UNIVERSITÀ CATTOLICA del Sacro Cuore

DIPARTIMENTO DI POLITICA ECONOMICA

A bridge over troubled water: Interdisciplinarity, Novelty, and Impact

Magda Fontana
Martina Iori
Fabio Montobbio
Roberta Sinatra

Quaderno n. 2/settembre 2018

VP VITA E PENSIERO

Università Cattolica del Sacro Cuore

DIPARTIMENTO DI POLITICA ECONOMICA

A bridge over troubled water: Interdisciplinarity, Novelty, and Impact

Magda Fontana
Martina Iori
Fabio Montobbio
Roberta Sinatra

Quaderno n. 2/settembre 2018

VP VITA E PENSIERO

Magda Fontana, Despina Big Data Lab, and Dept. of Economics and Statistics, University of Turin.

Martina Iori, Dept. of Economics and Statistics, University of Turin.

Fabio Montobbio, Dept. of Economic Policy, Università Cattolica del Sacro Cuore, Milan. BRICK, Collegio Carlo Alberto, Turin. ICRIOS, Bocconi University, Milan.

Roberta Sinatra. Dept. of Network & Data Science, and Dept. of Mathematics, Central European University, Budapest, Hungary. Center for Complex Network Research, Northeastern University, Boston, USA. Complexity Science Hub, Vienna.

✉ magda.fontana@unito.it

✉ martina.iori@unito.it

✉ fabio.montobbio@unicatt.it

✉ sinatrar@ceu.edu

Dipartimento di Politica Economica
Università Cattolica del Sacro Cuore
Largo Gemelli 1 – 20123 Milano
Tel. 02-7234.2921

✉ dip.politicaeconomica@unicatt.it

www.vitaepensiero.it

ISBN 978-88-343-3793-6

Abstract

This paper analyses the relationship between novelty, interdisciplinarity and impact of scientific research. We discuss the limitations of the established indicators of novelty and interdisciplinarity (Variety, Disparity and Balance) and how they overlap. We correct the biases with normalized indicators that exploit a Configuration Null Model and we propose a protocol to verify the reliability of the indicators and standardize the measurement. The proposed methodology and new indicators are illustrated exploiting 231450 articles in Physics, their references and citations, from 8 Journals of the American Physical Society (APS) from 1985 to 2015. We show that Novelty *à la* Uzzi et al. (2013) is highly correlated with Disparity and Novelty *à la* Wang et al. (2017) is deeply affected by the properties of the articles' citation network. We show the importance of taking into account the specific definition of the knowledge space and considering separately the different dimensions of interdisciplinarity. Using the new unbiased indicators we show that Variety has a positive effect on the articles' Impact while Disparity and Balance have a negative effect. Previous results tend to underestimate the negative Impact of Disparity and overestimate the positive effects of Novelty.

Keywords: measurement of interdisciplinarity; detection of novelty; impact; normalization of measures; Physics
JEL codes: I23; O31; O33; O38

1 Introduction

The last decades have seen an increasing importance of interdisciplinary and novel research in science policy objectives. The emphasis relies on the statement that the growing complexity of societies requires the integration of knowledge from different fields in order to obtain efficient and novel solutions to social and scientific challenges (Molas-Gallart and Salter, 2002; Rafols et al., 2012). The processes of knowledge integration are believed to be key drivers of progress and growth (Stirling, 1998; Stirling, 2007). The empirical evidence on this relationship, however, is mixed (Zeng et al., 2017). Variability in results has different sources. Interdisciplinarity and novelty are susceptible of a number of implementations that make results difficult to compare and to rely on. The number of proposed measures (e.g. Variety, Balance, Disparity, Integration Score, atypical combination, unprecedented combination), the different levels of analysis (e.g. paper, journal, institution, and fields), the types of bibliometric information (e.g., disciplinary memberships of authors, published journals, references) (Wagner et al., 2011) and its elaborations (co-citations and citations, cosine similarity) are likely to bring to different conclusions on the degree of interdisciplinarity, on the presence of novelty and on their impact.

In this paper, we focus on interdisciplinarity and novelty measures and on how their different implementations can affect the analyses of their impact and, subsequently, the evaluation of research. While the literature often misses motivating the rationale behind the choice of measures and seems to ignore differences in interpretations, we show that measures are not neutral with respect to the assessment of knowledge integration and production. Secondly, we show the measures are distorted by the presence of biases whose importance is nearly ignored by the literature (an exception is Uzzi et al., 2013). The calculation of the proximity of disciplines used to evaluate the degree of knowledge integration in interdisciplinarity and to appreciate the distance of recombined knowledge in the calculation of novelty is biased due to the difference - in size and time - of the composition of the knowledge space. We also argue that similar biases impinge on the reliability of interdisciplinarity and novelty measures. We propose a general procedure to obtain unbiased measures by introducing null configuration models, and we show its superiority to other normalization strategies such as cosine similarity. We then discuss the interpretation of a new set of reliable measures of interdisciplinarity and novelty. In addition, we show that, in spite of the conceptual differences between interdisciplinarity and novelty (i.e. interdisciplinary research is not always new and novelty can result from mono-disciplinary research or from the integration of close knowledge)(March, 1991; Hall et al., 2001; Uzzi et al., 2013), the measures applied in literature often fails to clearly distinguish between the two. Finally, we appreciate the importance of interdisciplinarity and novelty in the dynamics of science by relating them to the relevance of scholarly contributions, which is proxied by citation impact. The analysis exploits a sample of 231450 focal articles published on 8 journals of the American Physical Society between 1985 and 2005, 355092 citing articles (1985-2015), and 2439359 total citations. We use the codes of the Physics and Astronomy Classification Scheme (PACS) at the 2-digit level to compute the interdisciplinarity degree and to detect novelty. We conduct an econometric exercise estimating different models - OLS, Tobit Models, Negative Binomial Generalized Linear Models - to determine the effect of interdisciplinarity and novelty on impact. Regression analyses show that the effect of Interdisciplinarity on Impact is diversified across indicators. Balance and Disparity shows an inverted U-shaped effect: articles that have a prevalent (but not unique) field in references and recombine not too distant knowledge are more impactful. Results regarding novelty are not conclusive and evidence problems of novelty detection in focal articles.

The paper is organized as follows: Section 2 synthesizes the prevalent approaches to the measurement of IDR and novelty and their findings on the relationship with impact. Section 3 surveys and discusses the main IDR and novelty indicators highlighting differences in interpretations and introduces the relevance of the various structures of the knowledge space. Section 4 draws the attention to the distorting effect of disciplines size and growth variations on the reliability of proximity, IDR, and novelty indicator. The issue is dealt with a direct normalization procedure that avoids transformative effects on the knowledge space. Section 5 introduces the dataset and implement the proposed procedures. Section 6 presents regression analyses on the relationship between Interdisciplinarity, Novelty and Impact. Section 7 discusses the relevance of the findings for bibliometric analysis and policy evaluation and introduces the FIMS protocol for the measurement of interdisciplinarity in research, Section 8 concludes.

2 Theoretical background

In the economics of innovation, the production of new knowledge occurs through the recombination of previous knowledge (Nelson and Winter, 1982, p. 130; Schumpeter, 1939, p. 88). When the knowledge used in the process of integration comes from different disciplines or subfields, a research is referred to as interdisciplinary (IDR). The most diffused interdisciplinarity indicators in the literature are Variety, Balance, Disparity, and Integration Score. Variety counts the number of disciplines involved in knowledge production, Balance measures their distribution and Disparity evaluates their distance. The Integration Score compounds the other three measures in a synthetic indicator (for the detailed discussion of measures see section 3). Independently on the chosen measure, IDR indicators capture aspects of knowledge integration that are prerequisite for Novelty intended as exploration (March, 1991; Hall et al., 2001). This kind of Novelty is connected to the presence of new or unusual recombination of previous knowledge. For instance, Uzzi et al. (2013) define Novelty as atypical combinations of referenced journals, whereas Wang et al. (2017) define novelty in articles that make unprecedented combinations in the referenced journals. In some cases differences are minimal, and these indicators are used for measuring both IDR and for Novelty (e.g. Uzzi et al., 2013). The literature shows little interest on the property of measures and implications, while as the policy pushes for interdisciplinary and novel research increases (Section 7), scholars focus on their relationship with impact. 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. Adams et al. (2007) use the same measures on articles published by two Universities in the UK finding visual evidence for an inverted U-shaped relation between IDR and impact. Rinia et al. (2001) calculates Balance at the journal level for the publications of physicists in the Netherlands finding no effect of IDR on impact. Levitt and Thelwall (2008) examine all science and social science articles indexed in WoS and Scopus computing Variety at the journal level. Their results posit a negative relationship between IDR and impact in some disciplines. Larivière and Gingras (2010) compute Balance at the article level for all the papers published in WoS and identify an inverted U shape between IDR and impact. Larivière et al. (2015) measures IDR through 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. The same measures are applied by Yegros-Yegros et al. (2015) at the article level to a sample of 62408 articles indexed in WoS in cell biology, electrical and electronic engineering, food science and technology, and atomic, molecular and chemical physics. They find an inverted U-shaped relation between the different dimensions of IDR and impact, but no relationship between the Integration Score and impact. Porter and Rafols (2009) analyse the evolution of IDR between 1975 and 2005 in biotechnology and applied microbiology, engineering, electrical and electronic, mathematics, medicine, neuroscience and atomic, molecular and chemical physics. They detect IDR computing Variety, Balance and Integration Score on WoS subject categories in article references and conclude that the share of IDR is constant over time if the compound measure is considered. As for Novelty, Uzzi et al. (2013) depict the relationship between novelty and impact for 17.9 millions of article in all the disciplines included in the Web of Science (1950-2000). Novelty has a positive effect on impact only if it is counterbalanced by Conventionality. Wang et al. (2017) explore the link between novel research and impact for 661,643 papers in all disciplines published in the Web of Science in 2001. Novelty is defined for papers that make unprecedented combinations in the referenced journals and is computed for each paper as the sum of the distance of novel combinations. Papers making more distant combinations exhibit a higher variance in impact signalling the riskiness of undertaking novel research. Lee et al. (2015) study the relationship between research team size and novelty (measured as commonness of journal combinations) in 1493 articles in science, engineering and social science (2001-2006) finding an inverted U-shaped relationship. Carayol et al. (2018) propose a measurement of novelty based on the frequency of pairwise combinations of author keywords applied to the ten million research articles published over 1999-2013 by journals indexed in the Web of Science (WoS). Similarly to Wang et al. (2017), they find a positive effect of novelty on citations associated with a higher risk. In what follows the paper will show that such a variety in evidence can be generated by the absence of proper measurement procedures and by the presence of biases.

