



Revista Brasileira de Inovação

ISSN: 1677-2504

ISSN: 2178-2822

Universidade Estadual de Campinas

Correia, Marcos Paulo R.; Mueller, Bernardo
Pattern making and pattern breaking: measuring novelty in Brazilian economic research
Revista Brasileira de Inovação, vol. 21, e022015, 2022
Universidade Estadual de Campinas

DOI: <https://doi.org/10.20396/rbi.v21i00.8667409>

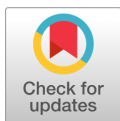
Available in: <https://www.redalyc.org/articulo.oa?id=641771991015>

- How to cite
- Complete issue
- More information about this article
- Journal's webpage in redalyc.org

UABM  redalyc.org

Scientific Information System Redalyc
Network of Scientific Journals from Latin America and the Caribbean, Spain and
Portugal

Project academic non-profit, developed under the open access initiative



Pattern making and pattern breaking: measuring novelty in Brazilian economic research

Marcos Paulo R. Correia* , Bernardo Mueller** 

* Universidade de Brasília (UNB), Brasília (DF), Brasil. E-mail: marcospaulocorreia@gmail.com

** Universidade de Brasília (UNB), Brasília (DF), Brasil. E-mail: [bmuellet@unb.br](mailto:bmueller@unb.br)

RECEIVED: 28 OCTOBER 2021 REVISED VERSION: 27 APRIL 2022 ACCEPTED: 04 JULY 2022

ABSTRACT

How do new ideas emerge in academic contexts and what forces determine which ideas get selected and which are forgotten? We analyze all papers presented at the ANPEC Brazilian Economics National Meetings from 2013 to 2019 using topic modeling and Kullback-Leibler divergence to measure novelty and resonance. In contrast to simply counting citations or reference combinations, these methods explore the Shannon information in the actual texts to detect the rise of new patterns and whether these patterns persist once they have been established. We find that novelty is highly correlated with transience so that most new ideas are quickly forgotten. However, of the ideas that persist, those that are more novel have higher impact. We show that our text-based measure of impact is correlated with subsequent citations.

KEYWORDS: Innovation; Shannon information; Economic research; Emergence of ideas; ANPEC

1. Introduction

Science is all about new ideas. Knowledge progresses as new ideas enable new understanding of the world. At the same time, science is profoundly conservative and resistant to ideas that differ from the current understanding (KUHN, 1996; MERTON, 1968; BOURDIEU, 2004). The production of new ideas is therefore subject to contradictory forces. This tension leads different individuals, research groups, and even entire disciplines to adopt different strategies for pursuing the advancement of knowledge. While novelty seems to be a necessary component of academic success, it is also the case that most new ideas, precisely because they are unfamiliar, remain obscure.

In this paper we empirically analyze the process of the advancement of science through the creation of novelty in a very specific environment: the annals of the Brazilian National Association of Postgraduate Programs in Economics (ASSOCIAÇÃO NACIONAL DOS CENTROS DE POS-GRADUAÇÃO EM ECONOMIA, 2021). This is an interesting context for such inquiry because ANPEC, and Brazilian economists in general, are a large and diverse group that is peripheral yet connected to the major centers of Economic research in the US and in Europe. While most research of this type focuses on the higher echelons of the profession in the major epicenters where research is done, there is less work on how the production of novelty takes place in subsidiary centers.¹ ANPEC was founded in 1973 and by 2022 was formed by 29 member centers, with another 24 centers seeking accession, making it the largest graduate association of economists in Brazil. ANPEC manages a large, unified exam used by many centers to select Master's students (PETTERINI, 2020). It also publishes an academic journal, *EconomiA* (2021), and holds two regional and one national conference each year. All the papers presented in the yearly national conference are published and made accessible in the meeting's annals (ASSOCIAÇÃO NACIONAL DOS CENTROS DE

¹ An exception is Castilla (2020) that analyzes the evolution of research in CEDE, a Colombian economic research center.

PÓS-GRADUAÇÃO EM ECONOMIA, 2021b). We use 1676 papers from the ANPEC meetings from 2013 to 2019 to analyze how novel ideas arise in this context and what is their impact.

Recent advances in machine learning and natural language procession allow large volumes of text to be processed by computers. While research that seeks to measure novelty in large corpora often resorts to counting words or tracking citations, we use a method that is based on information theory (SHANNON, 1948). It seeks, in Bayesian fashion, to identify pattern making and pattern breaking in the annals of the ANPEC meetings.² We first use topic modeling to identify patterns of word use in the papers at a given point in time. This technique yielded a surprisingly strong fit of the 30 topics that emerged from the data to the actual division of the National ANPEC meetings into 13 fields (e.g., labor economics, international economics, political economy, etc.). Each paper is then characterized as a probability distribution over each of the 30 topics. We then use Kullback-Leibler Divergence (KLD) measures to identify when the patterns established up to a point are broken by new, unexpected distributions. Following Barron et al. (2018), we call ‘novelty’ this measured surprise when we expect a certain pattern and instead get a different pattern. Similarly, we call ‘transience’ the KLD of the current text to the patterns that will emerge in the future, that is, when a new pattern that is established by a paper today is not repeated in the following periods. This allows us to then classify papers by ‘resonance’, which are patterns that are high in novelty and that, once created, continue to be repeated in the future. Resonant papers are how science evolves.

Boianovsky (2021) provides a panorama of the “slow coming to age” of Brazilian economists. Several other studies have analyzed the research strategies and productivity of Brazilian economists and linked them to the career and reputational incentives they face. (HADDAD et al., 2017; FARIA, 2004, 2005; FARIA; ARAUJO JUNIOR; SHIKIDA, 2007; ISSLER; FERREIRA, 2004; GUIMARÃES, 2011;

² We follow the approach and code used in Barron et al. (2018), that analyzed over 40,000 speeches in the debates of the National Constituent Assembly that followed the French Revolution.

NOVAES, 2008; among others). Most of this research, however, is focused on productivity and the choice between quantity versus quality in publishing, using information about citations and journal impact factors. Our paper is related to this literature, but our interest is in the choice between novelty and conventionality, based on the detection of patterns within the texts.

Our results allow us to classify papers, areas and departments in ANPEC meetings by novelty and by resonance. We validate these results through comparisons with the actual citation pattern of the papers in the ensuing years. Our first result is that more novel papers are generally more transient. This result, that what is new tends to disappear quickly, has been found in a wide variety of contexts (BARRON et al., 2018; BOUDREAU et al., 2016; WANG; VEUGELERS; STEPHAN, 2017; MURDOCK; ALLEN; DEDEO, 2017; JING; DEDEO; AHN, 2019). The second result is that resonant papers tend to have higher than average novelty. That is, if a researcher wants to make a mark, it is helpful to break with established patterns and display more novelty, however, thereby also risking obscurity. Of the 1679 papers presented from 2013-2019, 78% were not published in an academic journal by 2021 and 56% had no citations. These numbers suggest that overall, Brazilian economists take a conservative approach to topic choice and development, often eschewing the search for novelty to gain acceptance in the meetings, but at the risk of higher obscurity.

