articles per journal

Journal	**Number of articles**	**Average number of backward citations**	**Average number of forward citations**
PRA-Physical Review A	26363	8.92	9.40
PRB-Physical Review B	73748	10.43	9.45
PRC-Physical Review C	13433	8.10	8.35
PRD-Physical Review D	24690	9.79	11.73
PRE-Physical Review E	20838	8.57	7.01
PRL-Physical Review Letters	47638	7.46	19.29
PRSTAB-Physical Review Accelerators and Beams	405	4.71	5.53
RMT-Reviews of Modern Physics	276	55.56	98.81

level (846 PACS).

Tables 1a and 1b describe the composition of the sample.

The proximity of PACS at 2-digit level (see below for methodological details) in physics is represented in Figure 4. Edges represent co-occurrences in backward citations, the figure only includes co-occurrences that are higher than the average. Figure B.1 in Appendix B shows the value of proximity for all the levels of co-occurrence. Colors are assigned according to the 1-digit PACS and are used consistently throughout the figures of this section. Figure 4 shows that interaction of sub-fields is not homogeneous: Nuclear Physics, The Physics of Elementary Particles and Fields, and GeoPhysics, Astronomy and AstroPhysics have high proximity and so do Condensed Matter:

Structural, Mechanical and Thermal Properties and Condensed Matter: Electronic Structure, Electrical, Magnetic, and Optical Properties. It can be seen that General Physics co-occurs with all the fields.

Measures of novelty and interdisciplinarity NoveltyW is given by the sum of distances between new pairs of PACS in backward citations (Equation 1), while it is equal to zero for articles without new pairs of PACS (Wang et al., 2017).

To reproduce the Wang et al. (2017) setting, we create a buffer of 20 years within which we compute the occurrence of pairs of PACS in the backward citations of articles and we therefore compute Novelty$_W$ for the articles published in 2005.

In our data, the share of non-novel papers is around 88% (the same as in Wang et al., 2017), while in randomized data the share the novel articles amount to about 79%. Since this indicator is not bounded, in order to properly compare the randomized and observed samples, we scale the values between 0 and 1.

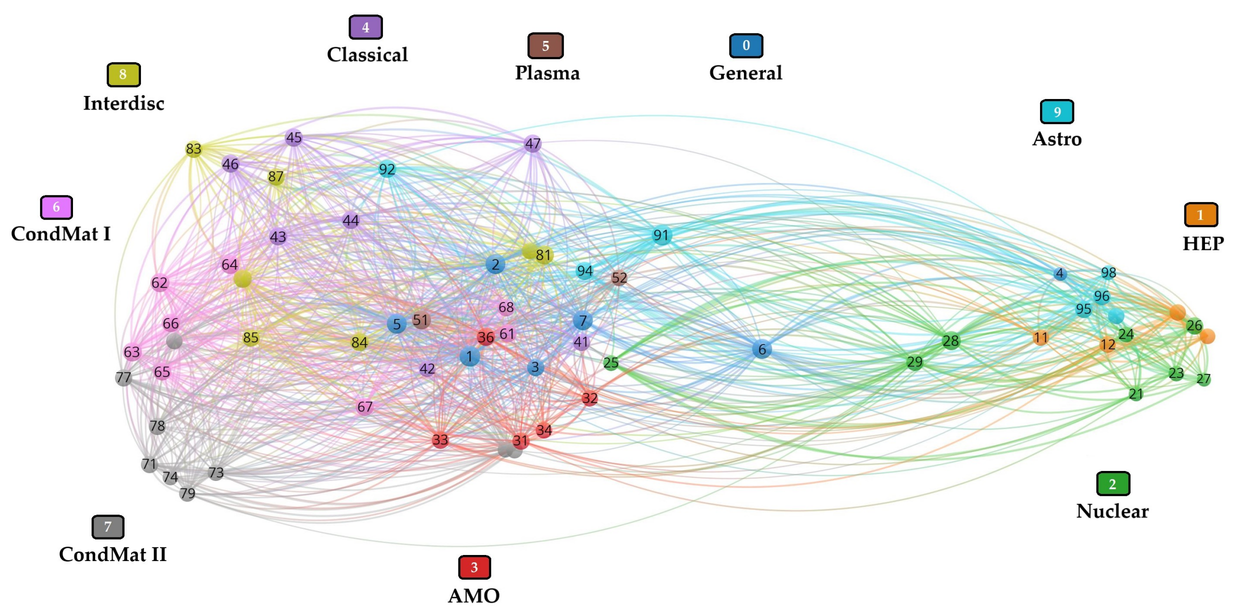


Figure 4: The proximity of PACS at 2-digit level (64 PACS). Colors indicate the sub-field of research at the 1-digit level (see Table 1a for details). The 1-digit PACS *Interdisciplinary Physics* (Interdisc) contains physics that does not belong to any other fields and the label does not refer to combinations of approaches from different fields. Its position in the proximity network confirms the results by Sinatra et al. (2015) and Battiston et al. (2019), where Interdisciplinarity Physics is mainly connected to Condensed Matter. The graph is built with VOSviewer.

NoveltyU is the 10th percentile of the distribution of the proximity of pairs between PACS in an article's backward citations and Conventionality is the median of the same distribution (Uzzi et al., 2013). To assign higher values to high-novelty articles, NoveltyU can be defined as 1 minus the tenth percentile of the proximity distribution:

$$Novelty_U = 1 - 10th\, percentile\, (F(p))\,, \tag{7}$$

where $F(p)$ is the cumulative distribution of proximity p between PACS pairs in backward citations of an article.

In the original papers by Wang et al. (2017) and Uzzi et al. (2013) the proximity between pairs is normalized, respectively with the cosine similarity and with the CNM.[12] In order to make results comparable we normalize proximity in both measures with the cosine similarity[13] because the application of the CNM for normalization purposes oversimplifies the analysis when applied to NoveltyW. Since the CNM is based on the co-citations of pre-existing pairs and NoveltyW considers unprecedented pairs, their CNM proximity is, by definition, equal to zero. It follows that NoveltyW is reduced to the count of new pairs. The cosine similarity, instead, compares the co-citation patterns between all fields and therefore returns positive values of proximity also for unprecedented pairs.

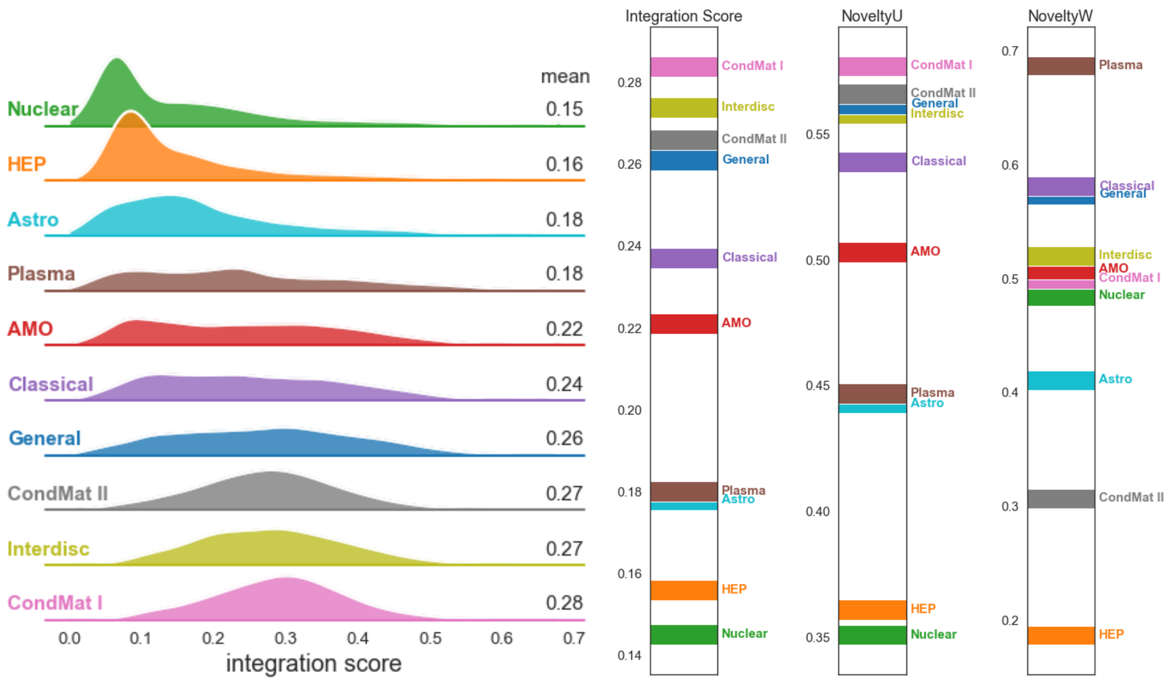
A preliminary investigation of our data shows that, even if we are referring to a single discipline, the sub-fields' tendency towards interdisciplinarity is quite variegated. Figure 5a shows the Integration Score, as a compound measure of interdisciplinarity, for the PACS at the 1-digit level. Sub-fields are ordered in increasing values of the measure. High Energy Physics (HEP), Nuclear Physics, and GeoPhysics and AstroPhysics (Astro) exhibit a low level of interdisciplinarity, whereas Condensed Matter I and II and Interdisciplinary Physics are more interdisciplinary. Figure 5b shows the average Integration Score, NoveltyU and NoveltyW per 1-digit PACS level. It

[12] It is worth noting that the values of proximity obtained with both methods are highly correlated (see Appendix B) and results are not significantly different in the two settings for interdisciplinarity indicators and NoveltyU.

[13] Cosine similarity is defined as

$$p_{ij}^{CS} = \frac{\sum_{k=1}^{N} c_{ik} c_{jk}}{\sqrt{\sum_{k_1}^{N} c_{ik}^2}\sqrt{\sum_{k_1}^{N} c_{jk}^2}}, \tag{8}$$

where N is the total number of PACS in the sample, c_{ik} is the vector of co-citations between PACS i and k and c_{jk} is the vector of co-citations between PACS j and k. The measure returns the cosine of the angle between two vectors of co-citations and it is bounded between 0 and 1.



(a) Probability distributions. (b) Average values.

Figure 5: Probability distributions of Integration Score and average values of Integration Score, NoveltyU, and NoveltyW by 1-digit PACS (for the extended PACS description see Table 1a). The connections among PACS codes, represented with the same colors, are shown in Figure 4. The Kendall's tau coefficient between NoveltyU and Integration Score is 0.85, between NoveltyU and NoveltyW is 0.11, and between NoveltyW and Integration Score is 0.18.

is worth noting that the ordering and the clustering of fields resulting from the Integration Score and NoveltyU are the same (apart from General Physics and Condensed Matter I). In contrast, NoveltyW generates results that are different from both NoveltyU and Integration Score. These patterns are confirmed by the Kendall's tau coefficient, a measure of rank correlation, which is high for NoveltyU and Integration Score (0.85) and low NoveltyU and NoveltyW (0.11), and for NoveltyW and Integration Score (0.18).

6 Results

The test of our hypotheses is conducted at the 4 and 2-digit classification for all the PACS and within a 1-digit PACS (CondMat II) at the 4-digit level. This strategy has a twofold objective. Firstly, it allows us to discuss the sensitivity of measures (NoveltyW and NoveltyU) to the level of aggregation of the specific fields (see Section 3). Secondly, when the analysis is possible at all levels, it shows that our results are robust to the level of aggregation of the PACS and, therefore, are scalable. Consequently, our results might reasonably apply also beyond the boundaries of physics.

As concerns the sensitivity of measures, we show that, at the 2-digit level, according to NoveltyW there are no novel papers in the year of interest and, therefore, we have to move at the 4-digit classification. In contrast, NoveltyU can be computed both at the 2 and 4-digit levels and the results are consistent. For these reasons, the analysis of the Nobel Prize and APS milestone articles is conducted at the 4-digit level.

We also test the robustness of our results by considering only the articles belonging to the largest sub-field of physics (CondMat II) and replicating the entire analysis within this PACS. We select this subset of articles, as physics is a subset of science, to suggest that our conclusions are not specific to physics, but depend on the properties of the novelty indicators. Since the results are totally confirmed also within a 1-digit PACS, we can conclude that the latter is the case and that our conclusions are general and mainly driven by the definition of these indicators.

In what follows, we present the results at the 4-digit level whereas the details concerning the 2-digit level and CondMat II can be found, respectively, in Appendix C and Appendix D.

6.1 Novelty and interdisciplinarity: observed vs. randomized data

The first part of our investigation tests whether the indicators of novelty and interdisciplinarity truly capture the properties of papers. To do so, we first create a baseline by measuring the same indicators in the randomized version of the citation network among papers. We then compare the measures in the observed and randomized data (H1a and H1b). Then we analyze the relationship between the indicators and the article impact (H2a and H2b).