3 Non-neutrality of Interdisciplinarity and Novelty measures

Despite the conceptual differences between interdisciplinarity and novelty, their operationalizations are similar and are based on the definition of knowledge proximity – or similarity – and representativeness of disciplines in the production of knowledge. In the next section we examine measures in details.

3.1 Interdisciplinarity

Figure 1 summarizes the three main dimensions of knowledge integration: Variety, Balance and Disparity. These dimensions are often compounded in the Integration Score (Stirling, 2007). Let a be the object of analysis (e.g. a focal article), v_a the Variety of the disciplines represented in references (R, Y, G are hypothetical disciplines), f_{ai} and f_{aj} respectively the frequency of discipline i and j in paper a, p_{ij} the proximity between discipline i and j.

Variety counts the number of different disciplines involved in knowledge production. Balance (1) refers to the evenness of their distribution and in figure 1 it is operationalized as the normalized Shannon entropy, returning a value between 0 and 1 (with this normalization the measure is independent of Variety). Disparity (2) 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 (see section 4.1). Disparity is defined for values between 0 and 1 and is independent of Variety and Balance. The Integration Score (also known as Rao-Stirling diversity index) (3) aggregates the previous dimensions; it returns values between 0 and 1. For all the measures higher values correspond to more interdisciplinary research. Intuitively, Integration Score is a stand-alone measure, but Variety, Balance and Disparity are distinct concepts. Imagine that the values in Figure 1 are computed for the references of a hypothetical article a. Variety is the prerequisite for all the other enquiries but it is not exhaustive. The three disciplines in the references of a are not equally distributed, the discipline red is the main domain from which knowledge has been drawn. A Balance equal to 0.83 shows that the distribution is not very uneven (lower values of Balance signal more skewed distributions). Notice that if we compute only Balance we lose information on how many

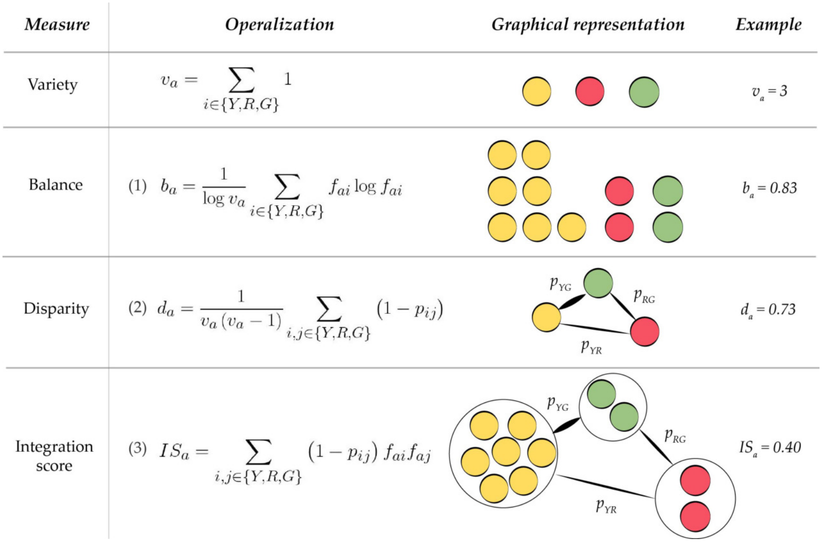


Figure 1: Measures of knowledge integration. In the graphical representation of the measures, circles of different colours represent the different disciplines in article a references, and thicker links between these disciplines indicate greater proximity

fields participate in the production of knowledge.[1] Disparity also adds to the information conveyed by Variety including the measurement of the distance between fields. In a a value of 0.73 shows that knowledge has been integrated starting from distant disciplines. Notice that the value of Disparity could be lower, for the same values of Variety and Balance, when average proximity between disciplines is high. The information on number and distribution of fields is lost. The Integration Score can be thought of as an alternative index in that it merges the previous three dimensions. Its interpretation must be cautious since it is not possible to ascertain which dimension among Variety, Balance and Disparity drive its value (for an example see 5.3).

[1] In literature Shannon entropy without normalization is also used as a compound measure of Variety and Balance (Leydesdorff and Rafols, 2011; Wang et al., 2015).

3.2 Novelty

Independently on the chosen measure, IDR indicators capture aspects of knowledge integration that are prerequisite in novelty, intended as exploration. This kind of novelty is usually connected to the presence of new or unusual recombination of previous knowledge. However, detecting actual breakthroughs among researches that recombine distant knowledge is not straightforward and it is not surprising that the operationalizations of novelty have several similarities with IDR measures. In some cases (e.g. Uzzi et al., 2013), differences are even minimal, and these measures are referred to as both IDR and novelty indicators in literature. Uzzi et al. (2013) define Novelty through the identification of atypical combinations of referenced journals. They consider the cumulative distribution of the proximity between pairs of knowledge combinations and heuristically define the degree of Novelty for each article as the tenth percentile of such distribution to express the idea that novelty resides in more distant combinations. Conventionality is the median of the same distribution.

For Wang et al. (2017) Novelty is defined for articles that make unprecedented combinations in the referenced journals and is computed for each paper as the sum of the distance of novel combinations:

$$Novelty_W = \sum_{i,j\ pairisnew} (1 - p_{ij}) \tag{4}$$

Where i, j is 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 profiles. They conduct a two-step identification of novelty in articles. Firstly, they select articles making an unprecedented combination of journals in references. These articles are defined as novel. Secondly, they compute proximity among journal pairs only for unprecedented combinations in novel articles. This approach is problematic. Conceptually it is based on the assumptions that novelty is more likely to spill from unprecedented recombinations and from recombination of distant knowledge. This approach rules out all rare knowledge combinations that appear in the papers. A possible contradictory outcome is that for Wang et al. (2017) an unprecedented combination of a relatively closer knowledge could have a higher novelty indicator than a rare combination of more distant knowledge.

Moreover, defining novelty as recombination of distant knowledge has several drawbacks. On the one side, it is questionable to assume that novelty emerges primarily from the integration of distant knowledge or from infrequent combinations. Focusing on atypical combinations, as admitted by scholars (Uzzi et al., 2013), underestimates the actual generation of novel ideas. Not only because establishing the distance above which knowledge can be considered novel is arbitrary but also because breakthroughs that come from mono-disciplinary research or from combinations of similar knowledge are not captured. On the other, even if one accepts that potential novelty derives mainly from recombination of bits of existing knowledge it is not possible to say whether the integration of new or distant inputs (journals, references, disciplines) will actually translate in articles that contain novel knowledge.

3.3 The structure of the knowledge space

Finally, IDR and novelty literature neglect that the variability in results can also be attributed to the definition of the knowledge space that is used to compute the distance between disciplines in IDR and Novelty (Eqs. 1, Figure 1, - 4). The structure of the knowledge space depends on how the underlying citation network is modelled since the detection of IDR and novelty is usually performed at the reference level. The choice determines both the value and the interpretation of the proximity measure. Proximity between disciplines can be computed as the number of co-citations (Uzzi et al., 2013) or as the number of citations (Leydesdorff and Rafols, 2011; Yegros-Yegros et al., 2015) among disciplines. These procedures create networks between disciplines, in which links are weighted by the proximities between the nodes. These two definitions of proximity are not strictly equivalent and the knowledge space that we obtain from them is different: proximity in co-citations reflects the heterogeneity of the referenced disciplines, whereas citations mirror the proximity between the disciplines of the focal article and those of the referenced articles. Co-citations generate a symmetric matrix of proximities, which measures how frequent is the recombination of two disciplines in the production of new knowledge. On the contrary, the use of citations results in an asymmetric matrix that reflects the source/target relationship. The asymmetry in the proximity matrix is a sign of the unbalance between the roles of knowledge source and target between two disciplines. It is straightforward to conclude that the two proximity matrices are likely to be sharply different.

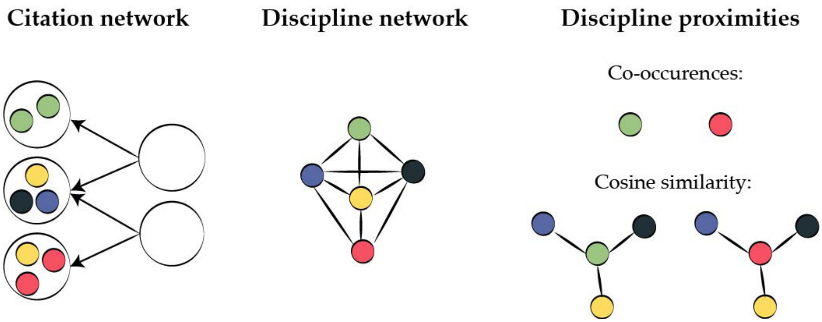


Figure 2: Measures of proximity among disciplines based on co-citations. Citation network (left) is used to define a discipline network (centre) where links are weighted by the number of co-citations between two disciplines. Direct and indirect computation of proximity, based on discipline network, may lead to discordant results. In the example, direct proximity (top right) is zero (no co-citations), whereas the indirect proximity (bottom right) is 1 (same co-citation behaviour)

Comparisons between proximities computed on different citation networks are also made difficult by the unbounded range of values of absolute numbers. A very diffused (Wang et al., 2017; Yegros-Yegros et al., 2015) form of normalization is cosine similarity, an indirect measure of proximity which compares co-citation or citation behaviour in different disciplines by considering the whole structure of the disciplines network. However, cosine similarity creates a new structure for the knowledge space that is likely to be different from the one derived from direct measures of proximity. An intuitive difference between these approaches is represented in Figure 2. Notice that very distant disciplines in terms of co-citations may have high cosine similarity proximity (Figure 2, bottom right). Cosine similarity on direct proximities returns normalized and comparable values, but measures citation patterns rather than the direct relationship between two disciplines and their easiness of recombination (for an example see Section 5.2 and Appendix A).