What are the reasons for this conservative academic stance of Brazilian economists? Although it may be due to ideological and cultural factors, it is also affected by the personal and career incentives faced in the Brazilian academic milieu of universities, departments, and scholarly associations, including ANPEC (CHECCHI; DE FRAJA; VERZILLO, 2021). These incentives can be both monetary and reputational. Having a paper accepted for presentation can bring benefits, especially to early career academics. The acceptance counts points for promotion in most universities, increases the likelihood of obtaining grants, and is favorably considered by the federal agency that ranks departments, which has consequences for how governmental

funds are distributed across and within universities. In addition, the meetings are a chance to see and be seen by other members of the profession, which contributes to prestige. Also, the meetings are often held in agreeable destinations.

The remainder of this paper is organized as follows. In the next section we briefly describe the National ANPEC meetings and the paper selection process. Section 3 then describes the concepts of information theory that we use to measure novelty and resonance. Section 4 reviews other papers that have used similar information-theoretic approaches to text analysis. Section 5 describes the data and programs used. In section 6 we describe the topic modeling of the ANPEC meeting annals and in section 7 the results from measuring the novelty, transience and resonance of the papers.

2. ANPEC and the National Economic Meetings

The first Economics departments in Brazilian universities date from 1946 (University of São Paulo and Federal University of Rio de Janeiro). Yet by the mid-1990s there was already the notion that maybe the proliferation of Economics departments had gone beyond the level of efficient resource use for Brazilian academia.³ The growth of Economics in Brazilian universities took off in the early 1970s together with the foundation of ANPEC, which sought to coordinate teaching and research at the graduate level. Possibly due to this higher degree of coordination and organization, Economics became a priority area for funding by the federal research councils (*CAPES* and *CNPq*), which resulted in a large number of Brazilian economists being trained and getting PhDs in American and European universities.

As the teaching of Economics grew in Brazil, ANPEC also expanded to accommodate the new centers that sought to join the association.

³ See discussion in Anuatti Neto (1997). On the early days of ANPEC and of the economics profession in Brazil, see Boianovsky (2021); Loureiro (1997); Bianchi (1997); Loureiro and Lima (1994).

Therefore, the yearly meetings also became larger.⁴ At some point a constraint for the size of the meetings was reached by the size of venue necessary to house the events. If the conference were to continue to grow, it would be too big for most conference halls or hotels, yet too small for the few really big venues. Once the conference reached this ceiling and the number of submissions continued to grow, the level of competition for acceptance at the conference increased significantly. Submissions come from young researchers to seasoned veterans in every area of economics. Increasingly papers and presentations have been delivered in English, despite enduring resistance in some quarters. Of the 1679 papers presented at ANPEC from 2013 to 2019, 1,101 were in Portuguese, 576 in English and 2 in Spanish. Slowly the conference is attracting the participation of economists from other countries.

This history of ANPEC is important for understanding the context in which the meetings occur and the incentives they provide for the creation of novelty versus conventionality. The high level of competition to get a paper accepted means that researchers must adopt strategies that involve trade-offs in terms of exploration versus exploitation, that is, whether to seek novelty or to stick with what is known and accepted. The nature of the meetings is also important to make clear what we mean by innovation in this context. Whereas it is a peripheral association of researchers in the global production of economic research - the highest ranked center in Brazil EPGE-FGV is ranked 287 in the Ideas-Repec global in Economics ranking (IDEAS, 2022) - it is nevertheless well integrated internationally and cannot be considered an isolated system. Innovation in such a context is not so much the creation of absolutely new ideas, but rather the early adoption and diffusion of ideas that flow from the center. This does not diminish the merit of innovating Brazilian economists as the natural resistance to change in academia still makes this type of innovation a risky career strategy.

⁴ The meetings are held in conjunction with the yearly meeting of the Brazilian Econometric Society - SBE. While ANPEC is more diverse and pluralistic in term of themes and approaches, SBE is more focused on quantitative approaches.

3. Information theory and measurement of novelty

We can understand the competition for scientific prestige through the lens of three variables: innovation, transience and resonance. Breaking previous patterns - high innovation - is a strategy that can only succeed if later research adopts this new trend - low transience. The difference between innovation and transience, which we call resonance, measures the ability of an article to impact scientific production.

In this paper we seek to measure the innovation, transience and resonance in the papers presented at the ANPEC meeting. With these indicators, it is possible to gain insights into the innovation process and in the frequency with which innovation becomes obscure. To quantify the statistical properties of a set of texts we use methods from Natural Language Processing by means of topic modeling framework.

Our application employs Latent Dirichlet Allocation (LDA) (BLEI; NG; JORDAN, 2003) as the topic modeling method. LDA is a generative statistical model that treats documents as bags of words generated by a mix of topics. When passing a corpus of texts as input, the algorithm estimates latent topics of the corpus and classifies each document according to the relevance of those topics.⁵ LDA topic modelling has already been used in many domains involving large quantity of documents (BOGDANOV; MOHR, 2013; BARRON et al., 2018; MURDOCK; ALLEN; DEDEO, 2017).

By applying the LDA algorithm on the papers in our database, we can describe each article as a probability distribution over the k topics. Our goal, then, is to find patterns and differences in the way co-authors choose their topic combination. We did this through Kullback-Leiber Divergence (KL), a measure of how one probability function diverges from another. To interpret the meaning of KL a brief explanation of the field to which it belongs, Information Theory, is necessary.

Introduced by Shannon (1948), Information Theory is a mathematical formulation designed to characterize the limits and possibilities of

⁵ See “Data and Programs” section for a detail of the method and requirements.

communication. The vast growth of the field, however, surpassed the analysis of communication and has made fundamental contributions to statistics, econometrics, physics, computer science and many other domains (COVER; THOMAS, 1991; MAASOUMI, 1993).

The most important concept in information theory is entropy. Entropy measures the amount of uncertainty or, equivalently, the randomness in a given context. The entropy, $H(X)$, of a random variable, X , is defined as:

$$H(X) = -\sum_{i=1}^n p(x_i) \log_2 p(x_i) \quad (1)$$

where the sum is over all values x that X can take, and $p(x)$ is the probability of the value x occurring. Measured in bits, $H(x)$ corresponds to the expected amount of information that the occurrence of an event in X produces, that is, the surprise of observing what happened in X . The more random the distribution, the more information resides in X .