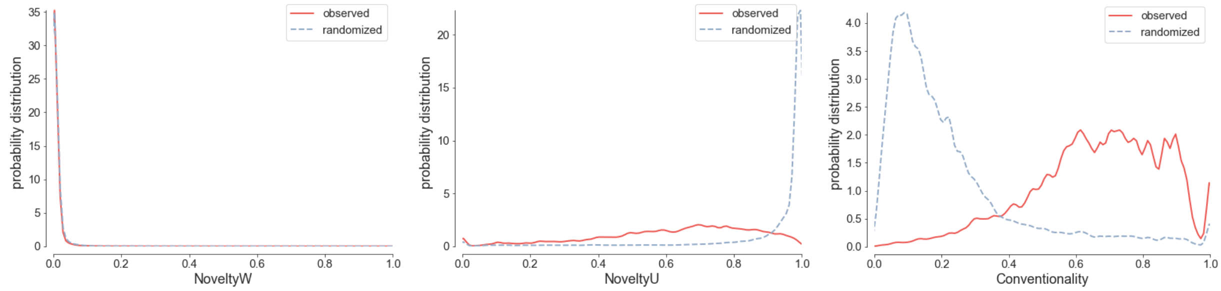
6.1.1 Testing H1a and H1b through the distribution of novelty and interdisciplinarity

We randomize the citation network by randomly selecting pairs of directed links and swapping the nodes they point to (see Figure 3). This randomization preserves the number of backward and forward citations of each paper, while it randomizes the distribution of paper attributes and removes the correlation of attributes among pairs of cited-citing papers. Since the journal and PACS numbers are a paper's attributes, all the measures based on these attributes are also randomized. This randomization generates a higher variety of journals and PACS recombinations than in the observed data and, as a consequence, different distributions of the measures based on the recombinations.

Taken together, in the randomized data we expect a higher number of articles with NoveltyW greater than zero, a higher NoveltyU, and lower Conventionality than in the observed data. Similarly, for interdisciplinarity indicators, we expect a higher average level of interdisciplinarity in the randomized sample. That is, we expect higher Variety, Balance to be close to 1, and Disparity to be close to the average distance between all the PACS.

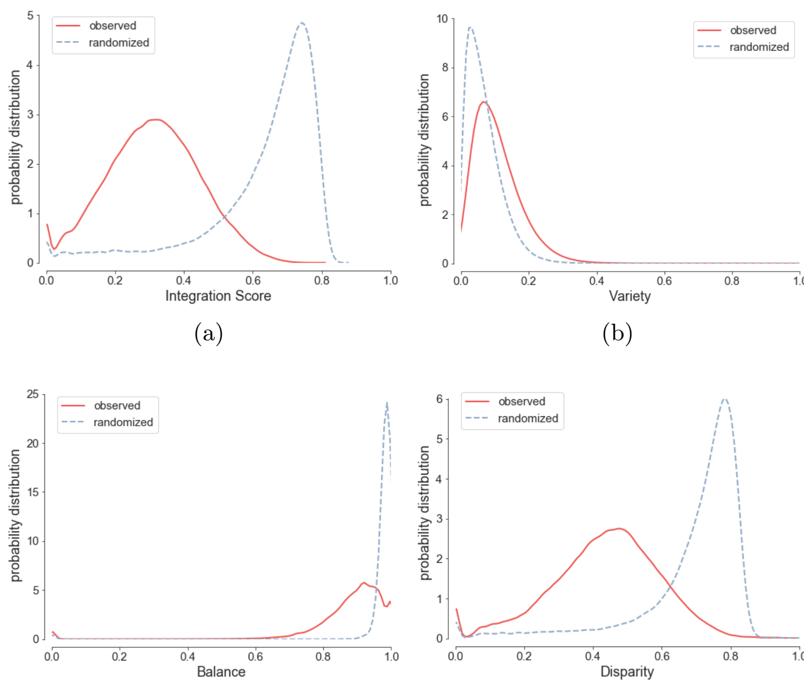
When we proceed with the comparison, we find that 21% of articles have NoveltyW greater than zero in the randomized sample versus 12% in the observed one. Given the relevance of zeros in determining the skewness of the distributions, we expected very different distributions of NoveltyW between the two sets of data. However, the observed and randomized distributions of NoveltyW are unexpectedly similar (Figure 6a). The other indicators exhibit pronounced differences between observed and randomized distributions (Figures 6b, 6c, and 7), confirming our expectations.

While the distributions of all indicators are not identical in the randomized and observed data, Figures 6 and 7 show a clear qualitative difference between NoveltyW and the other indicators: the distribution of NoveltyW in the randomized data is extremely close to that of the observed data. To capture quantitatively the difference between distributions in randomized and observed data, we use three different methods: (i) the Kolmogorov-Smirnov statistic (K-S), measuring the maximum distance between two observed cumulative distribution functions, (ii) the Hellinger distance, measuring the overall distance between two probability distributions as captured by the Euclidean norm of the difference of the distributions, and (iii) the Kullback-Leibler (KL) divergence, an information theory measure, quantifying the amount of information needed to reconstruct a distribution from a different given distribution. While we can test the significance of the K-S statistic, for the Hellinger distance and the KL divergence we observe and compare the absolute values across measures. These measures show that the distributions of all indicators are different in randomized and observed data (Table 2). However, the distances between randomized and observed are substantially lower (at least one order of magnitude) for NoveltyW than for all other indicators, indicating that NoveltyW mainly captures heterogeneities of the citation network rather than the properties of the single papers. Taken together H1a is confirmed for NoveltyU and Conventionality, but cannot be fully confirmed for NoveltyW. Notwithstanding the K-S statistic is significantly different from zero, the



(a) (b) (c)

Figure 6: Probability distributions of novelty indicators in observed (solid line) and randomized (dashed line) data.



(a) (b) (c) (d)

Figure 7: Probability distributions of interdisciplinarity indicators in observed (solid line) and randomized (dashed line) data. Variety is normalized between 0 and 1.

other two indicators for NoveltyW are very small both in absolute and relative value as compared with those of the other measures: the measured difference between the observed and randomized distribution seems to be a second-order effect, caused by the unavoidable noise of observed data (Gonzalez et al., 2008; Sinatra et al., 2016). We will conduct further investigation into this issue in the next section. Finally, H1b is confirmed for all interdisciplinarity indicators.

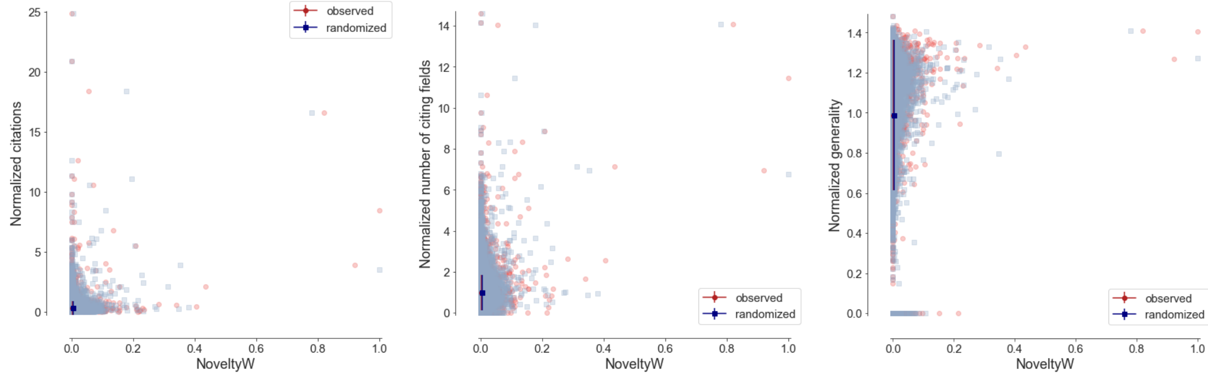
Table 2: Distances between randomized and observed distributions of novelty and interdisciplinarity indicators. K-S statistics are significantly different from zero for all indicators (p-values < 0.001).

	NoveltyW	NoveltyU	Conventionality	Integration Score	Variety	Balance	Disparity
K-S statistic	0.09	0.83	0.76	0.76	0.32	0.74	0.71
Hellinger distance	6.31	222.31	150.23	101.65	31.28	59.75	85.07
KL divergence	0.03	2.02	1.73	1.73	0.23	3.21	1.48

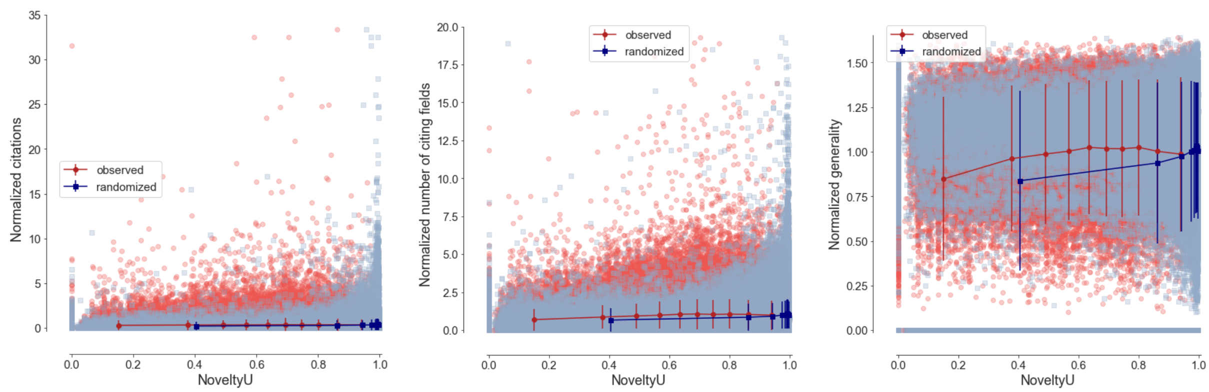
6.1.2 Relationship with impact (testing H2a and H2b)

Both Wang et al. (2017) and Uzzi et al. (2013) apply their indicators to the study of the effect of novelty on impact. We, therefore, move in the same direction observing whether the indicators computed on the simulated and observed data have different effects on the impact of articles. Figure 8 reports the relationship between novelty and impact. In order to account for differences in time and PACS, our measures of impact (the number of forward citations, the number of citing fields, and the generality of knowledge) have been normalized by year and 1-digit PACS. We observe that for NoveltyW (Figure 8a) the scatter plots overlap, signalling that the relationship between novelty and impact could be the same when NoveltyW is measured on the real data or with a randomized set of backward citations. Figure 8b instead shows that for NoveltyU the scatter plots are different suggesting that the relationship between impact and NoveltyU is not the same when the indicator is calculated on random backward citations. The same holds for Conventionality (Figure 8c).

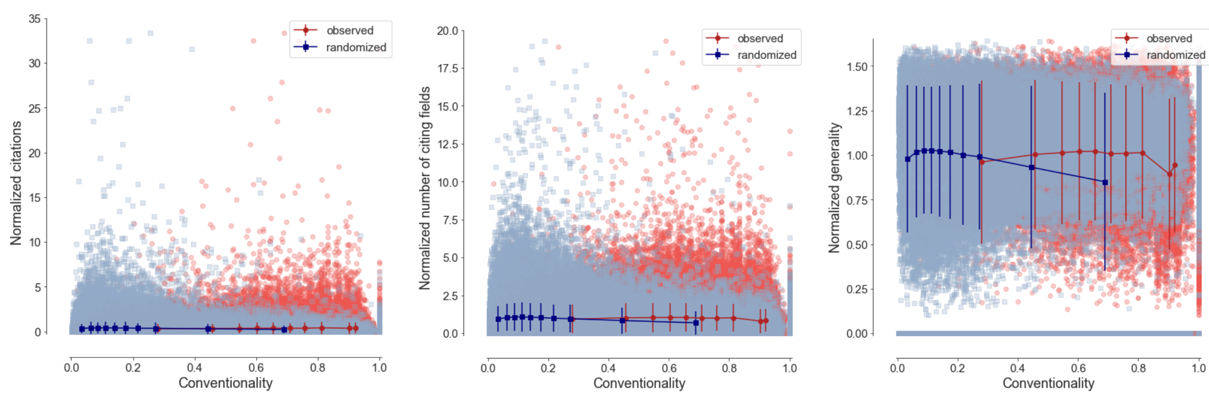
To further test the (di)similarities of these relationships, as explained in Section 4, we run a set of regressions (Negative Binomial when the dependent variable is the number of citations, as in Wang et al., 2017, and the number of citing fields, or OLS in case of generality), on the joint sample of randomized and observed data to analyze the effect of novelty and interdisciplinarity on impact. We include a set of dummy variables for fields (1-digit PACS codes), journals, and years. In addition, we control for the number of authors, the number of backward citations, and the presence of an international team (dummy), since these article characteristics are likely to positively affect the heterogeneity of topics discussed in an article and therefore its number of forward citations (Lee et al., 2015). Finally, when the articles cite just one field, Variety is equal to 1 and the other indicators are not defined. Therefore, we treat these articles with a separate intercept introducing a dummy variable as a control (Variety = 1).