4 Non-reliability of interdisciplinarity and novelty measures

Alstott et al. (2017, pp. 4–5) have shown that the probability of integration between two disciplines is affected by several factors that do not strictly relate to IDR or novelty and that vary considerably in time and across disciplines. Firstly, the probability that any discipline appears in the references of an article depends on its absolute and relative size and on how these change in time (hereafter, size effect). In addition, the numbers of papers per discipline, the references per paper, and the average citations have increased in time in all disciplines making intertemporal comparisons hardly reliable (hereafter, growth effect). Size and growth effects affect two aspects of the measurement of interdisciplinarity and novelty: the calculation of proximity and the probability of knowledge recombination/integration. Since the probability of cocitations and citation depends on the size of the disciplines, the resulting proximity also captures size and therefore it could be spurious. For instance, two relatively small fields i and j may be as proximate as two relatively big fields r and s, but their empirical proximity p_{ij} and p_{rs}, computed as the number of co-occurrences between fields, will be respectively smaller and higher, due to the discipline size. It follows that all the indicators that include proximity (Disparity, Integration Score, Novelty) could be unreliable. The size of the disciplines has the same effect on the probability of recombination in that bigger fields are more likely to be referenced. It is worth noting that all indicators, including Variety and Balance, are based on the probability of recombination. Biases in the computation of proximity and in the probability of recombination add up in the indicators of Disparity, Integration Score, Novelty, making the issue more critical.

In section 4.1 and 4.2, we propose to adopt Null Configuration Models that with respect to cosine similarity allow to obtain a non-transformed knowledge space on which to calculate an unbiased proximity measure, and serve as a benchmark to detect the importance of biases in the probability of recombination in IDR and novelty measures.

4.1 Introducing reliable proximity

A common way to deal with these biases is to build a knowledge space based on cosine similarity (Wang et al., 2017; Yegros-Yegros et al., 2015). This transformation partially accommodates size effects be-

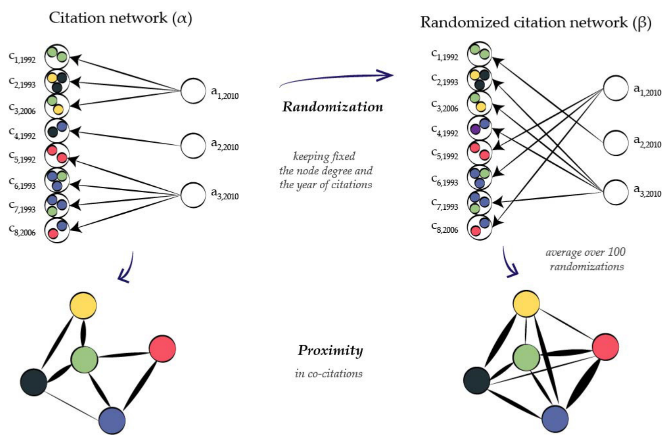


Figure 3: Configuration Null Model and Proximity computation. The randomized citation network (β) is obtained by applying a CNM to the empirical citation network (α). Both networks are used to derive a proximity between disciplines (coloured circles), defined as the number of co-occurrences between fields in article references, and the corresponding discipline network (bottom). Thicker links identify greater proximity

cause it compares two disciplines A and B through their direct proximity (computed as number of co-citations or citations) with every other discipline C: the size of C will affect both the proximity between A and C and the proximity between B and C, so the comparison between them is partially independent of the size of C. Cosine similarity, however, has the undesired effect of transforming the knowledge space (see the previous Section) and cannot deal with the growth effect. Eck and Waltman (2009) suggest that a direct and more intuitive normalization should be preferred.

A direct normalization of proximity is obtained through the definition of and the comparison to a Configuration Null Model (CNM) (Newman and Girvan, 2004). A CNM creates a randomized article-citation network that preserves the degree of each node in the original graph and controls for the changes in citation behaviour in time. The resulting randomized graph preserves all structural properties of the original graph and the sizes of disciplines over time. It is therefore used to define a randomized knowledge space and a proximity measure (based on the number of co-citations or citations, as in the empirical graph), characterized by the same impinging factors that affect the empirical direct measure. Through the comparison of the empirical and randomized proximity, we obtain an unbiased measure and a reliable knowledge space (Alstott et al., 2017).

Figure 3 synthesizes the procedure for the generation of the CNM. It applies to the computation of proximity in co-citations and citations (computed as disciplines co-occurrences), and for references, hereafter labelled - for the sake of simplicity - as citation network. Starting from the empirical citation network (α) for all papers a, links are randomly redistributed while preserving the degree of nodes and the years of citing and cited papers (Upper part of Figure 3). Randomization is repeated 100 times (β) to explore various random configurations. Proximity matrices are computed for the empirical (p_{ij} where p is proximity for discipline i and j) and the randomized networks (p_{ij}^r). Proximity is calculated by counting the number of co-citations or citations between disciplines of papers in the networks (lower part of Figure 3). We then obtain a stable value for p_{ij}^r by taking the average over the repetitions: $\overline{p_{ij}^r}$. We calculate an unbiased indicator of proximity by taking the ratio of the empirical proximity to the average randomized proximities:

$$p_{ij}^n = \frac{p_{ij}}{\overline{p_{ij}^r}}. \tag{5}$$

p_{ij}^n ranges between 0 and ∞. When it is equal to 1, the number of co-citations or citations in the empirical and randomized networks is the same. This amounts to say that the empirical proximity between discipline i and j is driven by the simultaneous presence of (relative) size and time effects. Values of p_{ij}^n between 0 and 1 identify atypical combinations of disciplines while values greater than 1 signal frequent combinations. A further transformation allows obtaining a symmetric proximity measure that ranges between 0 and 1:

$$p_{ij}^{nt} = \frac{\arctan p_{ij}^n}{\pi/2} \tag{6}$$

Values of p_{ij}^{nt} between 0 and 0.5 identify atypical combinations of disciplines, while values between 0.5 and 1 signal frequent combinations. The measures obtained from normalization are comparable since they are cleansed of the impinging factors.

In the remaining of the paper, all the calculations and the regressions use the proximity calculated as in 6, if not stated otherwise.

4.2 Introducing reliable IDR and novelty measures

As highlighted in Section 4, the definition of a reliable proximity controls only partially for the effects that size and growth of disciplines have on interdisciplinarity and novelty measures: the probability of recombination of previous knowledge is connected to a discipline's size at a given time, as well as to the structure of the citation network (i.e. the number of references in each article) and therefore, even after the normalization of proximity, measures cannot be taken as entirely reliable. The CNM normalization procedure proposed in 4.1 cannot be applied as such to solve this issue but the CNM is still useful to identify the importance of size and growth effects. In more detail, the proposed normalization procedure is applicable to proximity, since it is an aggregate indicator based on the total number of co-occurrence between disciplines (e.g. provides information on the entire network), therefore the comparison proposed in Equation 5 between the empirical and the randomized networks is meaningful. The same does not hold for IDR and novelty indicators that are computed at the article (node) level. While in the computation of proximity the coherence of comparisons is guaranteed by the fact the overall properties of the network are kept constant, extrapolating an article (node) from the network severs the logical connection between the empirical and the randomized references. It follows that comparisons in the style of the previous section are not possible when looking at the properties of a node i.e. for the IDR and Novelty of an article.

However, we can still use the randomized networks to control for size and growth effect that as explained in Section 4 are not ruled out by normalizing proximity. In this procedure, the randomized network is used as a benchmark to detect distortions in IDR and novelty measures and in their relationship with impact. We proceed in two steps. First, we compare the distributions of the measures computed on the empirical and on the randomized networks both with normalized proximity. If they are too similar, the measures cannot be taken as entirely reliable since the empirical distribution would be driven by growth and size effects (see Figures 7 and 10 below). For instance, articles with a prevailing field would have a low Balance. Without this procedure it would be impossible to say whether that specific field is actually important in the production of knowledge or simply over-represented due to its size in the sample. Second, we can use the randomized networks to corroborate the relationship between IDR and novelty indicators and an external variable. In our case, the obvious candidate is impact. We conduct our test visually comparing the plots of measures versus impact and/or by conducting regression analyses to check whether the effect of the measures is different in empirical and randomized networks. If there are no significant differences, the relationship between indicators and impact is not meaningful. In this case the results should be discussed carefully or discarded. We will return on this aspect in Sections 5.3, 5.4, and 6.

5 Data

We investigate narrow interdisciplinarity[2] and novelty in Physics (Pan et al., 2012; Sinatra et al., 2015) as an exemplification of the concepts illustrated above, and we expand the discussion including the relationship with IDR and novelty.

We exploit a sample of articles in Physics from the American Physical Society (APS) database. The original database includes 577870 articles published on the 13 APS journals from 1893 to 2015, their citations on APS journals and, starting from 1970 and widespread from 1985 (more than 90% of articles in the database has a PACS code, see Figure 15 in Appendix A), a 6-digit classification (self-attributed) known as PACS codes. The PACS is the subject classification system of the American Institute of Physics (AIP) for categorizing publications in physics and astronomy. Its consist of 10 top-level categories that represent broad fields. Its hierarchical structure allows to progressively identify more specific research areas and sub-fields up to five levels of successive specifications. PACS are better proxies for sub-fields with respect to 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., 2018). On the other hand, keywords are not standardized and the use of different terms to identify the same topic

[2] Within discipline interdisciplinarity is defined as narrow interdisciplinarity. In narrow interdisciplinarity the interaction between fields is in principle less challenging than in broad IDR at least in epistemological terms since the concepts, theories and/or methods are relatively similar in their presuppositions. However, the underlying bodies of knowledge still refer to specialized and distinct domains. Intuitively, the interaction between, say scholars in Condensed Matter Physics and in Astrophysics requires the merging of diverse expertise that is typical of IDR and novelty.

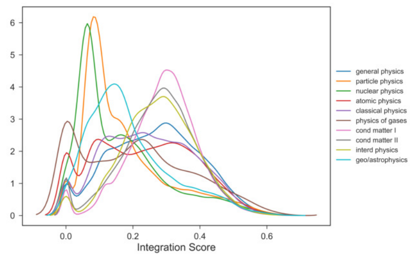


Figure 4: Distribution of Integration Score of articles by 1-digit PACS code

artificially expands the knowledge space. In both cases, the overestimation of IDR and novelty is likely to occur. The use of codes (e.g. JEL in Economics or PACS in Physics) is possibly a more appropriate choice if their structure is stable in time. [3]

In order to perform our analysis we select a subset of the APS database. Since PACS codes are available from 1970 and widespread from 1985 and we want to analyse a 10-year window of citation for each article, we focus on focal articles published between 1985 and 2005 with at least one reference and one PACS code in references. Our dataset is therefore composed by 231450 focal articles (1985–2005), 203910 referenced articles (1970–2005), 355092 citing articles in the first 10 years from publication of the focal articles (1985–2015), and 8 journals. Each article can have a maximum of 5 PACS. We take a ten-year window of citations in order to observe the same time span for all focal articles. We consider PACS at the 2-digit level. Only PACS that are represented in the entire period and that are not rare – i.e. occur at least fifty times - are taken into account (64 – over 79 in total - PACS at the 2-digit level, 10 PACS at 1-digit level). Tables 1a and 1b describe the composition of the sample. Impact factors are reported only for 2005. All IFs exhibit a similar moderately growing trend for 1985-2005 without any effect on journals ranking.