Two points should be highlighted here. First, the concept of uncertainty used in Information Theory does not make a distinction, as in the economic literature, between uncertainty and risk (KEYNES, 1937; KNIGHT FRANK, 1921). Algorithm Information Theory, an advanced approach in the area, allows the measurement of the informative quantity for computable contexts whose probability distributions are not accessible, so that this distinction can, in theory, be overcome.⁶

Second, information is defined in an unusual way, as equivalent to uncertainty, which is different from its definition in Economics (GARROUSTE, 2001). Specifically, Information Theory is related to Informational Economics, but they are quite separate fields. Arrow (1984), for example, defined information as an economic commodity whose payoff and cost functions could be modeled based on Information Theory. However, while Informational Economics studies economic behavior under conditions of incomplete information, Information Theory is a theoretical formulation designed to understand the transmission, encoding and compression of information, a very distinct analysis.

⁶ See the Data and Programs section for a complete description of these formulations.

The concept of Kullback-Leiber Divergence (KL), on which our analysis relies, measures the gain in information when, given a probability distribution, $p(x)$, we use instead an alternative distribution $q(x)$:

$$KL(q) = \sum_{i=1}^n p(x_i) \log_2 \frac{p(x_i)}{q(x_i)} \quad (2)$$

Due to Jensen's Inequality, $KL(p|q) \geq 0$. Also, $KL(p|q) = 0$ when $p(x) = q(x)$ (MAASOUMI, 1993). Kullback-Leiber Divergence is asymmetrical in relation to the distributions since it does not obey the triangle inequality. KL is also known as relative entropy.

In econometrics, KL has many applications. We can think about the Kullback-Leiber Divergence as the information - uncertainty - gain when we approximate the true distribution of data, $p(x)$ by $q(x)$, say, the normal distribution. Trying to approximate by other distributions would then give a selection criterion in which the best approximate distribution is the one with the lowest KL value. It can be shown that General information criterion (GIC), Akaike's information criterion (AIC) and Bayesian information criterion (BIC) are all derived from this measure (EVREN; TUNA, 2012).

From the perspective of Bayesian reasoning, the information gain from KL corresponds to the surprise that an agent has when expecting $q(x)$ and realizing that $p(x)$ has occurred (BARTO; MIROLI; BALDASSARRE, 2013). The measure of innovation, transience and resonance are developed from this interpretation. Replacing $p(x)$ with $s^{(j)}$ and $q(x)$ with $s^{(j-1)}$, the probability distribution of, respectively, the j^{TH} and $j-1^{TH}$ papers in the data, we have:

$$KL(s^{(j)} | s^{(j-1)}) = \sum_{i=1}^k s_i^{(j)} \log_2 \frac{s_i^{(j)}}{s_i^{(j-1)}} \quad (3)$$

where we sum over all the k topics. Equation 3 represents the surprise of a text given the topics combination of its predecessor. Innovative papers generate more surprise as they break the pattern of their previous ones. We define novelty, $N_w(j)$, of the j^{TH} paper by averaging the KL of all the w previous texts:

$$N_{\omega}(j) = \frac{1}{\omega} \sum_{d=1}^{\omega} KL(s^j | s^{j-d}) \quad (4)$$

Thus, novelty represents the Bayesian surprise of an article when all previous papers are known. The higher the value, the greater the evidence that a given paper has introduced a new way of dealing with the research question or a novel research subject. Similarly, we define transience, $T_z(j)$ of the j th paper by averaging the KL of all the z later texts:

$$T_z(j) = \frac{1}{z} \sum_{d=1}^z KL(s^j | s^{j+d}) \quad (5)$$

Transience is also Bayesian surprise, but in relation to all future texts. T determines the degree in which the pattern of an article is overlooked by future research. High values of transience represent a loss of interest in the way that a certain article combined the research topics.

While coauthors can decide in advance the novelty that their research will have, much less control can be had over the transience of their work. Even so, due to the parsimonious nature of scientific production (KUHN, 1996), we can expect a strong relationship between novelty and transience, so that novelty is punished by the high probability of being forgotten. In order to find a paper's ability to break previous patterns and influence future research, we define resonance, $R(j)$ of the $(j)^{TH}$ paper as:

$$R(j) = N_{\omega}(j) - T_z(j) \quad (6)$$

Resonance, therefore, estimates the frequency with which innovation can overcome obscurity. With novelty, transience and resonance, we are able to quantify the contradictory forces involved in the scientific progress.

4. Literature review

In this section we briefly describe the previous literature which uses topic modeling and Kullback-Leibler divergence on different types of text to analyze issues related to cultural and scientific evolution. In this paper we rely directly on the ideas and techniques developed by Barron et al. (2018). These authors use reconstructed transcripts of debates that took place during the first National Constituent Assembly of the French Revolution (NCA), involving thousands of speakers and over 40,000 speeches, to track the creation, transmission and destruction of new ideas, many of which would set the patterns to be followed by subsequent democracies. They found a strong relationship between novelty and transience, so that very innovative speeches, on average, tended to be quickly forgotten. The variance in the results, however, exposes different strategies between political groups.

Left-wing political representatives had more innovative speech patterns, while conservatives were responsible for keeping the debate consistent with the patterns already established - low transience. In this sense, the authors noted that both political spectra were important for the development of ideas concerning modern states, each one with different roles.

Besides analyzing patterns of collective speech, the authors also examined the role of individual speakers. Some stood out for systematically breaking established patterns with a high resonance value. Robespierre, a famous Jacobin, produced speeches of high innovation and low transience, which means that his speeches had the effect of determining the subject of debate. Conservative representatives such as Jean-Sifrein Maury and Jacques de Cazalès had speeches of low innovation, but even less transience, so that these politicians were able to stabilize the debate on the same issues (BARRON et al., 2018).

The measures uncovered the impact of newly created organizational functions such as the National Constituent Assemblies' president and the work committees. For example, the NCA presidents marked the discussions due to the high transience of speeches. This is because

these individuals repeatedly summarized the debates held throughout the day, without necessarily participating in the formulation of the arguments.

A similar method was used to analyze the readings that influenced Charles Darwin to create his theory of evolution (MURDOCK et al., 2017). Due to the dense reading journals that he annotated throughout his life, it was possible to evaluate Darwin's responses to the trade-off between exploration and exploitation. Exploitation refers to the deepening of knowledge in the same area and exploration refers to the search for different knowledge.

Murdock et al. (2017) used LDA to create a probabilistic structure of topics on the books cited in Darwin's journals, and then applied the Kullback-Leiver divergence to them. Exploration, here, was proposed as the KL values above the average, both between two close texts and between one text and all the previous ones. Conversely, exploitation refers to the KL values below the average, in the same context. The main technical difference between this paper and the previous one is the fact that the analysis of Darwin's readings did not assess transience. The authors found that Darwin changed his strategy between exploitation and exploration throughout his career, first engaging in exploitation and later exploring new topics. Remarkably, they show that these changes were related to important events in his career.