(a) NoveltyW.



(b) NoveltyU.



(c) Conventionality.

Figure 8: Relationship between novelty indicators and our indicators of impact in observed and randomized data. Blue and red lines represent the mean and the standard deviation of impact in papers binned by quantiles (ten bins). Bins of observed and randomized data are not aligned due to the differences in the underlying distributions of novelty indicators (see figure 6).

Table 3: Differences in the relationship between NoveltyW and impact in observed and randomized data.

	Dependent variable:		
	Number of citations	Number of citing disciplines	Generality
	Negative binomial	*Negative binomial*	*OLS*
	(1)	(2)	(3)
Novelty_W	1.255***	0.692**	−0.078
	(0.475)	(0.350)	(0.117)
Novelty_W*Observed	−1.033	0.548	0.217
	(0.631)	(0.462)	(0.156)
log(Number of backward citations)	−0.144***	−0.334***	−0.134***
	(0.042)	(0.033)	(0.009)
log(Number of authors)	0.080***	0.038***	0.018***
	(0.012)	(0.009)	(0.003)
International	0.120***	0.096***	0.029***
	(0.008)	(0.006)	(0.002)
Variety = 1	0.491***	0.296***	0.068***
	(0.008)	(0.006)	(0.002)
Constant	0.727***	1.000***	0.377***
	(0.032)	(0.024)	(0.007)
1-digit PACS	Yes	Yes	Yes
Journal	Yes	Yes	Yes
Year	Yes	Yes	Yes
Observations	35,300	35,300	35,300
R^2			0.126
Adjusted R^2			0.126
θ	1.022***	1.987***	
Akaike Inf. Crit.	229,233	225,439	
Residual Std. Error			0.249
Note:			*p<0.1; **p<0.05; ***p<0.01

Table 3 reports the regressions for NoveltyW. The effect of NoveltyW on the number of forward citations and the number of citing fields, as in Wang et al. (2017), is positive and significant, while its effect on the generality of knowledge is not significant. However, in any of these three cases, the relationship between NoveltyW and impact is the same in randomized and observed data. This can be observed in Table 3 where we identify observed data with a dummy variable and we study the interaction between this dummy variable and NoveltyW. The interaction term NoveltyW*Observed is not significantly different from zero, suggesting that the estimated impact of NoveltyW does not actually measure the effect of the intrinsic novelty of papers but captures the structure of the citation network instead.

It is worth noting that, in this perspective, Wang et al. (2017) show that novel articles display higher variance in impact. This result could depend upon the heterogeneity of articles rather than on their novelty.

Table 4 reports the same analysis for NoveltyU and Conventionality. As opposed to NoveltyW, the interaction between the dummy variable and the indicators is significant, signaling that the

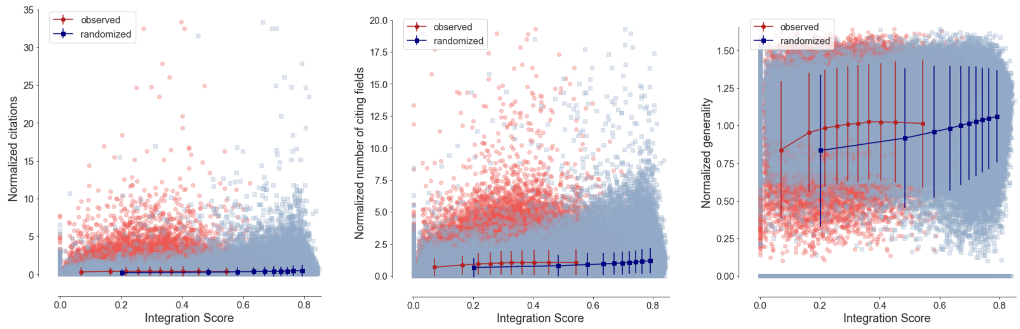
Table 4: Differences in the relationship between NoveltyU or Conventionality and impact in observed and randomized network.

	Dependent variable:					
	Number of citations		Number of citing fields		Generality	
	Negative binomial		*Negative binomial*		*OLS*	
	(1)	(2)	(3)	(4)	(5)	(6)
NoveltyU	−0.138***		0.075***		0.016***	
	(0.008)		(0.006)		(0.002)	
NoveltyU*Observed	−0.062***		0.045***		0.010***	
	(0.005)		(0.004)		(0.001)	
Conventionality		0.089***		0.003		−0.017***
		(0.013)		(0.010)		(0.003)
Conventionality*Observed		−0.023**		−0.189***		0.007***
		(0.010)		(0.014)		(0.002)
log(Number of backward citations)	0.472***	0.463***	0.285***	0.295***	0.070***	0.071***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.001)	(0.001)
log(Number of authors)	0.125***	0.126***	0.099***	0.099***	0.031***	0.031***
	(0.003)	(0.003)	(0.002)	(0.002)	(0.001)	(0.001)
International	0.037***	0.037***	0.019***	0.019***	0.009***	0.009***
	(0.004)	(0.004)	(0.003)	(0.003)	(0.001)	(0.001)
Variety=1	−0.151***	−0.083***	−0.082***	−0.100***	−0.055***	−0.062***
	(0.015)	(0.015)	(0.012)	(0.011)	(0.003)	(0.003)
Constant	1.633***	1.502***	1.352***	1.407***	0.457***	0.476***
	(0.015)	(0.015)	(0.011)	(0.011)	(0.004)	(0.003)
1-digit PACS	Yes	Yes	Yes	Yes	Yes	Yes
Journal	Yes	Yes	Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes	Yes	Yes
Observations	461,708	461,708	461,708	461,708	461,708	461,708
R^2					0.110	0.110
Adjusted R^2					0.110	0.110
θ	0.883***	0.882***	1.689***	1.690***		
Akaike Inf. Crit.	3,001,516	3,001,745	2,938,586	2,938,444		
Residual Std. Error					0.264	0.264
Note:						*p<0.1; **p<0.05; ***p<0.01

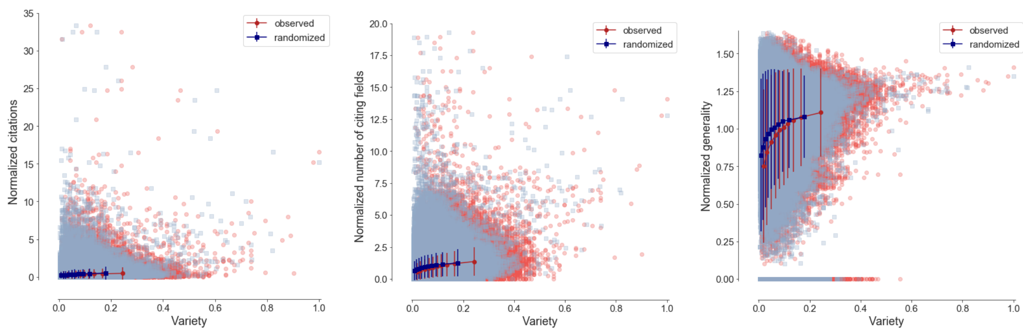
relationship with impact differs in observed and randomized data, as already evident from Figures 8b and 8c. Given the large difference in the distributions, the relationship between the indicators and the impact have different patterns in observed and randomized data. These indicators capture features of papers and do not uniquely depend on the structure of the citation network.

We replicate the analysis for interdisciplinarity indicators. Following the interdisciplinarity literature, we consider the three dimensions of interdisciplinarity (Variety, Balance, and Disparity) and the compound indicator (Integration Score). Figure 9 shows the scatter plots between Integration Score, Variety, Balance, or Disparity and the different measures of impact. As for NoveltyU and Conventionality, the plots reveal differences between the observed and randomized sample in the distribution of interdisciplinarity indicators and their relationship with impact. Unsurprisingly, the scatter plots display a higher level of similarity only in the case of Variety.[14] So, as reported in

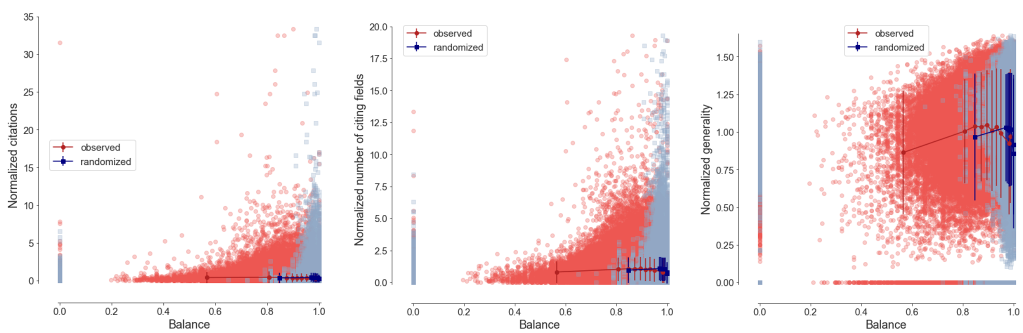
[14] The number of cited fields is, as expected, strongly correlated (77%) with the number of backward citations. For this reason, the number of backward citations is not included in regressions with Variety, which is one of the main



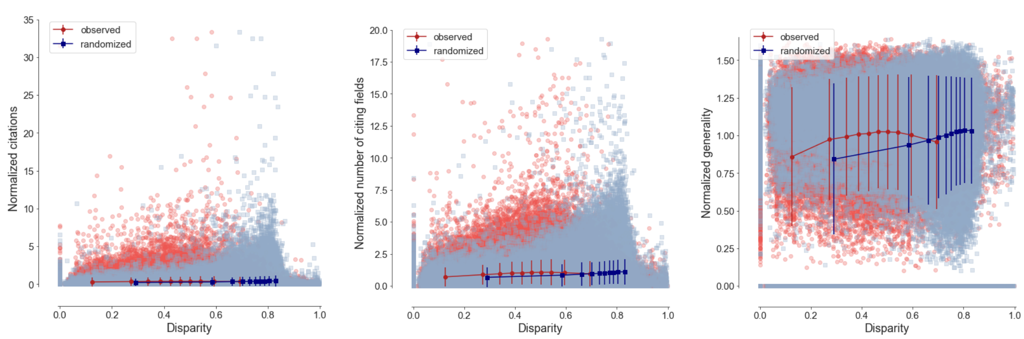
(a) Integration score.



(b) Variety.



(c) Balance.



(d) Disparity.

Figure 9: Relationship between interdisciplinarity indicators and our indicator of impact in observed and randomized data. Blue and red lines represent the mean and the standard deviation of impact in papers binned by quantiles (ten bins). Bins of observed and randomized data are not aligned due to the differences in the underlying distributions of interdisciplinarity indicators (see Figure 7).

Table 5: Differences in the relationship between interdisciplinarity indicators and impact in observed and randomized data.

	Dependent variable:											
	Number of citations				Number of citing fields				Generality			
	Negative binomial				*Negative binomial*				*OLS*			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Variety	0.024***	0.025***	0.025***		0.015***	0.017***	0.017***		0.004***	0.004***	0.004***	
	(0.0002)	(0.0002)	(0.0002)		(0.0001)	(0.0001)	(0.0001)		(0.00004)	(0.00004)	(0.00004)	
Variety*Observed	0.018***				0.023***				0.006***			
	(0.0003)				(0.0002)				(0.0001)			
Balance	−1.902***	−2.512***	−2.608***		−0.139***	−0.781***	−0.656***		0.051***	−0.093***	−0.073***	
	(0.029)	(0.027)	(0.029)		(0.022)	(0.021)	(0.022)		(0.007)	(0.006)	(0.007)	
Balance*Observed		0.102***				0.207***				0.057***		
		(0.005)				(0.004)				(0.001)		
Disparity	0.134***	0.075***	−0.042***		0.224***	0.224***	0.014		0.042***	0.048***	−0.009***	
	(0.012)	(0.013)	(0.011)		(0.009)	(0.010)	(0.008)		(0.003)	(0.003)	(0.003)	
Disparity*Observed			0.059***				0.303***				0.075***	
			(0.009)				(0.007)				(0.002)	
Integration Score				−0.179***				0.153***				0.041***
				(0.009)				(0.007)				(0.002)
Integration Score*Observed				−0.206***				0.181***				0.048***
				(0.011)				(0.008)				(0.002)
log(Number of backward citations)				0.475***				0.278***				0.068***
				(0.002)				(0.002)				(0.001)
log(Number of authors)	0.136***	0.133***	0.133***	0.124***	0.107***	0.105***	0.106***	0.099***	0.034***	0.033***	0.033***	0.031***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.002)	(0.002)	(0.002)	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)
International	0.037***	0.039***	0.039***	0.037***	0.018***	0.020***	0.019***	0.019***	0.009***	0.010***	0.010***	0.009***
	(0.004)	(0.004)	(0.004)	(0.004)	(0.003)	(0.003)	(0.003)	(0.003)	(0.001)	(0.001)	(0.001)	(0.001)
Variety = 1	−2.090***	−2.695***	−2.884***	−0.133***	−0.322***	−0.891***	−0.913***	−0.067***	−0.037***	−0.159***	−0.179***	−0.049***
	(0.030)	(0.031)	(0.031)	(0.015)	(0.023)	(0.023)	(0.024)	(0.012)	(0.007)	(0.007)	(0.007)	(0.003)
Constant	3.551***	4.132***	4.314***	1.610***	1.574***	2.116***	2.135***	1.346***	0.434***	0.549***	0.569***	0.454***
	(0.030)	(0.031)	(0.031)	(0.014)	(0.023)	(0.023)	(0.024)	(0.011)	(0.007)	(0.007)	(0.007)	(0.003)
1-digit PACS	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Journal	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	461,708	461,708	461,708	461,708	461,708	461,708	461,708	461,708	461,708	461,708	461,708	461,708
R^2									0.106	0.098	0.097	0.111
Adjusted R^2									0.106	0.098	0.097	0.111
θ	0.867***	0.862***	0.862***	0.883***	1.688***	1.656***	1.654***	1.691***				
Akaike Inf. Crit.	3,009,266	3,011,722	3,011,991	3,001,398	2,938,851	2,945,355	2,945,786	2,938,178				
Residual Std. Error									0.265	0.266	0.266	0.264

Note: *p<0.1; **p<0.05; ***p<0.01

Table 5, the estimated coefficients of the interaction terms (Variety*Observed, Balance*Observed, Disparity*Observed and, finally, Integration Score*Observed) show that the effect of all the interdisciplinarity indicators (including Variety) on impact is significantly different in the observed and randomized data.