Although we are referring to a single discipline, data show that the

[3] An interesting way to overcome the problem is to look for novelty within papers as in Kaplan and Vakili (2015). They use topic modelling to elicit the formation of new topics in patents. Those patents that originate new topics can be thought of as breakthroughs, Novelty here is based on the Kuhnian tradition according to which shifts in language signal new ideas that require a new language for illustration. This would allow decoupling IDR from Novelty and Impact.

Table 1: Descriptive statistics

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

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

(b) Descriptive statistics of articles per journal

Journal	Number of articles	Average number of references	Average number of citations	Impact factor (IF) 2005
PRA-Physical Review A	26363	8.92	9.40	3.00
PRB-Physical Review B	73748	10.43	9.45	3.19
PRC-Physical Review C	13433	8.10	8.35	3.61
PRD-Physical Review D	24690	9.79	11.73	4.85
PRE-Physical Review E	20838	8.57	7.01	2.42
PRL-Physical Review Letters	47638	7.46	19.29	7.49
PRSTAB-Physical Review Accelerators and Beams	405	4.71	5.53	1.70
RMT-Reviews of Modern Physics	276	55.56	98.81	30.25

sub-fields attitude towards IDR is quite variegated. PACS codes distinct behaviour in IDR corroborates the significance of our analysis. Figure 4 shows the Integration Score with normalized proximity for the PACS at the 1-digit level. Particle Physics, Nuclear Physics and Geophysics and Astrophysics exhibit a low level of IDR, whereas Condensed Matter I and II and Interdisciplinary Physics are more prone to

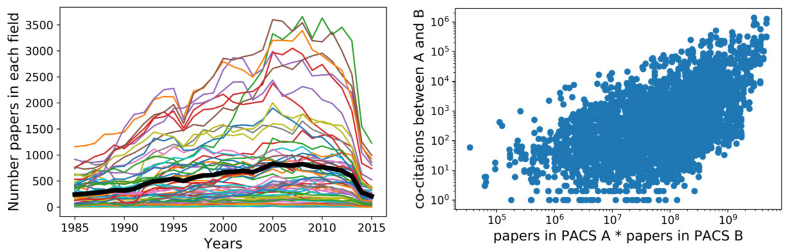


Figure 5: Impinging factors in Physics. Plots show the evolution of the number of articles per PACS codes (left) - black solid line represents the growth of article in time -, and the number of co-citations between PACS codes versus their size (right)

IDR.

5.1 Size and growth effects in Physics

Similarly to other scientific domains, Physics has experienced a growth in the number of published and indexed papers and an increase in the number of references and citations per paper. By plotting the trend of the number of articles per PACS and the correlation between the number of co-citations between any pair of PACS codes and their size (Figure 5) we confirm that the growth is heterogeneous across PACS codes (left panel) and that the likelihood of co-citation is affected by the size of the considered PACS code (right panel). In what follows we demonstrate the effects of normalization on the structure of the knowledge space, and we then discuss the non-neutrality of normalized IDR and novelty measures.

5.2 Knowledge space(s) in Physics

In section 3.3 we have argued that the structure of the knowledge space depends on how the information based on the underlying citation network is modelled and normalized (see Figure 2). The choice determines both the value and the interpretation of the proximity measure and therefore affects the calculation of IDR and Novelty.

Figure 6 shows the knowledge space in references computed on co-citation normalized with CNM (left panel) and with cosine similarity (right panel). We focus on co-citation network because in mapping the proximity of referenced knowledge it constitutes a more appropriate proxy for knowledge integration than the citation one. The latter is better suited to approximate the distance between knowledge used in the production of an article and the knowledge embedded in that article. Intuitively, the proximity between PACS codes results remarkably different. For instance, Solar system; Planetology and Solid earth Physics have proximity equal to 0.89 in co-citation space and to 0.64 in cosine similarity space. The discrepancy in values depends on different citation behaviour. On the other hand, the proximity between Acoustics and Equations of state, Phase equilibria, and Phase transitions is very low (0.20) when co-occurrence is considered and very high under the cosine similarity (0.78). For these PACS codes the opposite situation hold: they have similar citation patterns – they are often co-cited with other PACS codes – but seldom co-cited between themselves.

Data confirm the transformative effect of the normalization through cosine similarity with the subsequent interpretative differences and with an appreciable effect on the values of IDR and Novelty measures that include proximity (or distance). The proposed direct normalization through CNM is superior under many respects: it does not modify the knowledge space avoiding interpretative issues and also accounts for growth effects that are not considered in cosine similarity. Similar results hold for knowledge spaces computed with citations (see Appendix

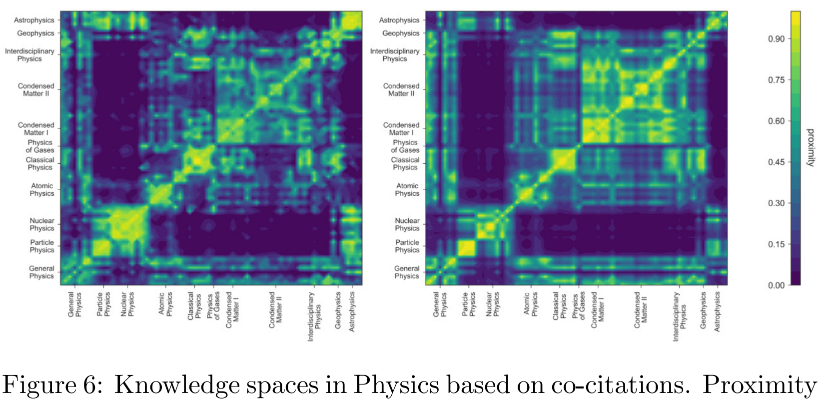


Figure 6: Knowledge spaces in Physics based on co-citations. Proximity among fields is computed with CNM (left) and cosine similarity (right). Knowledge spaces is calculated for the entire observed period. The robustness of the pattern is confirmed by comparison with knowledge spaces computed on five-year windows

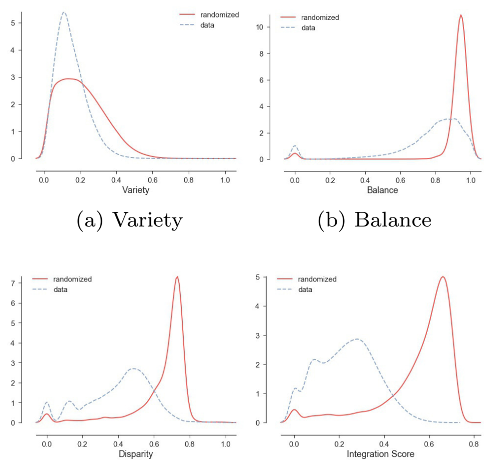
A).

5.3 Facets of interdisciplinarity in Physics

In this section we compute IDR indicators (see Figure 1) with the normalized proximity (Equation 6) . We apply our measures to the empirical and randomized citation networks and test their reliability by comparing their distributions in both networks. Figure 7 shows that distributions are different and therefore we can use them in our analyses.

In section 3 we argued that the choice of IDR indicators is non-neutral with respect to the information conveyed in that each measure emphasizes a facet of the phenomenon. Figure 8 confirms the statement by showing that for a given level of an indicator the others exhibit a remarkable variability.

The same figure also suggests that the immediacy of synthetic indicators implies a considerable loss of information. While for Low and High values of the Integration Score (Eq. 3, Figure 1) there is a sort



(a) Variety (b) Balance

(c) Disparity (d) Integration Score

Figure 7: Distribution of IDR measures in empirical (dashed line) and randomized (solid line) citation network. Variety is normalized between 0 and 1

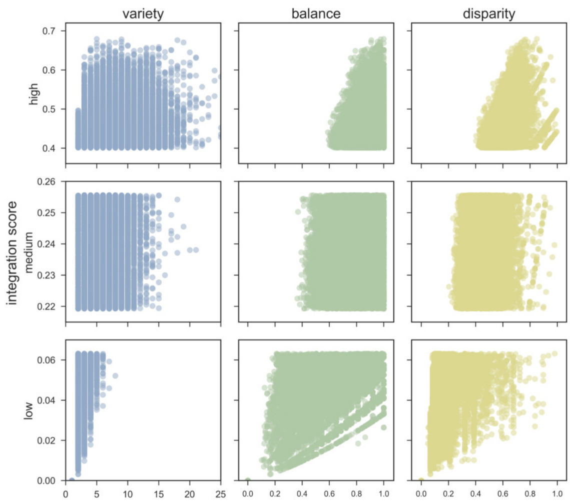


Figure 8: Variety, Balance and Disparity for high (greater than 90th percentile), middle (between 45th and 55th percentiles), and low (less than 10th percentile) levels of Integration Score

of regularity, Medium values correspond to articles that are highly heterogeneous in Variety, Balance and Disparity. In our sample, this has particular relevance since in the medium values of the Integration Score we find the most cited articles.

5.4 Novelty (or interdisciplinarity?)

In this section we also test for the reliability of Novelty measures, as computed by Uzzi et al. (2013) and Wang et al. (2017).

Figure 9 illustrates the procedure to obtain the main variables in Uzzi et al. (2013) analysis. Proximity is normalized with the CNM procedure. Conventionality is defined as the median of the cumulative distribution of proximity among PACS codes pairs in references. Novelty is related to the 10th percentile of the same distribution, bounded between 0 and 1. Low values of 10th percentile correspond to the presence of atypical (non-conventional) combinations in knowledge production and therefore novelty. For the sake of ease in interpretation we define Novelty as:

$$Novelty_U = 1 - 10\text{th percentile}. \tag{7}$$

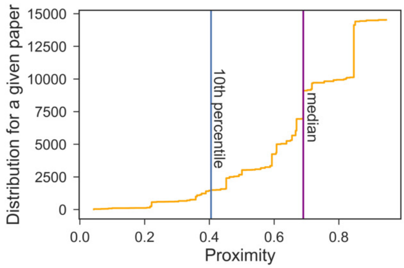


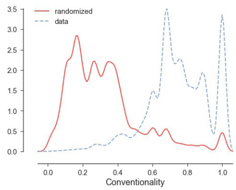
Figure 9: Distribution of proximity among PACS code pairs in article references. Uzzi et al. (2013) define as Conventionality the median of this distribution, whereas Novelty is related to the 10th percentile

While Novelty *à la* Wang et al. (2017) is defined for articles that make unprecedented combinations in the referenced journals and is computed for each paper as the sum of the distance of novel combinations (more details in Section 3.2):

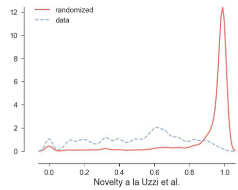
$$Novelty_W = \sum_{i,j\ pairisnew} (1 - p_{ij}) \,, \tag{1}$$

where the proximity between PACS codes is normalized with the cosine similarity, following the original work. In order to create a buffer of 20 years for articles as in Wang et al. (2017), we compute Novelty$_W$ for articles published in 2005. In our data the share of novel papers is the same as in Wang et al. (2017).