A related study by Jing et al. (2019) applied similar techniques to fanfiction, which is a genre of fiction where fans take existing characters and stories, such as Harry Potter or Sherlock Holmes, and write new stories that are linked to the original but posit new situations and contexts, for example making Watson an alien or crossing over with characters from other stories. They use a corpus of more than 500 thousand pieces of fanfiction and measure success as the number of kudos (similar to 'likes' in most social media) that each piece receives from other readers. Using topic modeling and KL-divergence, they classify each piece by novelty and then analyze how novelty relates to performance. This genre of text is particularly appropriate for this test because fanfiction often invites transgressions where the original stories

and characters are wildly modified and refitted. Their results show a U-shaped relationship between novelty and popularity. The fans of fanfiction overwhelmingly prefer conventional approaches. As novelty increases the number of kudos decreases monotonically. But at the extreme level of novelty a few pieces manage to reach very high levels of popularity.⁷

Degaetano-Ortlieb and Teich (2018) use KL-divergence on the corpus of the Royal Society of London to analyze linguistic change over time, with periods of change followed by periods of consolidation. Chang and Dedeo (2020) provide some discussion of the different ways to conceptualize novelty in the quantitative analysis of text and argue in favor of divergence (KL) as preferable to other distance measures. Finally, there is a large literature that analyzes text, music, patents, and other forms of cultural expression with the focus on measuring novelty and uncovering the relationship between novelty and success or performance, but these use other methods, such as tracking patterns of citations or using complex networks (ASKIN; MAUSKAPF, 2017; FOSTER et al., 2015; MUELLER, 2021; UZZI et al., 2013; YOUN et al., 2015).

5. Data and programs

Our analysis was implemented in Natural Language Processing using initial code made available by Barron et al. (2018).⁸ Initially, the *urls* containing the downloaded hyperlinks, authors' names and area of the paper in ANPEC's meeting were scraped and the results were saved as a structured database. Of the 1679 articles published between 2013 and 2019, only one had issues for downloading and was not used in the application.

⁷ Thus the title "Sameness attracts, novelty disturbs, but outliers flourish in fanfiction online".

⁸ See their supplementary material and the example at Barron (2021). Our code is available at Correia (2021). The appendix material for this paper is available at this link.

The downloaded documents were converted to text files. Subsequently, a textual language detection algorithm was performed. Identifying the language of the text is essential since the conventional LDA libraries were produced for single language corpus. Among the remaining 1678 articles, two of them were written in Spanish, which is too few for topic modeling, so they were then removed from the database, leaving 1100 articles in Portuguese and 576 in English. The corpus was divided by language into two corpora and the procedures were performed separately in each set.

We then performed the lemmatization of the texts, which is a process to find the root of the words. This is essential if we want the LDA to recognize terms like 'municipality' and 'municipalities' as coming from the same canonical word.⁹ Lemmatizing is an optional procedure but was applied due to its capacity to enhance topic modeling predictions in our corpora.

The next stage of implementation consists of pre-processing the texts. Bibliographic references, badly coded characters, special characters, digits, e-mails, references to websites, punctuation (with the exception of accents) and words with less than three letters were removed. The remaining words have been converted to lowercase. We also created a stopwords list - words that don't add information for topic modeling so that they are ignored by the LDA algorithm, such as 'the', 'a', 'an' - of about 1900 terms and only used words that were present in more than 3 papers.

We did the LDA topic modeling with a computationally efficient method for optimizing the words in the topics, the Online Variational Bayes (OVb). Based on stochastic optimization, OVb converges more quickly to an equilibrium compared to other versions of Bayesian calculations (HOFFMAN; BACH; BLEI, 2010). Following the best practices for LDA implementation (ASUNCION et al., 2012; WALLACH; MIMNO; MCCALLUM, 2009), we applied an optimized asymmetric

⁹ We implemented lemmatization through the Stanza library, a powerful neural network for Natural Language Processing. Specifically for lemmatization, is one of the best available currently (see UNIVERSAL DEPENDENCIES, 2018).

Dirichlet prior over the document–topic distributions and a symmetric prior over the topic–word distributions.

To determine the number of topics, k , we compared the Perplexity and Topic Coherence Score, which are the conventional selection criteria used in LDA applications, for a wide variety of numbers. We find that 30 topics was a good fit to our data both for English and Portuguese corpora.

Finally, with the probability distribution over the 30 topics, we calculate novelty (Equation 4), transience (Equation 5) and resonance (Equation 6) in relation to papers of the same area and of the same corpus. We control by publication area so that the measures used do not suffer from noise effects, in the sense that papers from different research areas may have high KLDs between them due to the different topics studied and not by the presence of novelty or transience.

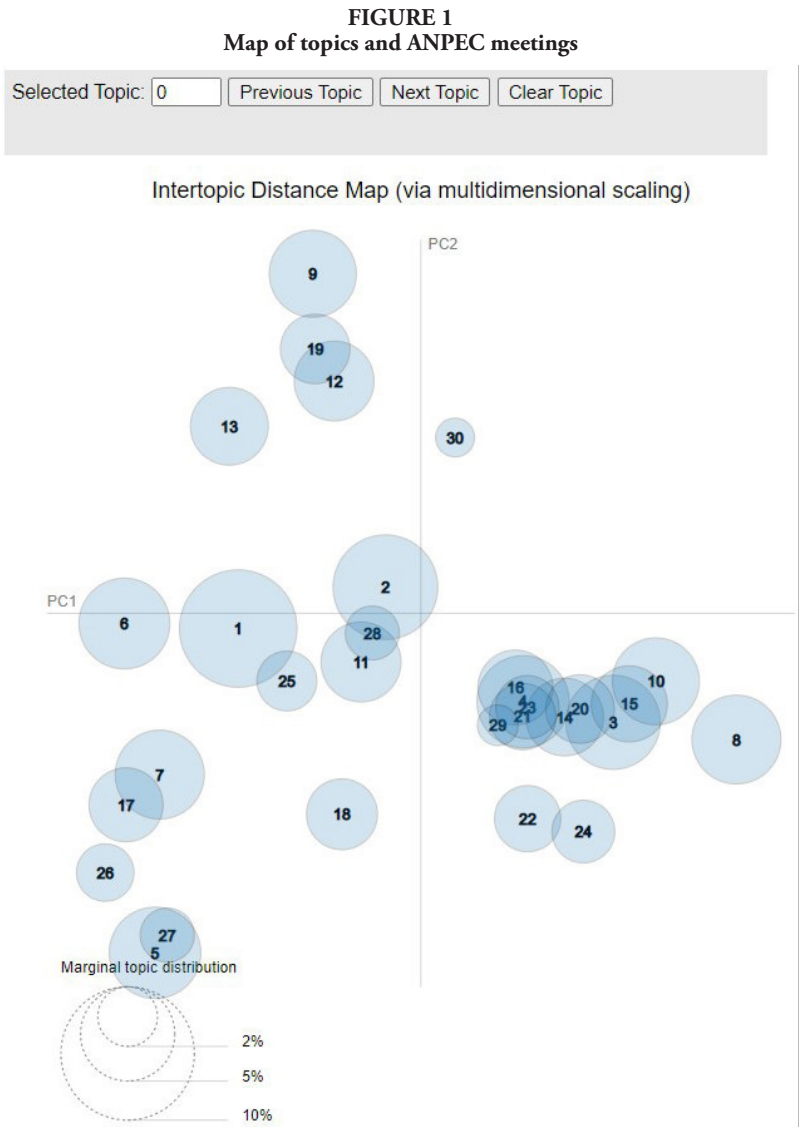
A limitation of the database used lies in the fact that it is not possible to know the order in which papers published in the same year were created. To overcome the limitation, we randomly sorted articles of the same year and area 100 times and took the median of all simulations as the final value. This solution is very similar to the one presented by Murdock et al. (2017) to deal with the same problem. After all the calculations, the corpora are again integrated into the same database.