In conclusion, we reject H2a for NoveltyW, while we confirm it for NoveltyU and Conventionality. Additionally, we confirm H2b for all interdisciplinarity indicators.

6.2 Novelty of Nobel Prizes and APS milestone papers (H3)

We compute NoveltyW and NoveltyU for articles related to researches awarded with Nobel Prizes and APS milestones to study the ability of these indicators to correctly capture novelty. Since we do not have, for all papers, a buffer of 20 years to identify the first appearance of a pair, we overestimate the value of NoveltyW for these articles. The average NoveltyW is slightly higher for Nobel Prize and APS milestone articles (0.0016 vs 0.0010), however, 44 out of 61 of these articles have zero NoveltyW (72%). When we look at the distribution of NoveltyW of the articles by year of publication, we find that 16 articles are in the top 20% and there is only one paper in the top 1% of NoveltyW of its cohort. We find no differences in the average value of NoveltyU in the Nobel Prize and APS milestones set and the entire database (0.62 in both samples). Only 16 papers are in the top 20% of NoveltyU (26% of the 61 articles), while there are no papers in the top 1% of the indicator. Moreover, the correlation between NoveltyW and NoveltyU is very low (25%) and only 5 papers are in the top 20% of both measures.

To verify H3 we test whether the mean values of NoveltyU and NoveltyW are different from the average value of the sample. We, therefore, run a simple multivariate analysis using as dependent variables NoveltyU and NoveltyW and including a dummy which is equal to one when the article belongs to a set of papers related to Nobel Prize winning researches. We perform the same exercise including a dummy which is equal to one when the article belongs to the set of APS milestones. Table 6 reports the results. Columns (1) refer to the regressions without additional controls, while Columns (2) refer to regressions that include a set of control variables (see Appendix E for the details). Table 6 shows the estimated values of the dummy variables and their p-values. If they are positive and statistically different from zero, it means that on average NoveltyU and NoveltyW

structural properties of the citation network.

Table 6: Nobel Prize and milestones effect on NoveltyW, NoveltyU, and Integration Score. (1) unconditional mean, (2) conditional mean. P-values in parenthesis. Full specifications in Table E.1.

	Nobel Prize (1)	Nobel Prize (2)	Milestones (1)	Milestones (2)
NoveltyW	0.071	-0.421	0.418	0.280
	(0.907)	(0.474)	(0.243)	(0.414)
NoveltyU	0.039	-0.011	-0.008	-0.019
	(0.455)	(0.817)	(0.790)	(0.479)
Integration Score	-0.017	-0.049	0.004	-0.010
	(0.594)	(0.071)	(0.805)	(0.543)

for the selected set of papers are larger than the sample average (i.e. the constant). Table 6 shows that this is never the case (we have performed the same exercise for the Integration Score and we discuss the results in Section 6.3.1). Hence, we reject H3 and show that the novelty indicators considered do not assign a higher value to a set of articles unequivocally considered novel by the scientific community.

6.3 Novelty or interdisciplinarity?

Given the analogies in the operalization of interdisciplinarity and novelty as an atypical combinations[15] discussed in Section 3, we ask whether NoveltyU in its two dimensions, Novelty and Conventionality, conveys information that is different from the one contained in interdisciplinarity measures.

We first compute the correlation between NoveltyU, Conventionality, and interdisciplinarity indicators. As expected, both NoveltyU and Conventionality are highly correlated with the interdisciplinarity indicators that consider proximity in their definition (the full correlation table can be found in Appendix A.2). In particular, NoveltyU overlaps significantly with Disparity (the correlation is 91.2%); there is also a very high correlation between NoveltyU and Integration Score (88.7%). In addition, the correlation between Conventionality and Integration Score is -87.6%. These correlations are also very high in the randomized dataset[16] supporting the idea that they depend upon the properties of the measures, rather than the specific features of the dataset.

[15]This issue is not accounted for in the literature, however, NoveltyU is sometimes used as a measure of interdisciplinarity, as in D'Este et al. (2019) and Yegros-Yegros et al. (2015).

[16]NoveltyU - Disparity: 93.1%; Conventionality - Integration Score: -91.6%; NoveltyU - Integration Score: 90.3%.

Table 7: Results of the principal component analysis on Novelty$_U$, Conventionality, and interdisciplinarity indicators.

(a) Principal components

	Components		
	RC1	RC2	RC3
Variety	0.20	0.97	0.02
Balance	0.23	0.02	0.97
Disparity	0.93	0.18	0.09
Integration Score	0.92	0.25	0.20
Conventionality	-0.87	0.04	-0.25
Novelty$_U$	0.89	0.27	0.14

(b) Importance of the three principal components

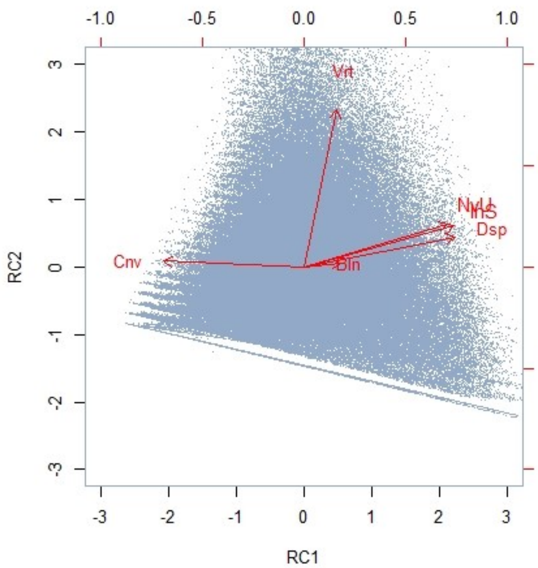
	Components		
	RC1	RC2	RC3
Proportion of variance	0.56	0.19	0.18
Cumulative variance	0.56	0.74	0.92
Proportion explained	0.60	0.20	0.19
Cumulative proportion	0.60	0.81	1.00

To further clarify these correlation patterns we report in Table 7 the results of a principal component analysis with orthogonal rotation (varimax) performed on the six variables (centered and scaled). The first three components explain more than the 92% of the total variance. Figure 10a shows how strongly the different variables influence the first two principal components (RC1 and RC2), whereas Figure 10b shows the relative importance of each variable in the definition of RC1, RC2, and RC3. As expected, RC1 – which explains the 56% of the total variance – combines, almost equally, Disparity, Integration Score, Novelty$_U$, and Conventionality (with the opposite sign). Variety and Balance, instead, are predominant, respectively, in RC2 and RC3.

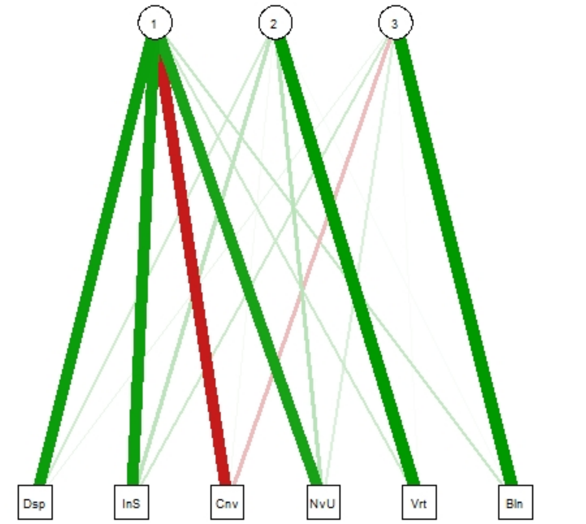
The principal component analysis confirms that, while Variety and Balance may capture different – but related – aspects of interdisciplinarity, the measures of novelty (Novelty$_U$ and Conventionality) strongly overlap with the interdisciplinarity indicators based on proximity (Disparity and Integration Score).

6.3.1 Interdisciplinarity and novelty in Nobel Prizes and APS milestone articles

In order to corroborate this evidence, we exploit the articles related to researches awarded with a Nobel Prize and the ones selected as APS milestones presented in Section 5 to suggest that novelty indicators measure interdisciplinarity and not novelty. In the Section 3 we have demonstrated that novelty indicators do not capture the novelty of these papers. As a consequence also their level of interdisciplinarity should not differ from the average of the entire sample since they are very much correlated with interdisciplinarity. We show here that this is the case.



(a) PCA biplot.



(b) Correlation between variables and PCA components. Edge width increases with correlation. Green and red edges indicate, respectively, positive and negative correlation.

Figure 10: Principal component analysis on Novelty$_U$, Conventionality, and interdisciplinarity indicators. Variables' abbreviation: Variety (Vrt); Balance (Bln); Disparity (Dsp); Integration Score (InS); NoveltyU (NvU); Conventionality (Cnv).

More specifically the level of interdisciplinarity (measured as Integration Score) of the Nobel Prizes and APS milestones is similar to the one of the entire sample (0.31 on average). 17 articles out of 61 (the 28%) are in the top 20% of Integration Score and only one paper is in the top 1% of this indicator. The regression analysis reported in Table 6 confirms that the average interdisciplinarity (Integration Score) of articles awarded with the Nobel Prize or recognized as APS milestones is not significantly different from the one observed in the entire sample. An exception is the conditional mean of the Integration Score of the Nobel Prize articles. In this case, it is slightly lower than the sample average, and this difference is statistically different from zero. This result confirms the idea that Nobel Prizes in physics do not have a high degree of interdisciplinarity and their impact barely extends beyond the specific disciplines in which they have originated (Szell et al., 2018).

Finally, Figure 11 confirms the high overlap between NoveltyU and Integration Score and shows that few articles differ in interdisciplinarity and novelty from the entire sample. Moreover, Figure 11 shows that several novel articles score low on NoveltyU and Integration Score. So, based on the assumption that Nobel Prize and APS milestone articles should be more novel than the other

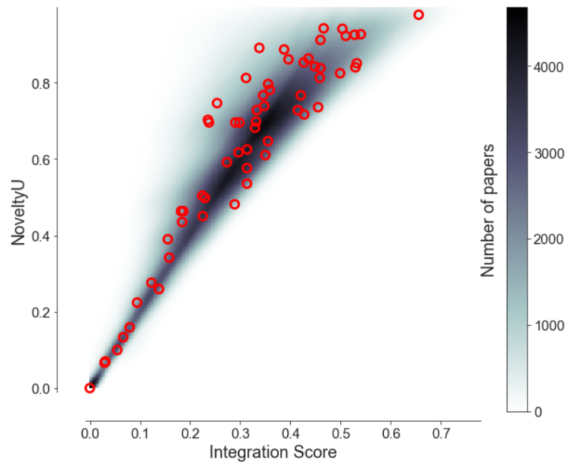


Figure 11: Density plot of the relationship between Integration Score and NoveltyU (grey). Red circles are Nobel Prize articles and APS milestones.

articles in the sample, we conclude that NoveltyU is more likely to measure interdisciplinarity than novelty.