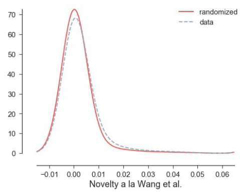
Figure 10 shows that in our dataset the empirical and randomized distribution of Novelty *à la* Wang et al. (2017) are very similar, signalling that their measure might be driven by unobserved factors that do not pertain to Novelty whereas Conventionality and Novelty *à la* Uzzi et al. (2013) are reliable. We can, therefore, rely on the latter to investigate the relationship between IDR and Novelty while we conduct further investigation on Wang et al. (2017) indicators in 6.2.2. Notwithstanding the conceptual differences, Novelty and Interdisciplinarity are measured in similar ways (see Section 3) casting doubts on the nature of information captured by novelty indicators. Intuitively, Uzzi's implementation of measures suggests that high Conventionality corresponds to low Interdisciplinarity while, due to the importance of distance in the determining novelty, high Novelty goes with high Disparity.



(a) Conventionality



(b) Novelty *à la* Uzzi et al.



(c) Novelty *à la* Wang et al. in range (0,0.06)

Figure 10: Conventionality and Novelty distributions in empirical (dashed line) and randomized (solid line) citation networks

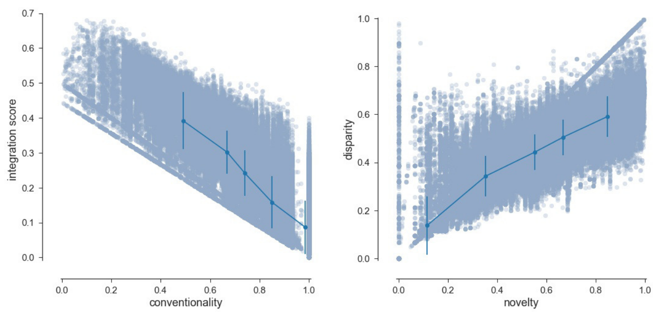


Figure 11: Correlation among IDR and Novelty measures. Pearson correlation between Integration Score and Conventionality is -0.85; between Disparity and Novelty is 0.88

The intuition is confirmed by Figure 11, in which Integration Score and Conventionality, and Disparity and Novelty are respectively negatively and positively highly correlated. The risk of confusing Novelty with the average distance of PACS codes is very high and further enquiries are required (see Section 6.2.1).

6 Impact

IDR research is typically considered more difficult to conduct than the mono-disciplinary one because it involves differences in the background of authors, in the methods, jargon etc. In addition, it is also considered more difficult to disseminate due to the heterogeneity of the potential audience. On the other hand, IDR research is also identified as the natural locus for novel and impactful research (Lee et al., 2015). It is therefore important to identify precisely the relationship between Interdisciplinarity, Novelty, and research Impact. As anticipated in Section 2 there is no definitive evidence on their relationship. In addition, we have discussed in the previous sections all the limitations of the main indicators of novelty and IDR used in the literature. In order to analyse the relationship between the IDR, Novelty and Impact, we calculate Impact using the normalized and absolute number of citations received by each article. In particular, we estimate how Interdisciplinarity and Novelty affect each article's Impact using a set of different regression models. Our dependent variable is the logarithm of the number of citations. This variable is preferred to the logarithm of citations normalized by fields and year because the interpretation of results is easier and more direct.[4] Since there is a set of factors that could be correlated with IDR and Impact confounding the identification of the relationship, we include a set of dummy variables for fields (1-digit PACS codes), journals and years and their interactions (results are reported in Appendix B). In addition, we control for the number of authors and the number of references, since these article characteristics are likely to positively affect the heterogeneity of topics discussed in an article and therefore its number of citations (Lee et al., 2015). Finally, the robustness of our results is tested by introducing different econometric models typically used in the literature: Tobit models and Negative Binomial generalized linear models. The first ones are used to control for the

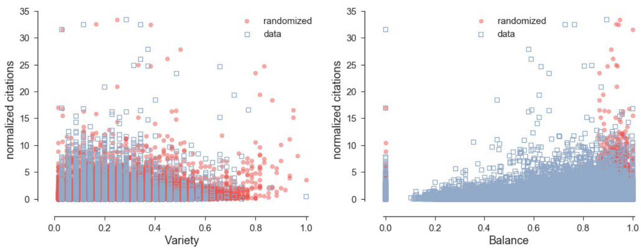
[4] However, to test the robustness of our results we estimated the same models with the normalized citations as dependent variables.

high percentage (around 10%) of zeros in the number of citations (as in Yegros-Yegros et al., 2015), whereas the second ones control for the overdispersion in the citation distribution (following Wang et al., 2017).

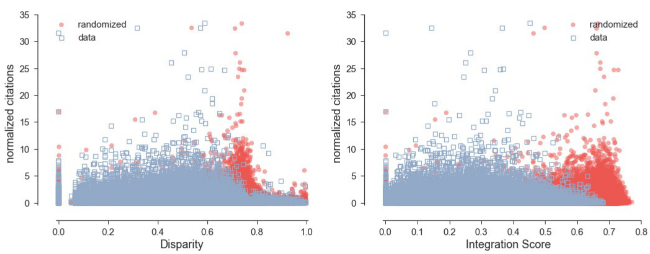
6.1 Interdisciplinarity

The first step is to control for the reliability of our indicators by comparing the empirical and randomized relationship with Impact, as anticipated in section 4.2. We calculate the IDR indicators using the normalized proximity (Equation 6) and we apply them to the real data and to the randomized network. Figure 12 plots the relationship, at the article level, between the IDR indicators (calculated over the two types of network) and the Impact of the article.

Figure 12 reassures that in our database the relationships of IDR measures and Impact in the empirical and randomized network (calculated at the paper level) are significantly different. As a consequence,



(a) Variety (b) Balance



(c) Disparity (d) Integration Score

Figure 12: Impact of IDR and Novelty in empirical (square) and randomized (circle) citation network

we can safely proceed and consider them as reliable indicators. We first run regressions with the proximity normalized by CNM and then with the traditional cosine similarity. In Figure 8 we show that IDR measures convey different information of knowledge integration and that it is not appropriate to collapse all the different dimensions into one single indicator (the Integration Score). As a result, we firstly regress all our three dimensions of IDR, i.e. Variety, Balance and Disparity, on the logarithm of the number of citations. Secondly, we compare these results with the effect of the Integration Score. When the articles have just a single field, Variety is equal to 1 and Balance and Disparity are not defined. So we treat these articles with a separate intercept introducing a dummy variable as a control (Variety = 1). Table 2 shows that, overall, Interdisciplinarity has a negative effect on Impact, but there is heterogeneity across the three dimensions of IDR. Variety has a positive effect whereas Balance and Disparity coefficients show a negative effect on Impact. In addition, when we add a quadratic term for Balance and Disparity we find an inverted U-shaped relationship. In this case, other things being equal, the maximum Impact corresponds to the following values of Balance and Disparity: 0.56 and 0.27 (we use specification (4)). So given the distributions of Balance and Disparity the relation is negative for the top 72% observations.[5]

This implies that articles that have a prevalent (but not unique) field in references have an advantage in citations. Strikingly even if we are restricting our analysis to a single discipline (Physics), the negative effect of Disparity emphasizes that the integrated knowledge must not be too distant in order to acquire citations. Differently from Yegros-Yegros et al. (2015), that find no relationship between the Integration Score and Impact, we identify a negative effect when it is introduced linearly and an inverted U-shape relationship when the squared term is considered. It is key to distinguish the different dimensions of Interdisciplinarity because it is evident that the negative coefficient in specification (5) is mainly due to the effect of Balance and Disparity.

[5] It is important also to control for multicollinearity, particularly after the introduction of quadratic terms. We use orthogonal polynomials to control for multicollinearity (Seber and Lee, 2012). Our results suggest that both linear and quadratic coefficients could be overestimated and the inverted U-shapes have a maximum for lower values of Balance and Disparity. So, there is the possibility that OLS regressions in Table 2 underestimate the magnitude of the negative effects of Balance and Disparity. The effects of the other variables are confirmed. Since orthogonal polynomials centre and scale independent variable values, coefficients are harder to interpret and we do not report these estimations.

Table 2: Impact in IDR. Proximity in interdisciplinarity measures is normalised by using a Configuration Null Model (CNM)

	Dependent variable:					
	log(number of citations + 1)					
	(1)	(2)	(3)	(4)	(5)	(6)
Variety	0.043*** (0.001)	0.038*** (0.001)	0.040*** (0.001)	0.036*** (0.001)		
Balance	−0.485*** (0.017)	1.887*** (0.104)	−0.471*** (0.017)	1.727*** (0.105)		
Balance^2		−1.657*** (0.072)		−1.537*** (0.072)		
Disparity	−0.406*** (0.016)	−0.397*** (0.016)	0.834*** (0.067)	0.704*** (0.067)		
Disparity^2			−1.443*** (0.076)	−1.282*** (0.076)		
Integration Score					−0.144*** (0.020)	0.945*** (0.073)
Integration Score^2						−2.017*** (0.131)
Variety = 1	−0.775*** (0.019)	0.003 (0.039)	−0.549*** (0.022)	0.147*** (0.039)	−0.347*** (0.010)	−0.236*** (0.013)
Number of references	0.030*** (0.0004)	0.030*** (0.0004)	0.030*** (0.0004)	0.030*** (0.0004)	0.041*** (0.0003)	0.041*** (0.0003)
Number of authors	0.006*** (0.0002)	0.006*** (0.0002)	0.007*** (0.0002)	0.007*** (0.0002)	0.006*** (0.0002)	0.007*** (0.0002)
Constant	2.035*** (0.023)	1.269*** (0.041)	1.814*** (0.026)	1.128*** (0.041)	1.581*** (0.017)	1.469*** (0.019)
1-digit PACS	Yes	Yes	Yes	Yes	Yes	Yes
Journal	Yes	Yes	Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes	Yes	Yes
Observations	231,450	231,450	231,450	231,450	231,450	231,450
R^2	0.187	0.189	0.189	0.190	0.181	0.181
Adjusted R^2	0.187	0.189	0.188	0.190	0.180	0.181
Residual Std. Error	0.974	0.973	0.973	0.972	0.978	0.977
F Statistic	1,240.552	1,227.191	1,222.457	1,207.674	1,243.971	1,221.251

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

Adding the squared term, the maximum Impact corresponds to Integration Score equal to 0.24, which is its median value. Therefore its effect is negative for the 50% of articles that have a value of the Integration Score larger than 0.24.