6. Topic modeling of ANPEC Meeting annals

In this and the next section we present the analysis of the 1,679 papers accepted to the ANPEC meetings from 2013 to 2019. Here we describe the results from the first step of this process, which involves using latent Dirichlet allocation to extract from the entire corpus 30 topics which are subsequently used in the next section to characterize each paper as a weighted combination of these topics. We opted for having 30 topics, after some experimentation, to get enough but not too much granularity. The program then selected the

content of each topic with no further information from us, simply by identifying co-occurring patterns.

Figure 1 shows a mapping of the topics on two dimensions by using multidimensional scaling. The topics cluster in groups as one would expect in any academic discipline, where some topics are related in object of study and/or methods, and others are more distant.



The 15 most relevant words for 10 topics are listed in Table 1.¹⁰ The overall most salient terms in the full corpus are shown in Figure 2. The program receives as inputs only the texts and no information about the structure of the Economics literature or the National ANPEC meetings. Yet for any economist the nature of each topic is instantly recognizable. And anyone who has participated in an ANPEC meeting will easily surmise to which of the 13 ANPEC areas each topic corresponds. In the last row of each column, we indicated our guess of the ANPEC area for each topic.

Once we know the topics, sense can be made of the clusters in Figure 1. The most salient words for topic 1, for example, can be seen in Figure 3. The topic is clearly related to international economics. By examining the topic mix in each area of the map in Figure 1 we can locate each of the ANPEC areas. The cluster in the top-left quadrant includes economic history (topics 9 [politician, Furtado, history ...] and topic 19 [coffee, slave, *paulista* ...]), history of economic thought (topic 12 [science, economist, institution ...]) and political economy (topic 13 [Marx, capitalist, money ...]).¹¹ A cluster of international economics is located in the lower-left quadrant, including topic 17 [elasticity, export, import ...], topic 26 [exchange, exchange rate, volatility ...] and 27 [opening, crisis, flux ...]. These are close to some macroeconomic topics, topic 5 [shock, inflation, regime ...], topic 18 [fiscal, expenditure, corruption ...], and topic 7 [interest, credit, monetary, ...]. The crowded cluster in the lower-right quadrant is composed mostly of applied microeconomics topics, such as labor, urban & regional, social & demographic. Interestingly, the topic modeling procedure is so discerning that it distinguishes between the macroeconomic analysis of labor and employment (topic 25 [unemployment, salary, inflation, worker, pay cut, wage ...]) and the microeconomic analysis (topic 3 [worker, wage, woman, occupation, man, age, earnings, ...]).

¹⁰ We only show 15 words and 10 topics because of space considerations. The full output is available upon request. The original words in Portuguese are in Table A1 in the Appendix 1.

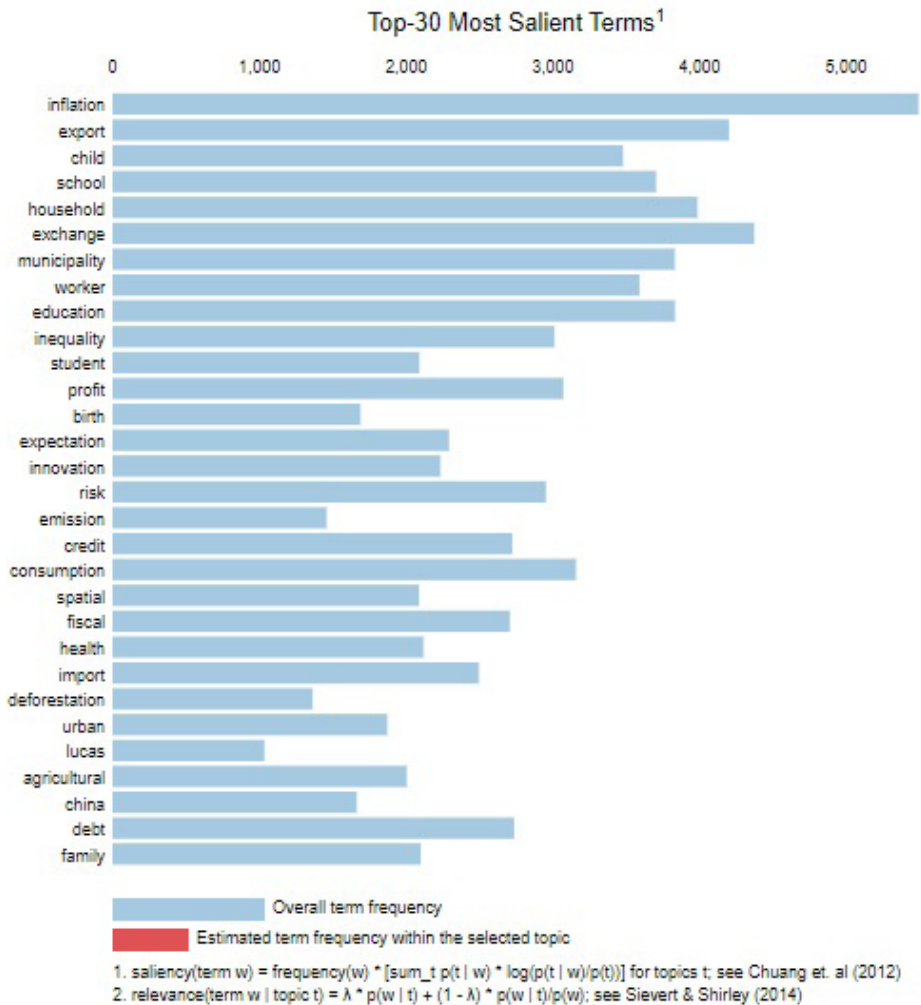
¹¹ In Brazil and especially in ANPEC, the term 'political economy' is often use to mean leftist economics.