6.3.2 Relationship with impact

One of the consequences of the correlation between interdisciplinarity and novelty measures is that they have the same effect on articles' impact. Given the doubts that we have raised on the capacity of novelty indicators to properly measure novelty, we suggest that results linking different forms of articles' impact to novelty indicators are mainly driven by their interdisciplinarity. In this section, we compare the effects of interdisciplinarity and novelty indicators on impact through a set of regressions. As in Section 6.1.2, we ran Negative Binomial and OLS regressions controlling for PACS, year, journal, number of authors, number of backward citations, international team, and Variety=1.

As a first step, we test the relationship in our data between interdisciplinarity and impact. We use three different indicators for impact: the number of citations, the number of citing fields, and generality. Table 8 (specifications 1, 2, 6, 7, 11, and 12) shows a positive effect of Variety and an inverted U-shaped relationship between the other interdisciplinarity indicators and impact that is consistent with the interdisciplinarity literature (Adams et al., 2007; Yegros-Yegros et al., 2015). Secondly, we consider the relationship between impact and NoveltyU and Conventionality. In specification (3), (8), and (13) we replace NoveltyU with Disparity, while in specifications (4), (5),(9),

Table 8: Effect of interdisciplinarity and novelty on the number of citations (negative binomial), the number of citing fields (negative binomial), and generality (OLS).

	Dependent variable:														
	Number of citations					Number of citing fields					Generality				
	Negative binomial					*Negative binomial*					*OLS*				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
Variety	0.049***		0.049***			0.043***		0.043***			0.010***		0.010***		
	(0.0005)		(0.0005)			(0.0004)		(0.0004)			(0.0001)		(0.0001)		
Balance	6.725***		7.426***			5.334***		5.508***			2.542***		2.613***		
	(0.331)		(0.332)			(0.260)		(0.260)			(0.076)		(0.076)		
$Balance^2$	−5.192***		−5.564***			−3.225***		−3.325***			−1.470***		−1.511***		
	(0.198)		(0.198)			(0.155)		(0.155)			(0.046)		(0.046)		
Disparity	0.358***					1.061***					0.321***				
	(0.075)					(0.056)					(0.017)				
$Disparity^2$	−1.082***					−1.391***					−0.443***				
	(0.083)					(0.062)					(0.019)				
Integration Score		−0.013					1.475***					0.451***			
		(0.078)					(0.059)					(0.018)			
Integration $Score^2$		−0.707***					−1.631***					−0.545***			
		(0.116)					(0.088)					(0.027)			
$Novelty_U$			0.233***	0.029				0.776***	0.838***				0.243***	0.270***	
			(0.055)	(0.055)				(0.041)	(0.042)				(0.013)	(0.013)	
$Novelty_U^2$			−0.473***	−0.210***				−0.709***	−0.581***				−0.231***	−0.208***	
			(0.046)	(0.046)				(0.035)	(0.035)				(0.011)	(0.011)	
Conventionality					0.619***					1.307***					0.440***
					(0.066)					(0.050)					(0.015)
$Conventionality^2$					−0.221***					−1.244***					−0.402***
					(0.054)					(0.041)					(0.013)
log(Number of backward citations)		0.464***		0.465***	0.445***		0.273***		0.274***	0.286***		0.066***		0.066***	0.069***
		(0.003)		(0.003)	(0.003)		(0.002)		(0.002)	(0.002)		(0.001)		(0.001)	(0.001)
log(Number of authors)	0.126***	0.120***	0.129***	0.122***	0.120***	0.104***	0.101***	0.105***	0.099***	0.100***	0.033***	0.032***	0.033***	0.031***	0.031***
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
International	0.041***	0.038***	0.040***	0.037***	0.038***	0.020***	0.019***	0.019***	0.020***	0.019***	0.010***	0.009***	0.009***	0.009***	0.009***
	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Variety=1	1.371***	−0.165***	1.712***	−0.158***	−0.204***	1.998***	0.006	2.082***	0.006	−0.035**	1.000***	−0.035***	1.037***	−0.035***	−0.047***
	(0.138)	(0.021)	(0.138)	(0.023)	(0.020)	(0.109)	(0.016)	(0.109)	(0.018)	(0.015)	(0.032)	(0.005)	(0.032)	(0.005)	(0.004)
Constant	0.126	1.579***	−0.223	1.587***	1.232***	−0.748***	1.211***	−0.835***	1.195***	1.179***	−0.611***	0.416***	−0.648***	0.410***	0.388***
	(0.138)	(0.022)	(0.138)	(0.024)	(0.027)	(0.108)	(0.017)	(0.108)	(0.018)	(0.021)	(0.032)	(0.005)	(0.032)	(0.006)	(0.006)
1-digit PACS	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Journal	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	230,854	230,854	230,854	230,854	230,854	230,854	230,854	230,854	230,854	230,854	230,854	230,854	230,854	230,854	230,854
R^2											0.125	0.115	0.124	0.113	0.116
Adjusted R^2											0.125	0.115	0.124	0.113	0.116
θ	0.875***	0.885***	0.873***	0.884***	0.885***	1.739***	1.701***	1.737***	1.696***	1.702***					
Akaike Inf. Crit.	1,502,636	1,500,346	1,503,141	1,500,608	1,500,238	1,464,430	1,468,169	1,464,631	1,468,668	1,468,025					
Residual Std. Error											0.262	0.263	0.262	0.264	0.263

Note: *p<0.1; **p<0.05; ***p<0.01

(10), (14) and (15) we consider NoveltyU and Conventionality alone. In specification (3), (8), and (13) we show that, for all our impact indicators, the inverted U-shaped relationship is preserved and the impact of NoveltyU is very similar to the one of Disparity. Besides, in specifications (4), (5),(9), (10), (14) and (15), the sign and the magnitude of the effect of NoveltyU and Conventionality replicate those of the Integration Score in the specification (2).[17] We can conclude that NoveltyU and interdisciplinarity are not only highly correlated but also have the same effect on impact, and they measure the same article features, as suggested in 3. These results are corroborated by considering the three main components resulting from the PCA (see the supplementary material). Disparity, Integration Score, NoveltyU, and Conventionality (the main elements of RC1) have an inverted U-shaped relationship with impact (scores are centered and scaled). Variety (the main element of RC2) has a positive effect on impact, while Balance (the main contributor to RC3) has an inverted U-shaped relationship with it.

Our evidence has an additional important implication on the analysis of the relationship between interdisciplinarity and impact. The measures proposed by Uzzi et al. (2013) can be used as alternative indicators of interdisciplinarity and the results shown in Table 8 confirm that the nature of this relationship is robust to the use of different indicators. The confirmed evidence is that articles that have a prevalent (but not unique) field in backward citations have an advantage in citations. Strikingly, even if we are restricting our analysis to a single discipline, the inverted U-shaped effect of Disparity (or NoveltyU) emphasizes that the knowledge integrated must not be too distant in order to acquire citations.

7 Conclusions

A deeper understanding of the forces that drive scientific discoveries is a key factor in effectively addressing important environmental, societal, and technological problems. The increased availability of digital data on scientific outputs allows exploration of the nature of the scientific activity, its novelty and interdisciplinarity, with the essential purpose of developing tools and policies with the potential to improve the organization of scientific activities. The recent empirical literature has

[17] It can be noted the negative effect of the Integration Score on the number of citations in Table 8. The same result can also be observed for NoveltyU in Table 8. This is coherent with the result shown in Table 6, where the Integration Score of the Nobel Prize articles is slightly lower than the sample average, and with the idea that NoveltyU measures interdisciplinarity and not novelty.

dedicated a substantial effort to building indicators that measure novelty and interdisciplinarity, to studying how they affect the importance of scientific research and, finally, to developing supporting policy.[18]

The primary objective of this study is thus to contribute to the debate on the measurement and impact of interdisciplinarity and novelty in science. In particular, we challenge the indicators that underpin the current literature on novelty (Uzzi et al., 2013; Wang et al., 2017) on the ground that novelty is a property of the outcome of the recombination process and not of the bits of knowledge that are recombined. More precisely, the features of the article backward citations are not related to the actual novelty of that paper. In addition, we study whether the indicators are able to distinguish between novelty and interdisciplinarity, which instead is related to the article references. We base our results on three methods: on the construction of a CNM, on the analysis of an external validation set, and, finally, on the analysis of the relationship between novelty, interdisciplinarity, and impact.

Our results show that defining novelty as the first appearance of a combination as in Wang et al. (2017) (NoveltyW) makes the measurement very sensitive to the level of aggregation of the recombined fields and dependant on the key structural properties of the citation network. For these reasons, NoveltyW fails in identifying as new a large share of papers that have led to discoveries awarded with the Nobel Prize or that have been deemed as new by the scientific community. With regard to impact, we find that NoveltyW calculated on randomized data has no significantly different effect on impact from the measure calculated on the observed data, casting further doubts on its ability to systematically detect novel contributions.

Novelty defined as an atypical combination as in Uzzi et al. (2013) (NoveltyU) shows a lower sensitivity to the level of aggregation of the recombined fields but is highly correlated with the indicators of interdisciplinarity and has the same effect as interdisciplinarity on impact. In the

[18] For instance, the European Research Council grants require that the applicants are engaged in frontier research which "stands at the forefront of creating new knowledge and developing new understanding. Those involved are responsible for fundamental discoveries and advances in theoretical and empirical understanding, and even achieving the occasional revolutionary breakthrough that completely changes our knowledge of the world" (see European Commission, *Frontier Research: The European Challenge. High-Level Expert Group Report 2005.* https://erc.europa.eu/sites/default/files/publication/files/high-level_expert_group_report_full_report_2005_en). In 1980, the US National Science Foundation has started to develop a mechanism to fund "highly creative or innovative" research for which there was "a high risk of failure". Since then there have been several support programs: the Expedited Awards for Novel Research (EANR), the Small Grants for Exploratory Research (SGER), and the Early-concept Grant for Exploratory Research (EAGER).

external validation set, this indicator assigns higher values of novelty to discoveries that exhibit higher interdisciplinarity thus confirming the problematic nature of disentangling interdisciplinarity and novelty and overlooking novelty that derives from monodisciplinary research.

We, therefore, conclude that the conceptualization of novelty à la Wang et al. (2017) finds it difficult to tell novel and non-novel articles apart, while the conceptualization of novelty à la Uzzi et al. (2013) hardly distinguishes between novel and interdisciplinary contributions.

The scalability of our results across levels of aggregation of fields and within fields shows that they derive from the design of measures and that they do not depend on the specific data set. Thus they might reasonably apply beyond the boundaries of physics.

Overall, our results suggest that novelty should be more profitably investigated in the outcome of the recombination process – the paper – rather than in its references.

The implications of these findings are twofold. On the one hand, they suggest that further effort should be dedicated to improving the measurement of novelty. Since novelty in research has long been recognized as one of the drivers of economic growth (Grossman and Helpman, 1991; Phelps, 1996; Romer, 1990) its precise measurement is a priority for the implementation and assessment of high impact policy schemes such as the Lisbon strategy (EU Lisbon Strategy, 2000).

On the other hand, the policy claims that are made on the basis of these indicators should be treated with caution. For policy based on novelty as the first occurrence of a combination, the problems with the distinction between novel and non-novel articles call into question the claim that organizations are more likely to fund projects with intermediate levels of novelty (Criscuolo et al., 2017) or that novel articles (Wang et al., 2017) and patents (Verhoeven et al., 2016) are riskier in terms of impact but more likely to make a hit. For policy based on novelty as an atypical combination, ignoring the high level of similarity between measures of interdisciplinarity and novelty might lead to tautological statements. For instance, it is claimed that increasing interdisciplinarity of research teams also increases the novelty of their output (Lee et al., 2015). The overlap between the measurement of interdisciplinarity and novelty that emerges from our analysis would suggest that what is actually found is that increasing the interdisciplinarity of research teams positively impacts on the interdisciplinarity of their work. Similarly, claims that research funded by competitive project funding are on average more novel could be read as a preference for research that is more interdisciplinary (Wang et al., 2018).

8 Acknowledgements

The authors wish to thank an anonymous reviewer for his/her careful reading of the manuscript and his/her many insightful comments and suggestions. The authors are grateful to the participants to the 29th and 30th EAEPE annual conferences, Budapest 2017 and Nice 2018, to the workshop on Data Science and Economics, Turin 2018, to the Innovation, Economic Complexity, and Economic Geography Workshop, MIT, Boston 2018, and to the 23th International Conference of Science and Technology Indicators (STI), Leiden 2018, for useful comments on previous versions of the paper. The authors are also indebted with the Colleagues of the Department of Network & Data Science of the Central European University of Budapest for the enlightening discussions at the initial stage of the research. M.I. has received financial support from a STSM Grant from the COST Action CA15109 (COSTNET). The usual disclaimer applies.

References

Adams, J., L. Jackson, and S. Marshall (2007). Bibliometric analysis of interdisciplinary research. Report to Higher Education Funding Council for England.

Arthur, W. B. (2009). The nature of technology: What it is and how it evolves. Simon and Schuster.