Therefore, as anticipated in section 3, interdisciplinarity indicators are not neutral in determining results. In addition, since the Integration Score does not allow for the identification of the separate effect of the other three drivers, it is difficult to use it as a guide to devise policies and to study the evolution of science. As we have shown in Figure 8, the same values of Integration Score can be generated by very het-

Table 3: Impact of IDR. Proximity in interdisciplinarity measures is normalized by using cosine similarity (CS)

	Dependent variable:					
	log(number of citations + 1)					
	(1)	(2)	(3)	(4)	(5)	(6)
Variety	0.041***	0.035***	0.037***	0.032***		
	(0.001)	(0.001)	(0.001)	(0.001)		
Balance	−0.487***	1.855***	−0.471***	1.777***		
	(0.017)	(0.105)	(0.017)	(0.105)		
Balance^2		−1.636***		−1.571***		
		(0.072)		(0.072)		
Disparity CS	−0.335***	−0.323***	0.380***	0.321***		
	(0.014)	(0.014)	(0.050)	(0.050)		
Disparity CS^2			−0.916***	−0.826***		
			(0.062)	(0.062)		
Integration Score CS					−0.187***	0.574***
					(0.019)	(0.062)
Integration Score CS^2						−1.578***
						(0.123)
Variety = 1	−0.741***	0.028	−0.635***	0.093**	−0.351***	−0.289***
	(0.018)	(0.038)	(0.019)	(0.039)	(0.010)	(0.011)
Number of references	0.030***	0.030***	0.030***	0.030***	0.041***	0.041***
	(0.0004)	(0.0004)	(0.0004)	(0.0004)	(0.0003)	(0.0003)
Number of authors	0.006***	0.006***	0.007***	0.007***	0.006***	0.006***
	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)
Constant	1.995***	1.238***	1.897***	1.180***	1.580***	1.522***
	(0.023)	(0.040)	(0.024)	(0.041)	(0.017)	(0.018)
1-digit PACS	Yes	Yes	Yes	Yes	Yes	Yes
Journal	Yes	Yes	Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes	Yes	Yes
Observations	231,450	231,450	231,450	231,450	231,450	231,450
R^2	0.187	0.189	0.188	0.190	0.181	0.181
Adjusted R^2	0.187	0.189	0.188	0.189	0.181	0.181
Residual Std. Error	0.974	0.973	0.974	0.973	0.978	0.978
F Statistic	1,239.585	1,225.844	1,217.602	1,203.526	1,245.221	1,220.389

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

erogeneous behaviours of the other indicators and therefore variability in the production of science, especially in time, is obscured. Signs and relative magnitude of coefficients are confirmed by estimations with the logarithm of normalized citations as the dependent variable (see Table 7 in Appendix B). Moreover, we tested for biases due to a large number of zeros in citations and to the overdispersion of citation distribution using respectively a Tobit model and a Negative Binomial generalized linear model. Both estimations confirm OLS results as shown in Table 8 (Appendix B).

Regressions with the cosine similarity normalization are reported

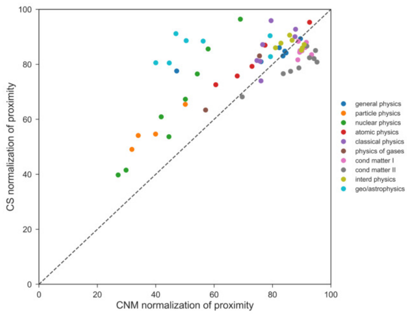


Figure 13: Differences in share of articles for which Disparity is negative under CNM and cosine similarity (CS) normalization. Black dashed line marks equal shares

in Table 3. Signs are confirmed however the magnitude of the effects is different. The negative role of Interdisciplinarity is overestimated since the maximum Impact is obtained for a lower Disparity, equal to 0.18 (corresponding to a negative effect for the 78% of articles - we use specification (4)), and for a lower Integration Score, equal to 0.18 (corresponding to a negative effect for the 54% of articles - we use specification (6)). The share of articles for which Disparity has a negative sign is similar for estimations with the cosine similarity and the CNM normalization, however, given the different topology of the underlying knowledge spaces, the article composition of the shares does not coincide. For 15% of the articles the effect of Disparity is negative or positive depending on the specification. Moreover, disaggregating data at the field level we show that the difference of share of articles for which Disparity has a negative effect is more evident for specific PACS. Figure 13 plots the share of articles for which Disparity has a negative sign in both the CNM and cosine similarity specifications. Condensed Matter I and Condensed Matter II have a higher share with the CNM normalization. This is due to the fact that cosine similarity overestimates the proximity between these fields. The opposite holds for the majority of fields and more markedly for Astrophysics. This shows that the topology of the knowledge space (see Figure 6) generated by the normalization of proximity also impacts at the article level. This evidence stresses the relevance of having an explicit structure of knowledge space in order to correctly characterize the disciplines. Even if the signs of the variables are not altered by cosine similarity, the negative importance of Disparity and Integration Score is overestimated with respect to the CNM normalization. This is due to the transformative effect of the cosine similarity normalization that reduces, in some cases, and amplifies, in others, the distance between fields (see Figure 6). It is worth noting that in the study of Yegros-Yegros et al. (2015) that uses cosine similarity, the Integration Score is not significant.

6.2 Novelty

In the following subsections we further investigate the limitations of Novelty measures, as introduced in Section 5.4. Section 6.2.1 deepens the analysis of the overlap between IDR and Novelty in the definition of Uzzi et al. (2013). Section 6.2.2 performs the tests suggested in 4.2 for Novelty *à la* Wang et al. (2017): visually investigate the relationship between Novelty and Impact and run regression to check whether the effect of this indicator is different in empirical and randomized network.

6.2.1 Novelty *à la* Uzzi et al.

Table 4[6] seems to confirm Uzzi et al.'s analysis according to which articles with high Conventionality and high Novelty are more impactful. Adding a squared term results in an inverted U-shape relationship for both variables, with maximum Impact for values of Conventionality and Novelty equal to 0.87 and 0.64. Therefore their effect is negative for the 20% of articles. Nevertheless, the two variables are percentiles of the same distribution and they are correlated (-0.64). It follows that the estimations can be difficult to interpret. Doubts on the interpretation of Conventionality and Novelty effects are corroborated by their high correlation with Integration Score and Disparity (see Figure 11). Consider first Conventionality: it compounds information in a way that is similar to the Integration Score. To test the actual effect of Conventionality and compare to the effect of the Integration Score, we have estimated a model with Conventionality as the exclusive explanatory variable (specification (3) and (4)). The effect of Conventionality and Integration Score (see specification (5) and (6) in Table 2) are opposite, confirming that articles with high Conventionality (and therefore low Interdisciplinarity) and with low Conventionality (and therefore high Interdisciplinarity) have low Impact.

[6] We control for Variety = 1 since we cannot define a proximity distribution if there is only one PACS code in article references.

Table 4: Impact of novelty *à la* Uzzi et al. (2013)

	Dependent variable:					
	log(number of citations + 1)					
	(1)	(2)	(3)	(4)	(5)	(6)
Conventionality	0.329***	1.379***	0.282***	1.469***		
	(0.016)	(0.071)	(0.013)	(0.066)		
Conventionality^2		−0.786***		−0.883***		
		(0.050)		(0.048)		
Novelty_U	0.065***	0.414***			−0.243***	0.215***
	(0.011)	(0.046)			(0.011)	(0.045)
Novelty_U^2		−0.323***				−0.428***
		(0.043)				(0.041)
Variety					0.045***	0.040***
					(0.001)	(0.001)
Balance					−0.415***	1.841***
					(0.017)	(0.106)
Balance^2						−1.575***
						(0.073)
Variety = 1	−0.366***	−0.242***	−0.386***	−0.330***	−0.670***	0.161***
	(0.011)	(0.014)	(0.010)	(0.011)	(0.017)	(0.038)
Number of references	0.041***	0.040***	0.041***	0.040***	0.029***	0.029***
	(0.0003)	(0.0003)	(0.0003)	(0.0003)	(0.0004)	(0.0004)
Number of authors	0.006***	0.007***	0.006***	0.006***	0.007***	0.007***
	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)
Constant	1.272***	0.888***	1.333***	0.977***	1.926***	1.105***
	(0.023)	(0.030)	(0.020)	(0.028)	(0.022)	(0.040)
1-digit PACS	Yes	Yes	Yes	Yes	Yes	Yes
Journal	Yes	Yes	Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes	Yes	Yes
Observations	231,450	231,450	231,450	231,450	231,450	231,450
R^2	0.182	0.184	0.182	0.183	0.187	0.189
Adjusted R^2	0.182	0.183	0.182	0.183	0.187	0.189
Residual Std. Error	0.977	0.976	0.977	0.976	0.974	0.973
F Statistic	1,226.605	1,182.046	1,255.525	1,235.377	1,236.926	1,200.055

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

Secondly, Figure 11 shows that Novelty *à la* Uzzi et al. has a strong and positive correlation with Disparity: instead of catching novelty, it measures one dimension of IDR. In order to compare the results with the previous specifications in Table 2, we have to control for all the dimensions of IDR and include in the regression also Variety and Balance (specification (6)). These estimations allow a direct comparison with standard IDR measures, reported in Table 2, as suggested by Yegros-Yegros et al. (2015). Specification (6) in Table 4 shows that the effect of Novelty, when Variety and Balance are added as controls, is now negative, exactly as it happens for Disparity in Table 2. This hints that measures chosen to detect novelty might be conceptually detached from their underlying theoretical framework and closer to the detection of interdisciplinarity. Firstly, Novelty *à la* Uzzi et al. (2013) is mainly driven by the distance of disciplines in references and therefore is more properly defined as Disparity. Secondly, controlling for other dimensions of interdisciplinarity (e.g. Variety and Balance), reverses the findings of Uzzi et al. (2013) highlighting a negative effect of "Novelty".