TABLE 1
Topics in the ANPEC annals

Topic 1	Topic 2	Topic 3	Topic 4	Topic 5
Trade	Innovation	Worker	Space	Shock
Export	Technological	Salary	City	Inflation
Industry	Firm	Women	Regional	Regime
Commercial	Growth	Occupation	Municipality	Cycle
Import	Technology	Man	Urban	Forecasting
Industrial	Patents	Age	Distance	Expectation
China	Industrial	Income	Center	Monetary
Exporter	Interaction	Time	Industry	IPCA
Input	Industry	Working	Worker	Hiatus
Trade	Innovative	Home	Density	Curve
Household	Intensity	Wages	Mobility	Structural
Export	Network	Differential	Concentration	Matrix
Sectoral	Innovation	White	Agglomeration	Breakage
Matrix	Effort	Boss	Location	Phillips
Worldwide	Innovator	Benefit	Transport	Target
Area 7 International Economics	Area 9 Industrial and Technology	Area 13 Labor Economics	Area 10 Regional & Urban	Area 4 Macro., Monetary & Finance
Topic 6	Topic 11	Topic 12	Topic 13	Topic 14
Currency	Emissions	Science	Marx	Inequality
Active	Energy	Economist	Capitalist	Poverty
Crisis	Scenario	Institution	Money	Regional
Risk	Family	Knowledge	Merchandise	Poor
Monetary	Environmental	Concept	Profit	Northeast
Portfolio	Ethanol	Scientific	Capitalist	Region
Return	Simulation	Veblen	Accumulation	Decomposition
Liquidity	Energy	Human	Strength	Rural
Flow	Transport	World	Category	Education
Global	Shock	Vision	Wealth	Income
Action	Fuel	Thought	Class	Southeast
Interest	Balance	Hayek	Expansion	North
Credit	Oil	Critical	Trading	Family
Title	Climate	Practice	World	Urban
Exchange	Input	Action	Infrastructure	Gini
Area 4 Macro., Monetary & Finance	Area 11 Agricultural & Environmental	Area 1 History Thought Methodology	Area 2 Political Economy	Area 12 Social & Demo-graphic Economics

Source: Authors' elaboration. Papers in Portuguese and table translated by the authors. Original table in Appendix 1 Correia (2021). Data from ANPEC Meeting annals (ASSOCIAÇÃO NACIONAL DOS CENTROS DE PÓS-GRADUAÇÃO EM ECONOMIA, 2021a). The words 'trade' in topic 1 and 'innovation' in topic 2 are repeated because they occur both in English and in Portuguese, though the texts are in Portuguese.

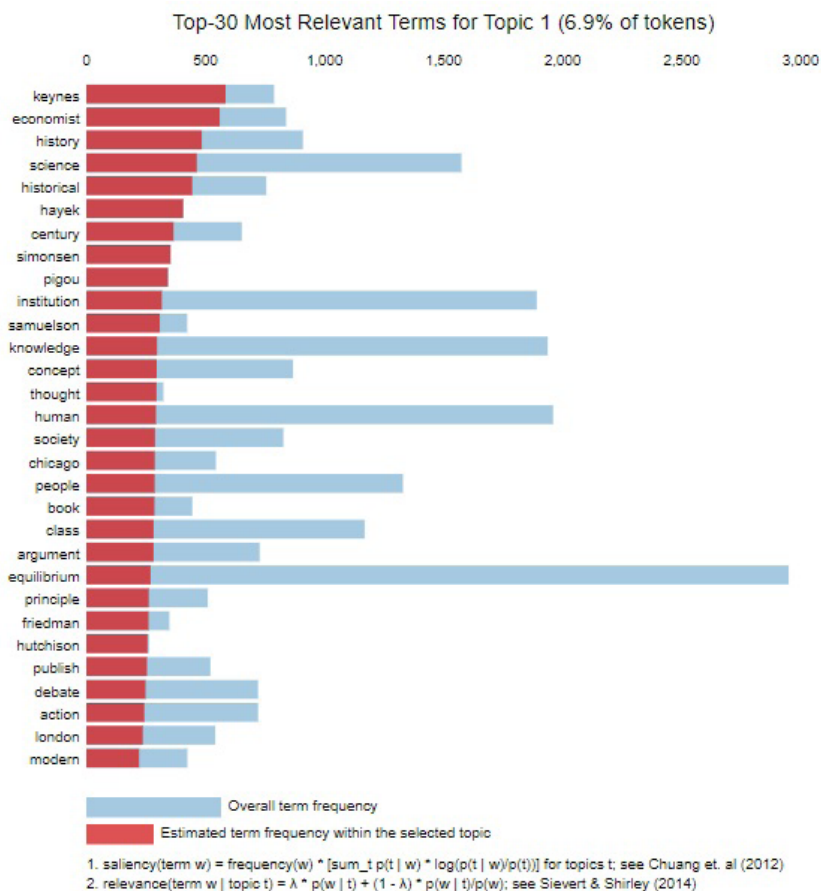
FIGURE 2
Most salient terms, full sample



Source: Created by the authors. Code and data in Correia (2021).

Several other examples of the uncanny precision of the topic modeling classification can be found by examining the map and topics. Based on these patterns we can speculate that the horizontal dimension is capturing the distinction between micro and macroeconomics, and the vertical dimension the greater or lesser use of mathematics in the texts.

FIGURE 3
Most salient terms, Topic 1



Source: Created by the authors. Code and data in Correia (2021)

7. Novelty, transience, and resonance

Once we have the list of topics which pervade Brazilian economic research, we can then decompose each paper into a set of patterns expressed as a probability distribution across topics. Novelty can then be measured as the divergence of a given paper, or set of papers, to the patterns established by previous papers. We use Kullback-Leibler Divergence (KLD) measures to do this, as described in section 3. Divergence is a measure of surprise, which can be interpreted as novelty. Given that you are used to a certain set of patterns from past

National ANPEC meetings, how surprised are you when you read a new paper from the latest meeting and find new patterns?