Baibich, M. N. et al. (1988). Giant magnetoresistance of (001)Fe/(001)Cr magnetic superlattices. Physical Review Letters 61 (21), 2472–2475. 10.1103/PhysRevLett.61.2472.

Battiston, F., F. Musciotto, D. Wang, A.-L. Barabási, M. Szell, and R. Sinatra (2019). Taking census of physics. Nature Reviews Physics 1(1), 89. 10.1038/s42254-018-0005-3.

Bennett, C. H., G. Brassard, C. Crépeau, R. Jozsa, A. Peres, and W. K. Wootters (1993). Teleporting an unknown quantum state via dual classical and Einstein-Podolsky-Rosen channels. Physical Review Letters 70 (13), 1895–1899. 10.1103/PhysRevLett.70.1895.

Bennett, C. H., G. Brassard, and N. D. Mermin (1992). Quantum cryptography without Bell's theorem. Physical Review Letters 68 (5), 557–559. 10.1103/PhysRevLett.68.557.

Binasch, G., P. Grünberg, F. Saurenbach, and W. Zinn (1989). Enhanced magnetoresistance in layered magnetic structures with antiferromagnetic interlayer exchange. Physical Review B 39 (7), 4828–4830. 10.1103/PhysRevB.39.4828.

Bollobás, B. and B. Béla (2001). Random graphs. 73. Cambridge University Press.

Boudreau, K. J., E. C. Guinan, K. R. Lakhani, and C. Riedl (2016). Looking across and looking beyond the knowledge frontier: Intellectual distance, novelty, and resource allocation in science. Management Science 62(10), 2765–2783. 10.1287/mnsc.2015.2285.

Carayol, N., L. Agenor, and L. Oscar (2019). The right job and the job right: Novelty, impact and journal stratification in science. Cahiers du GREThA, 2019-05. URL: `ideas.repec.org/p/grt/wpegrt/2019-05.html`.

Cedrini, M. and M. Fontana (2017). Just another niche in the wall? How specialization is changing the face of mainstream economics. Cambridge Journal of Economics 42(2), 427–451. 10.1093/cje/bex003.

Criscuolo, P., L. Dahlander, T. Grohsjean, and A. Salter (2017). Evaluating novelty: The role of panels in the selection of R&D projects. Academy of Management Journal 60(2), 433–460. 10.5465/amj.2014.0861.

D'Este, P., O. Llopis, F. Rentocchini, and A. Yegros (2019). The relationship between interdisciplinarity and distinct modes of university-industry interaction. Research Policy. 10.1016/j.respol.2019.05.008.

Diddams, S. A. et al. (2000). Direct link between microwave and optical frequencies with a 300 THz femtosecond laser comb. Physical Review Letters 84 (22), 5102–5105. 10.1103/PhysRevLett.84.5102.

Ekert, A. K. (1991). Quantum cryptography based on Bell's theorem. Physical Review Letters 67 (6), 661–663. 10.1103/PhysRevLett.67.661.

EU Lisbon Strategy (2000). Lisbon European Council 23 and 24 March 2000 - Presidency conclusions. URL: http://www.europarl.europa.eu/summits/lis1_en.htm.

European Union Research Advisory Board (2004). Interdisciplinarity in Research. 04.009-FINAL.

Fortunato, S. et al. (2018). Science of science. Science 359(6379). 10.1126/science.aao0185.

Gonzalez, M. C., C. A. Hidalgo, and A.-L. Barabasi (2008). Understanding individual human mobility patterns. Nature 453(7196), 779. 10.1038/nature06958.

Grossman, G. and E. Helpman (1991). Innovation and growth in the global economy. MIT Press.

Hall, B. H., A. B. Jaffe, and M. Trajtenberg (2001). The NBER patent citation data file: Lessons, insights and methodological tools. NBER Working Paper 8498. 10.3386/w8498.

Hopfield, J. (2017). Reflections on the APS and the evolution of Physics. APS News 16(8).

Huutoniemi, K., J. T. Klein, H. Bruun, and J. Hukkinen (2010). Analyzing interdisciplinarity: Typology and indicators. Research Policy 39(1), 79–88. 10.1016/j.respol.2009.09.011.

Iori, M. and M. Fontana (2019). Novelty as recombination of knowledge. ISSI 2019 conference proceedings.

Jones, B. F. (2009). The burden of knowledge and "the death of the renaissance man": Is innovation getting harder? The Review of Economic Studies 76(1), 283–317. 10.1111/j.1467-937X.2008.00531.x.

Kaplan, S. and K. Vakili (2015). The double-edged sword of recombination in breakthrough innovation. Strategic Management Journal 36(10), 1435–1457. 10.1002/smj.2294.

Kelly, B., D. Papanikolaou, A. Seru, and M. Taddy (2018). Measuring technological innovation over the long run. NBER working paper 25266. 10.3386/w25266.

Kelly, J. (1996). Wide and narrow interdisciplinarity. The Journal of General Education 45(2), 95–113.

Kim, D., D. B. Cerigo, H. Jeong, and H. Youn (2016). Technological novelty profile and invention's future impact. EPJ Data Science 5(1), 8. 10.1140/epjds/s13688-016-0069-1.

Klein, J. (2005). Humanities, culture, and interdisciplinarity: The changing American academy. SUNY Press.

Larivière, V., S. Haustein, and K. Börner (2015). Long-distance interdisciplinarity leads to higher scientific impact. PLOS ONE 10(3), e0122565. 10.1371/journal.pone.0122565.

Lee, Y.-N., J. P. Walsh, and J. Wang (2015). Creativity in scientific teams: Unpacking novelty and impact. Research Policy 44(3), 684–697. 10.1016/j.respol.2014.10.007.

Levitt, J. M. and M. Thelwall (2008). Is multidisciplinary research more highly cited? A macrolevel study. Journal of the Association for Information Science and Technology 59(12), 1973–1984. 10.1002/asi.20914.

Mansfield, E. (1991). Academic research and industrial innovation. Research Policy 20(1), 1–12. 10.1016/0048-7333(91)90080-A.

March, J. G. (1991). Exploration and exploitation in organizational learning. Organization Science 2(1), 71–87. 10.1287/orsc.2.1.71.

Maslov, S. and K. Sneppen (2002). Specificity and stability in topology of protein networks. Science 296(5569), 910–913. 10.1126/science.1065103.

Molas-Gallart, J. and A. Salter (2002). Diversity and excellence: considerations on research policy. IPTS Report 66.

Nelson, R. R. and S. G. Winter (1982). An evolutionary theory of economic change. Harvard University Press.

Newman, M. E. J., S. H. Strogatz, and D. J. Watts (2001). Random graphs with arbitrary degree distributions and their applications. Physical Review E 64 (2). 10.1103/PhysRevE.64.026118.

Pan, R. K., S. Sinha, K. Kaski, and J. Saramäki (2012). The evolution of interdisciplinarity in physics research. Scientific Reports 2, 551. 10.1038/srep00551.

Phelps, E. (1996). Models of technical progress and the golden rule of research. Review of Economic Studies 33(2), 133–145. 10.2307/2974437.

Porter, A. and I. Rafols (2009). Is science becoming more interdisciplinary? Measuring and mapping six research fields over time. Scientometrics 81(3), 719–745. 10.1007/s11192-008-2197-2.

Rafols, I., L. Leydesdorff, A. O'Hare, P. Nightingale, and A. Stirling (2012). How journal rankings can suppress interdisciplinary research: A comparison between innovation studies and business & management. Research Policy 41(7), 1262–1282. 10.1016/j.respol.2012.03.015.

Reichert, J., M. Niering, R. Holzwarth, M. Weitz, T. Udem, and T. W. Hänsch (2000). Phase coherent vacuum-ultraviolet to radio frequency comparison with a mode-locked laser. Physical Review Letters 84 (15), 3232–3235. 10.1103/PhysRevLett.84.3232.

Rinia, E. J., T. N. Van Leeuwen, H. G. Van Vuren, and A. F. J. Van Raan (2001). Influence of interdisciplinarity on peer-review and bibliometric evaluations in physics research. Research Policy 30(3), 357–361. 10.1016/S0048-7333(00)00082-2.

Romer, P. (1990). Endogenous technological change. Journal of Political Economy 98(5), S71–S102. 10.1086/261725.

Schumpeter, J. A. (1934). The Theory of Economic Development.

— (1939). Business cycles: A theoretical, historical, and statistical analysis of the capitalist process. McGraw-Hill.

Sinatra, R., P. Deville, M. Szell, D. Wang, and A.-L. Barabási (2015). A century of physics. Nature Physics 11(10), 791. 10.1038/nphys3494.

Sinatra, R., D. Wang, P. Deville, C. Song, and A.-L. Barabási (2016). Quantifying the evolution of individual scientific impact. Science 354(6312), aaf5239. 10.1126/science.aaf5239.

Steele, T. W. and J. C. Stier (2000). The impact of interdisciplinary research in the environmental sciences: A forestry case study. Journal of the Association for Information Science and Technology 51(5), 476–484. 10.1002/(SICI)1097-4571(2000)51:5¡476::AID-ASI8¿3.0.CO;2-G.

Stephan, P., R. Veugelers, and J. Wang (2017). Reviewers are blinkered by bibliometrics. Nature 544, 411–412. 10.1038/544411a.

Stirling, A. (2007). A general framework for analysing diversity in science, technology and society. Journal of the Royal Society Interface 4(15), 707–719. 10.1098/rsif.2007.0213.

Strumsky, D. and J. Lobo (2015). Identifying the sources of technological novelty in the process of invention. Research Policy 44(8), 1445–1461. 10.1016/j.respol.2015.05.008.

Szell, M., Y. Ma, and R. Sinatra (2018). A Nobel opportunity for interdisciplinarity. Nature Physics 14, 1075–1078. 10.1038/s41567-018-0314-6.

The National Academies (2005). Facilitating interdisciplinary research. National Academies Press.

Tian, Y. and S. Yan (2013). Giant magnetoresistance: history, development and beyond. Science China Physics, Mechanics and Astronomy 56(1), 2–14. 10.1007/s11433-012-4971-7.

Uzzi, B., S. Mukherjee, M. Stringer, and B. Jones (2013). Atypical combinations and scientific impact. Science 342(6157), 468–472. 10.1126/science.1240474.

Verhoeven, D., J. Bakker, and R. Veugelers (2016). Measuring technological novelty with patent-based indicators. Research Policy 45(3), 707–723. 10.1016/j.respol.2015.11.010.

Wagner, C. S. et al. (2011). Approaches to understanding and measuring interdisciplinary scientific research (IDR): A review of the literature. Journal of informetrics 5(1), 14–26. 10.1016/j.joi.2010.06.004.

Wang, J., Y.-N. Lee, and J. P. Walsh (2018). Funding model and creativity in science: Competitive versus block funding and status contingency effects. Research Policy 47(6), 1070–1083. 10.1016/j.respol.2018.03.

Wang, J., B. Thijs, and W. Glänzel (2015). Interdisciplinarity and impact: Distinct effects of variety, balance, and disparity. PLOS ONE 10(5), e0127298. 10.1371/journal.pone.0127298.

Wang, J., R. Veugelers, and P. Stephan (2017). Bias against novelty in science: A cautionary tale for users of bibliometric indicators. Research Policy 46(8), 1416–1436. 10.1016/j.respol.2017.06.006.

Wu, M. K. et al. (1987). Superconductivity at 93 K in a new mixed-phase Y-Ba-Cu-O compound system at ambient pressure. Physical Review Letters 58 (9), 908–910. 10.1103/PhysRevLett.58.908.

Wuchty, S., B. F. Jones, and B. Uzzi (2007). The increasing dominance of teams in production of knowledge. Science 316(5827), 1036–1039. 10.1126/science.1136099.

Yegros-Yegros, A., I. Rafols, and P. D'Este (2015). Does interdisciplinary research lead to higher citation impact? The different effect of proximal and distal interdisciplinarity. PLOS ONE 10(8). 10.1371/journal.pone.0135095.

Zeng, A. et al. (2017). The science of science: From the perspective of complex systems. Physics Reports. 10.1016/j.physrep.2017.10.001.

A Further information on data

A.1 PACS codes in APS database

A PACS code is composed of three elements: a pair of 2-digit numbers separated by . and followed by two characters that may be letters or + or − signs.

The first digit of the first 2-digit number identifies the main field out of the 10 broad fields specified at the first level (first hierarchy level) and the second digit specifies a sub-field within that field (second hierarchy level). The second 2-digit number further specifies a narrower category within the field given by the first two digits (third hierarchy level). The last two characters specify even more detailed categories up to the fifth level of the hierarchy. We use the PACS codes up to the third level of hierarchy (4 digits) since they represent the sub-fields of physics and are more stable over time compared to deeper levels of hierarchy (see Pan et al. (2012) for further details).