6.2.2 Novelty *à la* Wang et al.

As done for IDR measures we plot the relationship between Novelty scaled between 0 and 1 and Impact, controlling for PACS code and year. Figure 14 shows that, in our data, the relationship between Novelty *à la* Uzzi et al. (2013) is the similar in empirical and randomized citation network. To further investigate these relationships we perform a regression analysis, using Novelty in the randomized network as a control and testing whether the effect of Novelty on Impact is different in the empirical network. In particular, we run a OLS regression with usual control variables on a sample composed of the articles published in 2005 and their analogous in the randomized citation network (the control group). We add a dummy variable that identifies empirical data, used to test whether the relationship is different in the two networks. Table 5 shows that the interaction term between Novelty and this dummy variable is not significant. It follows that the relationship between Novelty *à la* Wang et al. (2017) and Impact is the same in empirical and randomized citation network, thus regressions are not

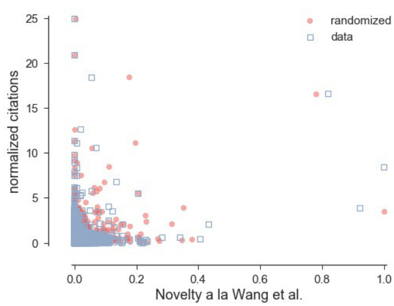


Figure 14: Impact of Novelty *à la* Wang et al. (2017) in empirical (square) and randomized (circle) network

Table 5: Impact of Novelty *à la* Wang et al. (2017), computed on empirical and randomized network

	Dependent variable:
	log(number of citations + 1)
Novelty_W	-3.164^{***}
	(0.471)
Empirical	-0.007
	(0.010)
Novelty_W*Empirical	0.989
	(0.603)
Variety = 1	-0.425^{***}
	(0.034)
Number of references	0.035^{***}
	(0.001)
Number of authors	0.006^{***}
	(0.0004)
Constant	1.274^{***}
	(0.023)
1-digit PACS	Yes
Journal	Yes
Year	Yes
Observations	35,418
R^2	0.225
Adjusted R^2	0.224
Residual Std. Error	0.939
F Statistic	446.573
Note:	OLS regression
	*p<0.1; **p<0.05; ***p<0.01

meaningful and the link between novelty and riskiness (variance in citations) as established by Wang et al. (2017) is not reliable. When this occurs generally it is likely that the relationship between the variables is due to unobserved properties of the data. Specifically, in our case, the overlap between the Novelty-Impact relationship in empirical and randomized data is probably due to the implementation of the measure of Novelty in Wang et al. (2017) (see Section 3.2), that translates in the prevalence of zero novelty articles (88% in our data and 89% in Wang et al. (2017)) and in the presence of both false positive and false negative in the detection of novelty.

7 Discussion

Interdisciplinary research is increasingly seen and consequently promoted as the appropriate way to cope with the growing complexity of the objects of scientific enquiries. The National Academies (2005) and

European Union Research Advisory Board (2004) record an increase in IDR and encourage the creation of interdisciplinary research networks and the funding of interdisciplinary projects. Subsequently, interdisciplinarity appears as a target in the strategic plans of Universities and research centres and ranks as a reward criterion for several funding bodies. Similarly, novelty in research has long been recognized as one of the drivers of economic growth (Grossman and Helpman, 1991; Phelps, 1996; Romer, 1991). The paper highlights that the detection and measurement of IDR and novelty encounter several problems. Results are sensitive to the definition of the concepts (i.e. Novelty *à la Uzzi et al.* vs. Novelty *à la Wang et al.*), indicators are not equivalent, and analyses are likely to be biased due to the presence of size and growth effects in the datasets. Moreover, it also demonstrates that IDR and novelty measures are often correlated stressing that, in spite of the differences, the two concepts are not neatly separated in analyses.

This raises the issue of the reliability of policies designed starting from IDR and novelty indicators and of the ex-post evaluation of their impact. The same holds when the assessment of individual careers is at stake. Moreover, imprecise or poor indicators are likely to introduce distortion in individual and institutional incentives in the balance between exploration and exploitation. The literature is aware of the general relevance of these issues. The Leiden Manifesto (Hicks et al., 2015) proposes ten principles for research evaluation that are inspired by flexibility, transparency and simplicity, while the San Francisco Declaration on Research Assessment (2012)[7] suggests good practices for funding agencies, research institutions, researchers, publishers and agency providing metrics. However, Stephan et al. (2017) that science panels still rely on poor proxies to judge the quality and the impact of research. In spite of the conviction that interdisciplinary and novel research drive economic progress (see for instance the Lisbon Agenda 2000), within the IDR and novelty framework little is done to establish standards on the use of indicators, and the discussion on their absolute and relative goodness languishes.

This paper tries to bridge the gap by identifying and exemplifying the main issues related to the use of IDR and novelty indicators and by proposing a protocol (see Table 6) to investigate IDR and, more generally, to obtain unbiased IDR and novelty measures. We have shown that IDR indicators convey different qualitative and quantitative information on interdisciplinarity and we have warned against the blind use

[7] sfdora.org/read/

of synthetic measures such as the Integration Score (see Figure 8). IDR analyses should extract all the available information computing Variety, Balance and Disparity. If this not done, analyses should motivate the choice of the measure and interpret results accordingly. The paper has also shown that different structures of the knowledge space on which the proximity of the integrated knowledge is computed both for Novelty and Interdisciplinarity emphasise different aspects of the relationship between disciplines. Proximity in co-citations reflects the heterogeneity of the referenced disciplines, whereas citations mirror the proximity between the disciplines of the focal article and those of the referenced articles. It is therefore suggested to explicit the rationale behind the selection of the knowledge space and to propose a coherent interpretation of the resulting proximity. A similar homogeneity is also desirable in the normalization procedures. The presence of size and growth effects in data compels scholars to operate on the structure of knowledge space to obtain unbiased measures of proximity. The paper demonstrates that cosine similarity while partially accommodating for the size effect, creates, de facto, a third type of knowledge space structure that is not coherent with the original one (See Figure 6). Cosine similarity defines proximity as similarity in citation patterns rather than as a direct relation between disciplines or source and target articles. The paper shows that Configuration Null Models avoid the transformative effect of cosine similarity and also accommodates for both size and growth effect. We maintain that, should the proposed normalization procedure be generally adopted, reliability or analyses would substantially improve. Moreover, the paper also draws the attention to an unexplored facet of IDR and novelty measurement. While different types of normalization are commonly used to measure proximity (or distance) it is neglected that the size and growth effects also affect IDR and novelty measures even if proximity is normalized. Intuitively, co-occurrences of disciplines also depend on their relative size and trend. We suggest that comparing the empirical and randomized distribution of measure provides useful information on the relevance of biases and therefore on the reliability of measures. Finally, the paper stresses an open issue that is ripe with implications on the evaluation of research and related policy. Regression analyses show that the effect of Interdisciplinarity on Impact is sensitive to the choice of measures and of the knowledge space and is diversified across indicators. We find that Balance and Disparity have a mixed effect: articles that have a prevalent (but not unique) field in references and recombine not too distant knowledge have an advantage in citations. The empirical analysis reveals that novelty measures are often deprived of explanatory power (Wang et al., 2017) and somehow detached from their theoretical aim (Uzzi et al., 2013). Evaluating novelty in articles is very useful to identify the rate, the loci, the determinants, and the impact of innovation. However, we have shown that the measures adopted in the literature by relying mainly on recombination and distance of knowledge tend to overlap with the measurement of interdisciplinarity. Indeed, according to the theory of recombinant knowledge mixing distant bits of knowledge is a pre-requisite for innovation, but if indicators prevalently capture this facet the measure should be better defined as 'potential novelty' since not all interdisciplinary research is new. If this is the case, and our results witness positively in this direction, establishing a relationship between potential novelty and actual impact is logically inconsistent. We suggest that further investigations in the style of what is done in Section 5.4 should be conducted to evaluate the scope of the problem and to devise more rigorous measures of novelty.

8 Conclusions

The paper investigates the reliability of interdisciplinary and novelty measures in science and of their relationship with impact. In the face of a growing emphasis and intensity in funding of interdisciplinary and potentially novel research, we show that the indicators that evaluate the degree of interdisciplinarity and its impact in terms of citations are still imprecise and under-analysed. The related risks concerning the misallocation of resource and of scholarly rewards based on the relationship between IDR, novelty and impact inspired the analyses conducted in this paper. The vast use of bibliometric indicators based on impact to evaluate the career of scholars and the productivity of institutions[8] suggests that the issue is extremely relevant and must be carefully addressed. The articles show that considerable improvement in reliability of measures and results could be obtained by introduc-

[8] Stephan et al. (2017) p. 412 gives some examples. In Spain, increases of salary based on productivity depend heavily on journal Impact Factors. In Italy, the habilitation is bestowed on the basis of candidates bibliometric profiles and the ordinary. In Europe, the United States and in China, there are formal or informal lists of which journals are more important in assessing candidates for promotion. These measures are also used to allocate resources to universities and, in turn, universities use them to distribute resources to departments.

Table 6: Fontana, Iori, Montobbio, Sinatra (FIMS) Protocol for IDR and Novelty measurement

	Equation	Section	Figure
1. Control for the presence of growth and size effects in data		4, 5.1	5
2. Choice of the structure of the knowledge space		3.3	2, 6
3. Generation of the randomized knowledge space via Configuration Null Model		4.1	3
4. Computation of normalized and unbiased proximity	5,6	4.1	
5. Calculation of diverse IDR measures	(1),(2),(3) in Fig. 1	3.1	8
6. Control for the incidence of growth and size effect on measures		4	12, 14, 17
7. Elimination of measures that fail the comparison with randomized data		4	

ing some refinements in measures interpretation and operalization and proposes a protocol to standardize analyses and improve comparability of results. The proposed protocol is discussed both theoretically and empirically on a dataset of articles published in Physics in the period 1985-2005. The paper shows that the mixed evidence that characterizes the literature can be attributed to the use of different non-equivalent measures of IDR. Their analysis suggests that they should all be computed in that they have different meaning with respect to the facets of knowledge integration. We then bring the attention to the same problem in the choice of the structure of knowledge spaces and we propose a non-transformative representation of the knowledge space that also accommodates for biases generated by the presence of size and growth effect in data. Finally, we demonstrate that while IDR analyses though heterogeneous and uncoordinated rely on a solid translation of the theoretical background in empirical measurement, novelty detection is much more imprecise. We maintain that in looking at atypical of unprecedented knowledge combinations scholars only identify potential innovation and mainly measure a subset of interdisciplinary research. We demonstrate that in some operationalizations novelty is highly correlated with interdisciplinarity, e.g. Uzzi et al. (2013), or is mainly driven from unobserved effects, e.g. Wang et al. (2017). Further investigations are required to improve the detection of novelty in order to reduce potential misallocation of resources or incentive distortion. Overall the paper builds a bridge between different approaches by proposing a standard to measures IDR and novelty in the hope that further discussions and improvements would ensue.