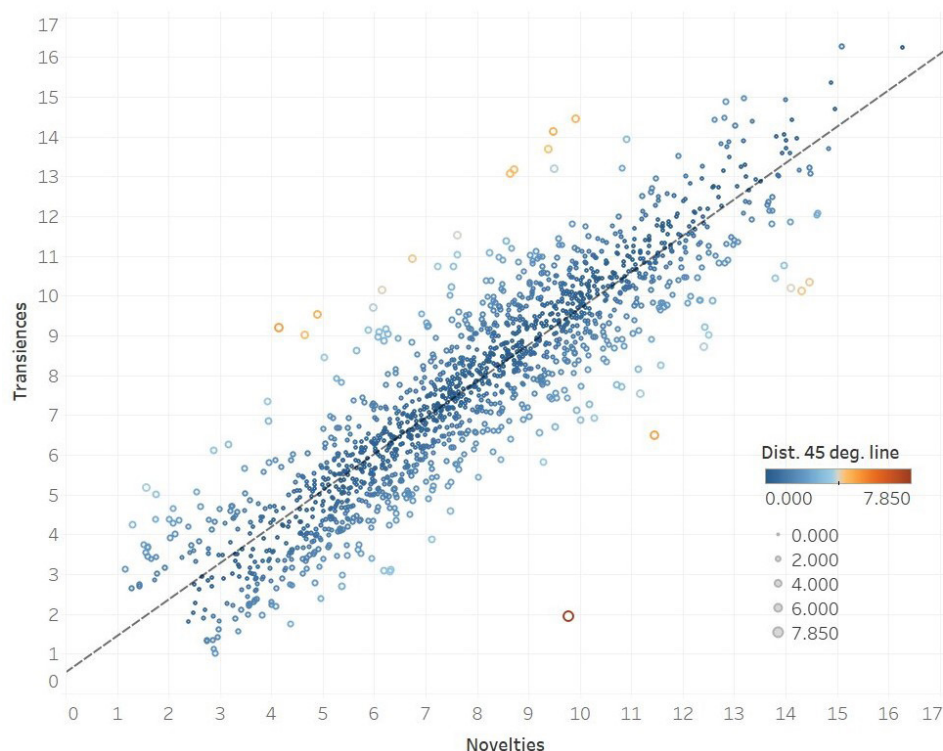
Note that this approach has some similarities with the standard approach of conceiving of novelty as the recombination of existing ideas (ASKIN; MAUSKAPF, 2017; MUELLER, 2021; UZZI et al. 2013; YOUN et al. 2015). But it is also different because it does not think in terms of atomistic recombination of lone ideas, but rather of whole distributions of ideas and their interactions, which gives it a Bayesian nature that provides a better representation of the pattern-making and pattern-breaking dynamic that is the evolution of science.

Besides measuring the surprise for the patterns in a paper given the patterns in past papers (novelty), we also measure the surprise compared to future papers, which we take as a measure of transience. If a new pattern appears at a given ANPEC meeting, but then does not appear in subsequent meetings, this means that the pattern did not catch on or diffuse. If, on the contrary, the new pattern subsists in subsequent meetings, then we can think of it as having, in a sense, changed the conversation. Following Barron et al. (2018), we call the difference between novelty and transience, resonance.

In Figure 4 we show the results for the set of ANPEC papers in Portuguese. The figure plots each paper by novelty on the horizontal and transience on the vertical axis. There is a tight fit along the 45-degree line. This indicates that papers that are high in novelty tend to also be high in transience. Most novel ideas tend to be ephemeral. This may be because many new ideas are simply uninteresting, but it can also be due to the natural conservatism of scientific inquiry. In many cases this is how it should be. Researchers should only embrace new ideas once these ideas prove themselves to be valid and valued, and it is not always obvious whether any given new idea is one or the other. But this resistance is often misplaced, and good ideas can be wrongly dismissed. There is no guarantee that all good ideas will eventually prevail.

Close examination of the graph reveals that although the fit is tight, there are several papers that are significantly below the 45-degree line.

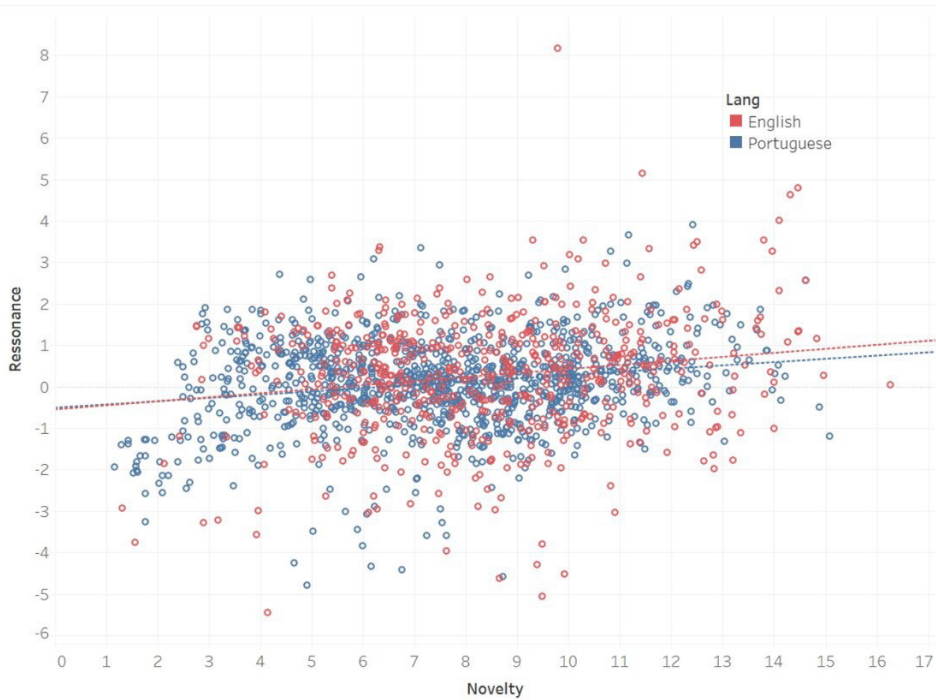
FIGURE 4
Novelty vs. Transience at ANPEC meetings



Source: Authors' elaboration using data from Associação Nacional dos Centros de Pós-Graduação em Economia (2021) and code from Barron et al. (2018). The size and color of the points changes as they become further from the 45-degree line.

This means that they have considerably less transience than would be expected given their level of novelty. Figure 5 explores these relations by plotting novelty against resonance. Although it may be difficult to visualize, there is a positive correlation of 0.19 between novelty and resonance (statistically significant at 1%). A regression line reveals a positive slope of 0.78 for the papers in Portuguese and a slope of 0.98 for English, also statistically significant at 1%. This result means that in the National ANPEC meetings the papers that contribute most to changing the conversation in the economics profession, tend to be papers that contain novel patterns. Though positive and statistically significant, the correlation is not that high, so other characteristics besides novelty

FIGURE 5
Novelty vs. Resonance at ANPEC meetings



Source: Authors' elaboration using data from Associação Nacional dos Centros de Pós-Graduação em Economia (2021) and code from Barron et al. (2018). The slope for English is 0.98 and for Portuguese 0.78, both statistically significant at 1%.

are probably also honored by the profession. Nevertheless, novelty is also valued.