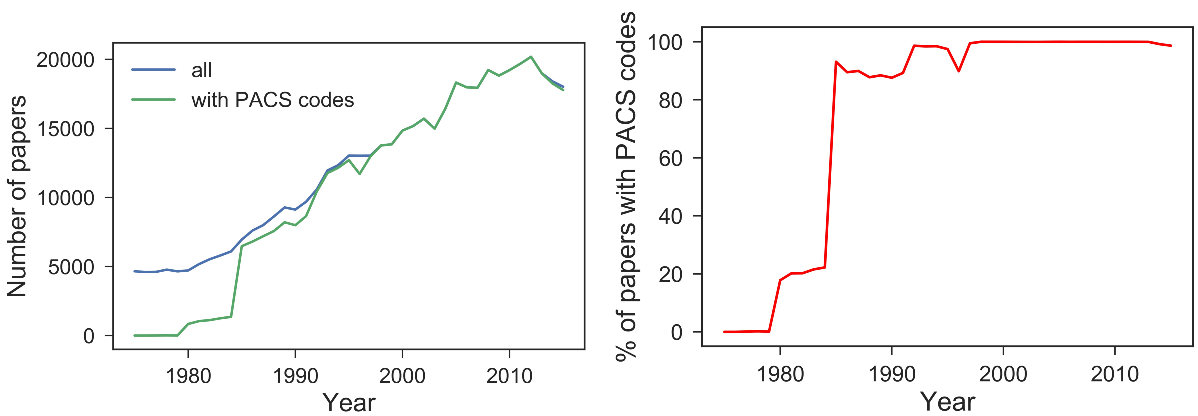


Figure A.1: Articles with PACS codes compared to the number of articles in the APS database in time (left) and share of articles with PACS codes in time (right).

A.2 Correlation between variables

Table A.1 shows the correlation between the variables used in the regression analysis.

Table A.1: Correlation table between variables used in regressions (4-digit level).

	NoveltyW	NoveltyU	Conventionality	Variety	Balance	Disparity	Integration Score	Number of authors	Number of backward citations	International	Number of forward citations	Number of citing fields	Generality
NoveltyW	1.00	0.13	-0.10	0.29	0.02	0.13	0.15	-0.02	0.31	0.00	0.17	0.16	0.04
NoveltyU	0.13	1.00	-0.67	0.40	0.36	0.91	0.89	-0.06	0.17	0.02	0.03	0.12	0.11
Conventionality	-0.10	-0.67	1.00	-0.19	-0.41	-0.75	-0.88	0.07	0.07	0.02	0.05	-0.05	-0.04
Variety	0.29	0.40	-0.19	1.00	0.08	0.35	0.44	-0.02	0.76	0.08	0.16	0.31	0.26
Balance	0.02	0.36	-0.41	0.08	1.00	0.32	0.40	-0.01	-0.16	-0.01	-0.02	0.05	0.10
Disparity	0.13	0.91	-0.75	0.35	0.32	1.00	0.88	-0.07	0.20	0.01	0.02	0.09	0.08
Integration Score	0.15	0.89	-0.88	0.44	0.40	0.88	1.00	-0.07	0.13	0.01	0.00	0.13	0.13
Number of backward citations	0.31	0.17	0.07	0.76	-0.16	0.20	0.13	0.00	1.00	0.09	0.22	0.25	0.18
Number of authors	-0.02	-0.06	0.07	-0.02	-0.01	-0.07	-0.07	1.00	0.00	0.20	0.05	0.04	0.05
International	0.00	0.02	0.02	0.08	-0.01	0.01	0.01	0.20	0.09	1.00	0.04	0.05	0.06
Number of forward citations	0.17	0.03	0.05	0.16	-0.02	0.02	0.00	0.05	0.22	0.04	1.00	0.77	0.26
Number of citing fields	0.16	0.12	-0.05	0.31	0.05	0.09	0.13	0.04	0.25	0.05	0.77	1.00	0.57
Generality	0.04	0.11	-0.04	0.26	0.10	0.08	0.13	0.05	0.18	0.06	0.26	0.57	1.00

Table B.1: Correlations between the values of novelty and interdisciplinarity indicators computed applying CNM and cosine similarity.

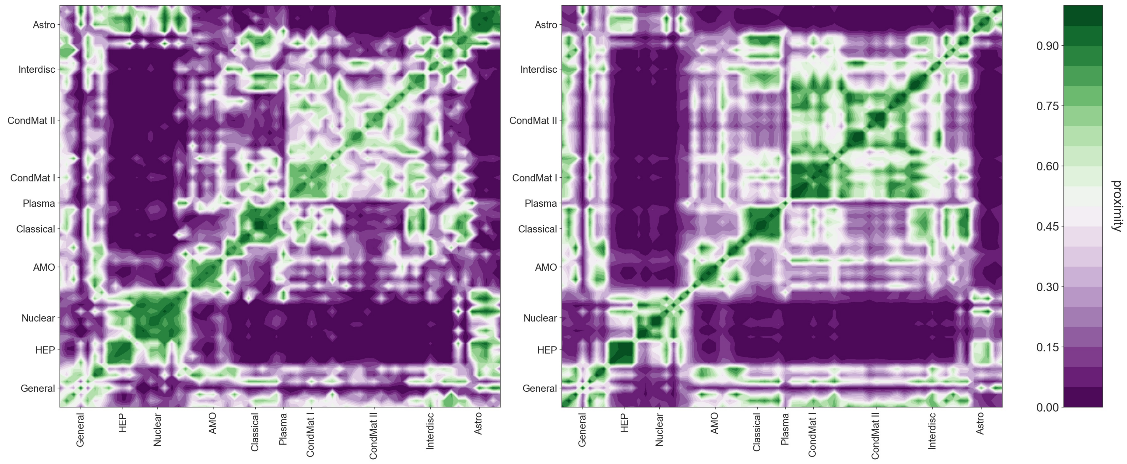
	NoveltyU	Conventionality	Disparity	Integration Score
2-digit	0.87	0.80	0.83	0.85
4-digit	0.83	0.77	0.82	0.86

B Proximity normalization

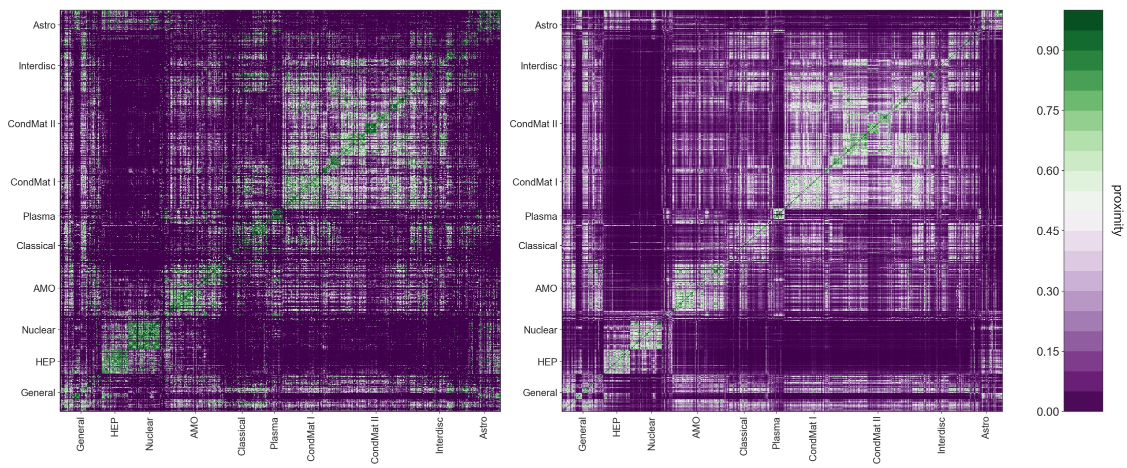
In this appendix, we compare the values of proximity obtained normalizing the number of co-citations with CNM and cosine similarity. The correlation between the two sets of values is high both at the 2-digit and 4-digit levels (respectively 0.81 and 0.72). Figure B.1 shows the values of proximity obtained through the application of CNM (left) and cosine similarity (right) at the 2-digit and 4-digit levels. Differences between the two approaches are evident for Nuclear Physics, Condensed Matter, and AstroPhysics. As explained in Section 5, we do not compute NoveltyW using the proximity obtained applying CNM since it will be reduced to the number of new pairs in the forward citations. Table B.1 shows that the correlation between the measures computed with CNM and cosine similarity is high for all indicators involving proximity in their definition. Therefore, we can conclude that the choice of normalising proximity with cosine similarity also for the indicators devised by Uzzi et al. (2013) does not affect our results. The high correlation between interdisciplinarity and novelty indicators computed with CNM supports our rejection of H4: the correlation of NoveltyU with Disparity and Integration Score (4-digits - CNM) is respectively 92% and 92%, and the correlation between Conventionality and Integration Score (4-digits - CNM) is -85%.

C Results at the 2-digit level

We replicate the analysis considering PACS at the 2-digit level to compute novelty, interdisciplinarity, and impact indicators. At this level of analysis, NoveltyW is equal to zero for all articles published in 2005, since all possible pairs between the 64 PACS at the 2-digit levels appear at least once in the backward citations of papers published before that year. This result highlights the sensitivity of NoveltyW to the unit of analysis and the level of classification codes. Therefore, we focus only on NoveltyU, Conventionality, and interdisciplinarity indicators. All the results concerning



(a) 2-digit level.



(b) 4-digit level.

Figure B.1: Proximity of sub-fields in physics computed applying CNM (left) and cosine similarity (right).

these indicators are confirmed at the 2-digit level.

Testing H1a and H1b The values of NoveltyU, Conventionality, and interdisciplinarity indicators are significantly different in the randomized and observed samples, as confirmed by the plots of their distributions in Figure S1 in supplementary materials. Moreover, Table C.1 shows that the distance between these distributions is high (K-S statistic and Hellinger distance) and they provide different information (KL divergence). Overall, H1a is confirmed for NoveltyU and Conventionality, and H1b is confirmed for interdisciplinarity indicators.

Table C.1: 2-digit level. Distances between randomized and observed distributions of novelty and interdisciplinarity indicators. K-S statistics are significantly different from zero for all indicators (p-values < 0.001).

	NoveltyU	Conventionality	Integration Score	Variety	Balance	Disparity
K-S statistic	0.67	0.70	0.72	0.28	0.62	0.60
Hellinger distance	172.40	124.60	116.50	42.17	89.84	100.53
KL divergence	1.20	1.43	1.53	0.30	2.30	1.00

Testing H2a and H2b We confirm H2a for NoveltyU and Conventionality and H2b for interdisciplinarity measures since the effect of these indicators on impact (Number of Citations, Number of Citing Fields, and Generality) is significantly different in randomized and observed data. Figures S2 and S3 in supplementary materials show the differences in the relationship between the indicators and impact in the two samples. Tables S2 and S3 in supplementary materials, instead, summarise the regression analysis that confirms that these relationships are significantly different in randomized and observed data (the coefficients of NoveltyU*Observed, Conventionality*Observed, Variety*Observed, Balance, Disparity*Observed, and Integration Score*Observed are significantly different from zero).

Testing H3 To verify H3, we test whether the (conditional) mean of NoveltyU is higher in novel articles (Nobel Prize and APS milestones). Table C.2 (full specifications in Table S4 in supplementary materials) shows that this is not the case, so we reject H3. Also, it shows that the conditional mean of the Integration Score is significantly different from zero only for Nobel Prize articles.

Table C.2: 2-digit level. Nobel Prize and milestones effect on NoveltyW, NoveltyU, and Integration Score. (1) unconditional mean, (2) conditional mean. P-values in parenthesis.

	Nobel Prize (1)	Nobel Prize (2)	Milestones (1)	Milestones (2)
NoveltyU	0.044 (0.463)	-0.056 (0.280)	-0.009 (0.807)	-0.027 (0.365)
Integration Score	0.005 (0.857)	-0.050 (0.040)	0.016 (0.319)	-0.002 (0.880)

Table C.3: 2-digit level. Results of the principal component analysis on Novelty$_U$, Conventionality, and interdisciplinarity indicators.

(a) Principal components

	Components		
	RC1	RC2	RC3
Variety	0.24	0.10	0.95
Balance	0.17	0.95	0.12
Disparity	0.91	0.03	0.20
Integration Score	0.89	0.30	0.26
Conventionality	-0.75	-0.45	0.04
Novelty$_U$	0.90	0.12	0.29

(b) Importance of the three principal components

	Components		
	RC1	RC2	RC3
Proportion variance	0.51	0.20	0.19
Cumulative variance	0.51	0.72	0.90
Proportion explained	0.57	0.22	0.21
Cumulative proportion	0.57	0.79	1.00

Novelty or interdisciplinarity? The novelty indicators defined in Uzzi et al. (2013) are highly correlated with interdisciplinarity measures. The correlation of NoveltyU with Disparity and Integration Score is, respectively, 89% and 89%, whereas the correlation between Conventionality and Integration Score is -83%. Table C.3 reports the result of a PCA analysis with orthogonal rotation (varimax) performed on the six variables. The first three components explain more than 90% of the total variance. As expected, RC1, which explains the 51% of the total variance, combines Disparity, Integration Score, NoveltyU, and Conventionality. Variety and Balance, instead, are predominant, respectively, in RC3 and RC2. More details on RC1, RC2, and RC3 are in Figure S4 in the supplementary material. The PCA analysis confirms that the measures of novelty (NoveltyU and Conventionality) strongly overlap with the interdisciplinarity indicators based on proximity (Disparity and Integration Score).