Acknowledgements

The Authors are grateful to the participants to the 29th EAEPE annual conference, Budapest 2017 and to the workshop on Data Science and Economics, Turin 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 in the initial stage of the research. This work was supported by a STSM Grant from the COST Action CA15109 (COSTNET).

A Additional descriptive statistics and knowledge spaces

A.1 Representativeness of articles with PACS codes in APS database

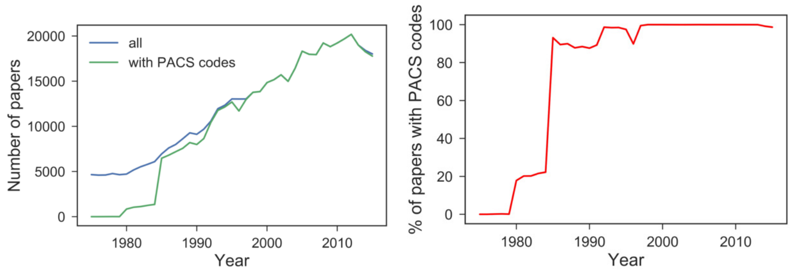


Figure 15: Articles with PACS codes compared to the number of articles in the APS database in time (left) and share of articles with PACS codes in time (right)

A.2 Knowledge space based on citation

Knowledge space based on proximity computed on citations between PACS codes (Figure 16) is asymmetric and slightly differ from the symmetric knowledge space based on co-citations (Figure 6). In both cases normalisations with CNM and cosine-similarity lead to different results.

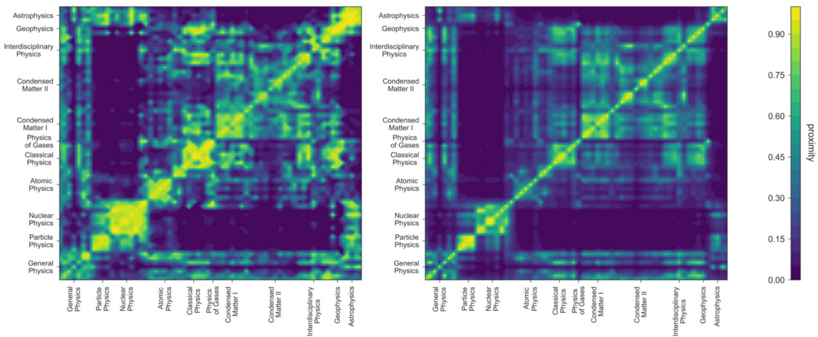


Figure 16: Knowledge spaces in Physics based on citations. Proximity among fields is computed with CNM (left) and cosine similarity (right). Knowledge spaces have been calculated for the entire observed period. The robustness of the pattern is confirmed by comparison with knowledge spaces computed on five-year windows)

B Additional regressions

Table 7: Impact of IDR. Impact is measured both as the number of citations and the number of citations normalized by year and sub-fields

	Dependent variable:	
	log(number of citations + 1)	log(normalized citations + 1)
	(1)	(2)
Variety	0.043***	0.010***
	(0.001)	(0.0003)
Balance	−0.485***	−0.081***
	(0.017)	(0.004)
Disparity	−0.406***	−0.075***
	(0.016)	(0.004)
Variety = 1	−0.775***	−0.126***
	(0.019)	(0.005)
Number of references	0.030***	0.007***
	(0.0004)	(0.0001)
Number of authors	0.006***	0.001***
	(0.0002)	(0.0001)
Constant	2.035***	0.257***
	(0.023)	(0.006)
1-digit PACS	Yes	Yes
Journal	Yes	Yes
Year	Yes	Yes
Observations	231,450	231,450
R^2	0.187	0.161
Adjusted R^2	0.187	0.161
Residual Std. Error	0.974	0.250
F Statistic	1,240.552	1,036.061
Note:		OLS regressions
		*p<0.1; **p<0.05; ***p<0.01

Table 8: Impact of IDR. Comparison among OLS, Tobit model, and Negative Binomial generalized linear model. Proximity in interdisciplinarity measures is normalized by using a CNM

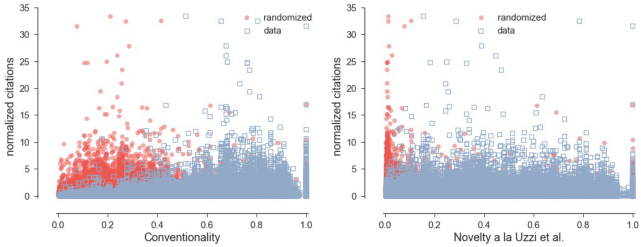
	Dependent variable:		
	log(number of citations + 1)		
	OLS	*Tobit*	*Negative Binomial*
	(1)	(2)	(3)
Variety	0.043***	0.048***	0.044***
	(0.001)	(0.001)	(0.001)
Balance	−0.485***	−0.551***	−0.379***
	(0.017)	(0.019)	(0.017)
Disparity	−0.406***	−0.453***	−0.434***
	(0.016)	(0.018)	(0.016)
Variety = 1	−0.775***	−0.886***	−0.701***
	(0.019)	(0.021)	(0.018)
Number of references	0.030***	0.032***	0.031***
	(0.0004)	(0.0004)	(0.0004)
Number of authors	0.006***	0.007***	0.006***
	(0.0002)	(0.0003)	(0.0002)
Constant	2.035***	2.039***	2.468***
	(0.023)	(0.026)	(0.022)
1-digit PACS	Yes	Yes	Yes
Journal	Yes	Yes	Yes
Year	Yes	Yes	Yes
Observations	231,450		231,450
R^2	0.187		
Adjusted R^2	0.187		
Log Likelihood		−337,462.800	−776,795.400
θ			1.277 (0.004)
Akaike Inf. Crit.			1,553,679.000
Residual Std. Error	0.974		
F Statistic	1,240.552		
χ^2		45,457.420	
Note:			*p<0.1; **p<0.05; ***p<0.01

Table 9: Impact of IDR. OLS with additional control variables

	Dependent variable:		
	log(number of citations + 1)		
	(1)	(2)	(3)
Variety	0.043***	0.046***	0.045***
	(0.001)	(0.001)	(0.001)
Balance	−0.485***	−0.511***	−0.509***
	(0.017)	(0.017)	(0.017)
Disparity	−0.406***	−0.407***	−0.396***
	(0.016)	(0.016)	(0.016)
Variety = 1	−0.775***	−0.786***	−0.785***
	(0.019)	(0.019)	(0.019)
Number of references	0.030***	0.030***	0.030***
	(0.0004)	(0.0004)	(0.0004)
Number of authors	0.006***	0.006***	0.006***
	(0.0002)	(0.0002)	(0.0002)
Constant	2.035***	2.232***	2.084***
	(0.023)	(0.044)	(0.046)
1-digit PACS	Yes	Yes	Yes
Journal	Yes	Yes	Yes
Year	Yes	Yes	Yes
Journal*Year	No	Yes	No
1-digit PACS*Year	No	No	Yes
Observations	231,450	231,450	231,450
R^2	0.187	0.195	0.195
Adjusted R^2	0.187	0.194	0.194
Residual Std. Error	0.974	0.970	0.970
F Statistic	1,240.552	370.223	230.809

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

C Impact of Novelty in randomized network



(a) Conventionality (b) Novelty a la Uzzi et al.

Figure 17: Impact of Conventionality and Novelty *à la* Uzzi et al. (2013) in empirical (square) and randomized (circle) network

References

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

Alstott, J., G. Triulzi, B. Yan, and J. Luo, 2017. Mapping technology space by normalizing patent networks. Scientometrics 110(1), 443–479. 10.1007/s11192-016-2107-y.

Carayol, N., A. Lahatte, and O. Llopis, 2018. The right job and the job right: Novelty, impact and journal stratification in science. Retrieved from: conference.druid.dk/acc˙papers/.

Eck, N. J. v. and L. Waltman, 2009. How to normalize cooccurrence data? An analysis of some well-known similarity measures. Journal of the American Society for Information Science and Technology 60(8), 1635–1651. 10.1002/asi.21075.

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

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.

Hicks, D., P. Wouters, L. Waltman, S. de Rijcke, and I. Rafols, 2015. Bibliometrics: The Leiden Manifesto for research metrics. Nature 520, 429–431. 10.1038/520429a.

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.

Larivière, V. and Y. Gingras, 2010. On the relationship between interdisciplinarity and scientific impact. Journal of the Association for Information Science and Technology 61(1), 126–131. 10.1002/asi.21226.

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, a.-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.

Leydesdorff, L. and I. Rafols, 2011. Indicators of the interdisciplinarity of journals: Diversity, centrality, and citations. Journal of Informetrics 5(1), 87–100. 10.1016/j.joi.2010.09.002.

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

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. and M. Girvan, 2004. Finding and evaluating community structure in networks. Physical Review E 69 (2), 026113. 10.1103/PhysRevE.69.026113.

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.

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., 1991. Endogenous technological change. Journal of Political Economy 98(5), S71–S102. 10.1086/261725.

Schumpeter, J. A., 1939. Business cycles: a theoretical, historical, and statistical analysis of the capitalist process. McGraw-Hill.

Seber, G. A. and A. J. Lee, 2012. Linear regression analysis. Vol. 329. John Wiley & Sons. DOI: 10.1002/9780471722199.

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.

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., 1998. On the economics and analysis of diversity. SPRU Electronic Working Papers Series 28, 1–156.

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.

The National Academies, 2005. Facilitating interdisciplinary research. Washington, DC: National Academies Press.

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

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., 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.

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.

Quaderni pubblicati

1. *Innovation, jobs, skills and tasks: a multifaceted relationship.* M. Piva, M. Vivarelli, Vita e Pensiero, maggio 2018 (ISBN 978-88-343-3654-0)
2. *A bridge over troubled water: Interdisciplinarity, Novelty, and Impact.* M. Fontana, M. Iori, F. Montobbio, R. Sinatra, Vita e Pensiero, settembre 2018 (ISBN 978-88-343-3793-6)

Printed by
Gi&Gi srl - Triuggio (MB)
September 2018

9788834337936