The overlap between NoveltyU and Integration Score is confirmed in Figure C.1, which shows the density plot between the two indicators and highlights novel articles (Nobel Prize articles and APS milestones as red circles). As displayed by the figure, NoveltyU and Integration Score highly correlate also in this case and a large share of novel articles has a low value of NoveltyU.

These indicators are not only highly correlated but also have the same effect on the different measures of impact, as shown by Table S5 in supplementary materials.

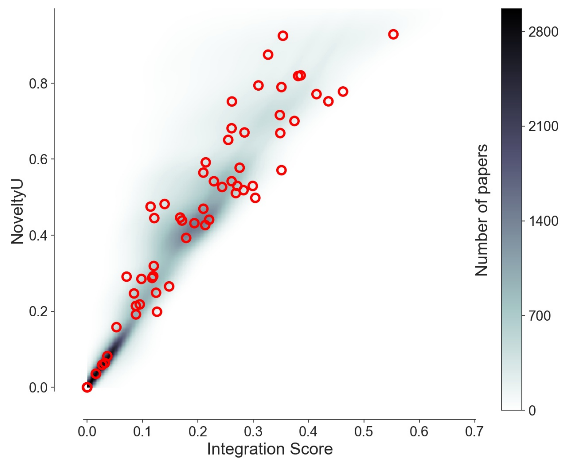


Figure C.1: 2-digit level. Density plot of the relationship between Integration Score and NoveltyU (grey). Red circles are Nobel Prize articles and APS milestones.

D Results of the analysis within a 1-digit PACS

We replicate the entire analysis within the largest 1-digit PACS, CondMat II. We select only focal, cited, and citing articles belonging to this sub-field. The number of focal articles in this sample is 91,058. Since the results of the paper are confirmed, in this section we can claim that our conclusions are robust to different specifications of the database and are independent of the scientific domain of analysis.

Testing H1a and H1b Figures S5 and S6 in supplementary materials show the distributions of novelty and interdisciplinarity indicators in randomized and observed data. The distributions of NoveltyW in the two samples highly overlap, whereas they differ for the other indicators. The high similarity between the distributions of NoveltyW in randomized and observed data is confirmed by the values of the K-S statistic, the Hellinger distance, and the KL divergence reported in Table D.1. It is worth noting that, for NoveltyW only, the K-S statistic is not significantly different from zero (p-value $>$ 0.05). Therefore, H1a is confirmed for NoveltyU and Conventionality, but it is rejected for NoveltyW. Finally, H1b is confirmed for all interdisciplinarity indicators.

Testing H2a and H2b Figures S7 and S8 in supplementary materials show the relationship between novelty or interdisciplinarity indicators and impact for papers in Condensed Matter II. While the relationships involving NoveltyW overlap in the observed and randomized data, the ones

Table D.1: Condensed Matter II. Distances between randomized and observed distributions of novelty and interdisciplinarity indicators.

	NoveltyW	NoveltyU	Conventionality	Integration Score	Variety	Balance	Disparity
K-S statistic	0.02	0.75	0.73	0.79	0.43	0.53	0.69
Hellinger distance	4.89	62.26	33.91	52.21	34.92	49.41	45.28
KL divergence	0.06	1.63	1.75	1.89	0.45	2.37	1.42

concerning the other indicators differ in the two samples. Table D.2 confirms that the relationships between NoveltyW and impact indicators are the same in randomized and observed data since NoveltyW*Observed is not significantly different from zero. Tables S6 and S7 in supplementary materials report the same analysis for NoveltyU, Conventionality, and interdisciplinarity indicators and show that, for these indicators, the relationships with impact are significantly different on randomized and observed data.

We can conclude that, as for the entire database, H2a is rejected for NoveltyW and confirmed

Table D.2: Condensed Matter II. Differences in the relationship between NoveltyW and impact in observed and randomized data.

	Dependent variable:		
	Number of citations	Number of citing disciplines	Generality
	Negative binomial	*Negative binomial*	*OLS*
	(1)	(2)	(3)
Novelty_W	0.476 (0.570)	0.371 (0.378)	0.115 (0.133)
Novelty_W*Observed	0.881 (0.657)	0.221 (0.435)	0.074 (0.154)
log(Number of backward citations)	-0.720^{***} (0.065)	-0.766^{***} (0.053)	-0.255^{***} (0.014)
log(Number of authors)	0.098^{***} (0.021)	0.015 (0.014)	0.015^{***} (0.005)
International	0.166^{***} (0.018)	0.190^{***} (0.012)	0.058^{***} (0.004)
Variety = 1	0.422^{***} (0.015)	0.276^{***} (0.011)	0.066^{***} (0.003)
Constant	1.231^{***} (0.086)	0.682^{***} (0.063)	0.307^{***} (0.019)
1-digit PACS	Yes	Yes	Yes
Journal	Yes	Yes	Yes
Year	Yes	Yes	Yes
Observations	13,970	13,970	13,970
R^2			0.220
Adjusted R^2			0.218
θ	0.767^{***}	2.023^{***}	
Akaike Inf. Crit.	112,471	82,237	
Residual Std. Error			0.270
Note:			*p<0.1; **p<0.05; ***p<0.01

Table D.3: Condensed Matter II. Nobel Prize and milestones effect on NoveltyW, NoveltyU, and Integration Score. (1) unconditional mean, (2) conditional mean. P-values in parenthesis.

	Nobel Prize (1)	Nobel Prize (2)	Milestones (1)	Milestones (2)
NoveltyW	0.355	0.128	0.010	0.017
	(0.115)	(0.558)	(0.940)	(0.893)
NoveltyU	0.004	0.024	-0.046	0.009
	(0.973)	(0.813)	(0.507)	(0.879)
Integration Score	-0.021	-0.025	-0.015	0.009
	(0.617)	(0.656)	(0.640)	(0.748)

for NoveltyU and Conventionality. Moreover, H2b is confirmed for all indicators.

Testing H3 We test the degree of NoveltyW and NoveltyW of papers related to Nobel Prize winning researches or in APS Milestones lists. In this sample, we have 2 Nobel Prize related articles and 9 milestones (of which 2 are Nobel Prize related articles). As shown in Table D.3 (full specifications in Table S8 in supplementary materials), the (conditional) mean of NoveltyW and NoveltyU of these articles is not significantly different from the (conditional) mean of the other articles in Condensed Matter II. Therefore, we can reject H3.

Novelty or interdisciplinarity? Interdisciplinarity and novelty indicators defined by Uzzi et al. (2013) are highly correlated. The correlation of NoveltyU with Disparity and Integration Score is, respectively, 90% and 92%, while the correlation of Conventionality with Integration Score is -82%. These values are similar to the ones obtained on the entire database. Table D.4 reports

Table D.4: Condensed Matter II. Results of the principal component analysis on Novelty_U, Conventionality, and interdisciplinarity indicators.

(a) Principal components

	Components		
	RC1	RC2	RC3
Variety	0.20	0.96	0.10
Balance	0.29	0.11	0.95
Disparity	0.89	0.23	0.21
Integration Score	0.89	0.34	0.23
Conventionality	-0.87	0.08	-0.26
Novelty_U	0.85	0.37	0.18

(b) Importance of the three principal components

	Components		
	RC1	RC2	RC3
Proportion of variance	0.53	0.21	0.18
Cumulative variance	0.53	0.74	0.92
Proportion explained	0.58	0.22	0.20
Cumulative proportion	0.58	0.80	1.00

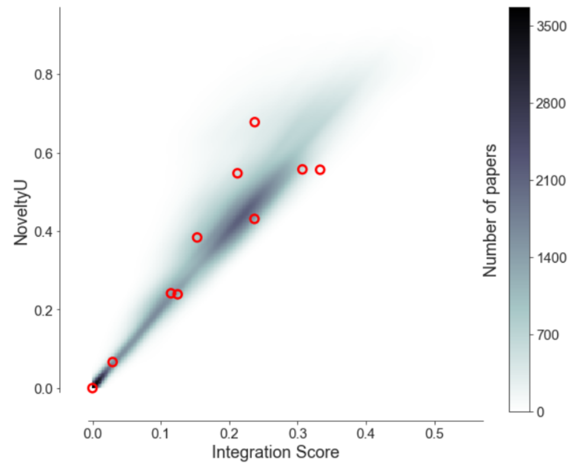


Figure D.1: Condensed Matter II. Density plot of the relationship between Integration Score and NoveltyU (grey). Red circles are Nobel Prize articles and APS milestones.

the result of a PCA analysis performed on the six variables. The first three components explain more than 92% of the total variance. As expected, RC1, which explains the 53% of the total variance, combines Disparity, Integration Score, NoveltyU, and Conventionality. RC2 and RC3 capture, respectively, Variety and Balance. More details on RC1, RC2, and RC3 are in Figure S9 in the supplementary material. The PCA analysis confirms that the measures of novelty (NoveltyU and Conventionality) strongly overlap with the interdisciplinarity indicators based on proximity (Disparity and Integration Score).

The overlap between NoveltyU and interdisciplinarity is also confirmed in Figure D.1, which reports the density plot between the two indicators and highlights articles related to Nobel Prize winning research and milestones (red circles). Also in the case of Nobel Prize and APS milestone articles, NoveltyU and Integration Score are highly correlated. Moreover, a large share of these articles has a low value of NoveltyU, confirming that this indicator is capturing interdisciplinarity rather than novelty.

Furthermore, the effect of NoveltyU or Conventionality and impact is similar to the effect of interdisciplinarity on the same variables, as confirmed by regressions in Table S9 in supplementary materials.

E Novelty and interdisciplinarity of Nobel Prize research and APS milestone articles

Table E.1 shows the full specifications of the regression analysis that tests whether novelty and interdisciplinarity of Nobel Prize research and APS milestones articles is different from the ones of "ordinary" articles.

Table E.1: Effect of Nobel Prize and APS milestones on novelty and interdisciplinarity.

	Dependent variable:											
	Novelty$_W$				Novelty$_U$				Integration Score			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Nobel Prize	0.071	−0.421			0.039	−0.011			−0.017	-0.049^{*}		
	(0.614)	(0.588)			(0.053)	(0.046)			(0.031)	(0.027)		
Milestones			0.418	0.280			−0.008	−0.019			0.004	−0.010
			(0.358)	(0.343)			(0.031)	(0.027)			(0.018)	(0.016)
log(Number of backward citations)		0.025^{**}		0.025^{**}		0.001		0.001		0.001^{**}		0.001^{**}
		(0.012)		(0.012)		(0.001)		(0.001)		(0.001)		(0.001)
log(Number of authors)		0.216^{***}		0.216^{***}		-0.538^{***}		-0.538^{***}		-0.263^{***}		-0.263^{***}
		(0.040)		(0.040)		(0.003)		(0.003)		(0.002)		(0.002)
International		-0.055^{***}		-0.055^{***}		-0.018^{***}		-0.018^{***}		-0.011^{***}		-0.011^{***}
		(0.008)		(0.008)		(0.001)		(0.001)		(0.0004)		(0.0004)
Variety = 1		0.642^{***}		0.642^{***}		0.050^{***}		0.050^{***}		0.021^{***}		0.021^{***}
		(0.007)		(0.007)		(0.001)		(0.001)		(0.0003)		(0.0003)
Constant	0.570^{***}	-0.328^{***}	0.570^{***}	-0.328^{***}	0.627^{***}	0.417^{***}	0.627^{***}	0.417^{***}	0.309^{***}	0.173^{***}	0.309^{***}	0.173^{***}
	(0.006)	(0.046)	(0.006)	(0.046)	(0.0005)	(0.004)	(0.0005)	(0.004)	(0.0003)	(0.002)	(0.0003)	(0.002)
1-digit PACS	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Journal	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	230,854	230,854	230,854	230,854	230,854	230,854	230,854	230,854	230,854	230,854	230,854	230,854
R^2	0.00000	0.083	0.00001	0.083	0.00000	0.244	0.00000	0.244	0.00000	0.221	0.00000	0.221
Adjusted R^2	−0.00000	0.083	0.00000	0.083	−0.00000	0.243	−0.00000	0.243	−0.00000	0.221	−0.00000	0.221
Residual Std. Error	2.676	2.563	2.676	2.563	0.230	0.200	0.230	0.200	0.135	0.120	0.135	0.120

Note: *p<0.1; **p<0.05; ***p<0.01