These results provide some assistance to Brazilian economists contemplating the trade-off between exploration and exploitation. Exploration (seeking new areas) is a risky endeavor that often ends badly. The guardians of the profession, including editors, peer reviewers, grant proposal evaluators, dissertation committees and others, value traditional concepts, well-established topics and familiar methods and do not easily engage with the new. On the other hand, there is a cost to exploitation (sticking to what already works), as having a greater impact seems to require at least some novelty.¹²

¹² An important concern regarding our results is whether the early years (in the case of novelty) and the latter years (in the case of transience) could be in some way biased

The results above were all derived passing the corpus of ANPEC papers through topic modeling and KL divergence procedures. One might wonder if our measurements of novelty, transience and resonance actually capture what we claim they measure. To validate our interpretations of the results we compare these measurements with the number of citations each of these papers has received since the meeting. This information is available in Google Scholar, which also identifies when the paper has been subsequently published in a journal or whether it has been published only in the ANPEC meetings annals.¹³ Because citations are an external measurement of the papers' content, it would not be tainted by any bias or misinterpretation in our procedures, so it is a good counterpoint to compare our results and evaluate our claims.

We do not expect resonance and citations to be the same thing. Resonance is an information theoretic measure based solely on the papers' content. Citations are subject to subsequent choices of the authors after writing the paper (submission, presentation, networking, etc.). They are also subject to the sociology of science that involves the reaction of the network to the paper, based on many other criteria besides the actual content of the paper (friendships, rivalries, the vagaries of peer review, institutional policies, etc.). Nevertheless, if our interpretations of novelty, transience and resonance are to make sense, we expect that there should be some relation between them and citations.

One way to make this comparison is to divide the full set of papers in two groups according to the number of citations and check if those with more than the median number of citations have a different level of novelty, transience and resonance than those below the median. Doing this, however, does not show any difference in these measures

because they are calculated based on a smaller set of preceding/following texts. To investigate whether this is the case, we replicate Figures 4 and 5 in the appendix without the data for 2013 and 2019 (see Figures A4, A5 and A6). We also present the descriptive statistics for the main variables by year in Table A2. There does not seem to be any systematic difference across years. Nevertheless, in future work it would be useful to extend the dataset back in time to use an even larger sample of papers.

¹³ We retrieved the number of citations for each paper from Google Scholar in December 2020.

(see Figure A1 in the Appendix 1 for a histogram for resonance).¹⁴ The problem with doing the comparison in this way is that the year of the meeting makes a big difference in terms of citations, as papers presented closer to 2013 have had more time to garner citations than those presented closer to 2019. Therefore, in Table 2, we regress the number of citations against each of our three measures (separately), while controlling for each paper's meeting year. We use negative binomial regression because the dependent variable, citations, is an over-dispersed count variable.¹⁵ In addition, we control for whether the paper was eventually published in a Brazilian or a foreign journal (base category: published only in the ANPEC annals), whether the paper was written in English or Portuguese, and the number of co-authors. We also add ANPEC area dummies and dummies for all departments that had more than 10 papers presented in the full 2013-2019 period.

The results show our measures of novelty, transience and resonance are associated with citations in accordance with our interpretations above. Column I shows that novelty is negatively but not statistically related to citations. As expected, novelty tends to be quickly forgotten. Yet transience, in column II, is significantly negatively associated with citations. Papers that do not endure in our information theoretic measure also do not make a subsequent impact through citations. Most importantly, resonance in column III is positively associated with citations. Resonant papers are those that introduce novel patterns and these patterns do not immediately fade but rather have some endurance in subsequent meetings (all else constant). The coefficient magnitude implies that a one unit increase in resonance is associated with 1.06 more citations ($e^{0.058}$).

The control variables provide interesting information on which kinds of papers in the ANPEC meetings receive most citations. As expected, the year dummies are more negative the closer to 2019. It takes time for papers to be known and to be cited. Papers published

¹⁴ The appendix is available at Correia (2021).

¹⁵ We limit the sample to observations with at least one citation, thus focusing on those papers that had at least some impact by this measure. OLS results, full sample results, and results without using 2013 and 2019 are shown in Table A3 in the Appendix 1.

TABLE 2
Citations and papers' characteristics

Dep. Var.	I Citations	II Citations	III Citations
Novelty	-0.011 (-0.71)		
Transience		-0.030* (1.94)	
Resonance			0.058* (1.87)
Brazilian journal	0.894*** (11.31)	0.893*** (11.31)	0.901*** (11.41)
Foreign journal	1.394*** (10.22)	1.372*** (10.16)	1.372*** (10.42)
English	0.019 (0.21)	0.029 (0.32)	0.002 (0.02)
Number authors	0.095** (2.43)	0.098** (2.52)	0.099** (2.56)
2014	-0.181* (-1.68)	-0.185* (-1.74)	-0.203 (-1.91)
2015	-0.381*** (-3.57)	-0.387*** (-3.62)	-0.406*** (-3.76)
2016	-0.688*** (-5.82)	-0.683*** (-5.79)	-0.703*** (-5.92)
2017	-0.867*** (-6.79)	-0.867*** (-6.86)	-0.903*** (-6.98)
2018	-1.099*** (-7.36)	-1.104*** (-7.38)	-1.129*** (-7.58)
2019	-1.757*** (-8.09)	-1.763*** (-8.28)	-1.772*** (-8.29)
Constant	0.722*** (3.38)	0.838*** (3.88)	0.697*** (3.37)
Observations Pseudo R-squared	733 0.12	733 0.12	733 0.12
Wald chi ² (55), p-value	536.1, 0.000	548.4, 0.000	543.1, 0.000

Source: Negative binomial regression. Sample limited to papers with at least one citation. Robust errors. *10%, **5%, ***1%. Base year is 2013. Foreign and Brazilian publications compared to papers only published in the ANPEC annals. ANPEC area effects and department effects were included in the estimation but are not shown in the table.

in Brazilian journals have on average 2.46 more citations than those that never progress further than the meetings' annals, and for papers published in foreign journals that number is 3.94. Papers published in English are not more cited than those in Portuguese. In addition, papers with more coauthors are more highly cited. This is a trend that has been observed more generally in science (WUCHTY; JONES; UZZI, 2017). We cannot, however, distinguish if this is because the content is better or if more coauthors are better able to promote the paper, through more presentations and networking.

The results in Table 2 are evidence that our measures of novelty, transience and resonance capture characteristics of the papers that, at least in part, explain their subsequent performance in terms of citations. These characteristics are by no means the sole or even the main

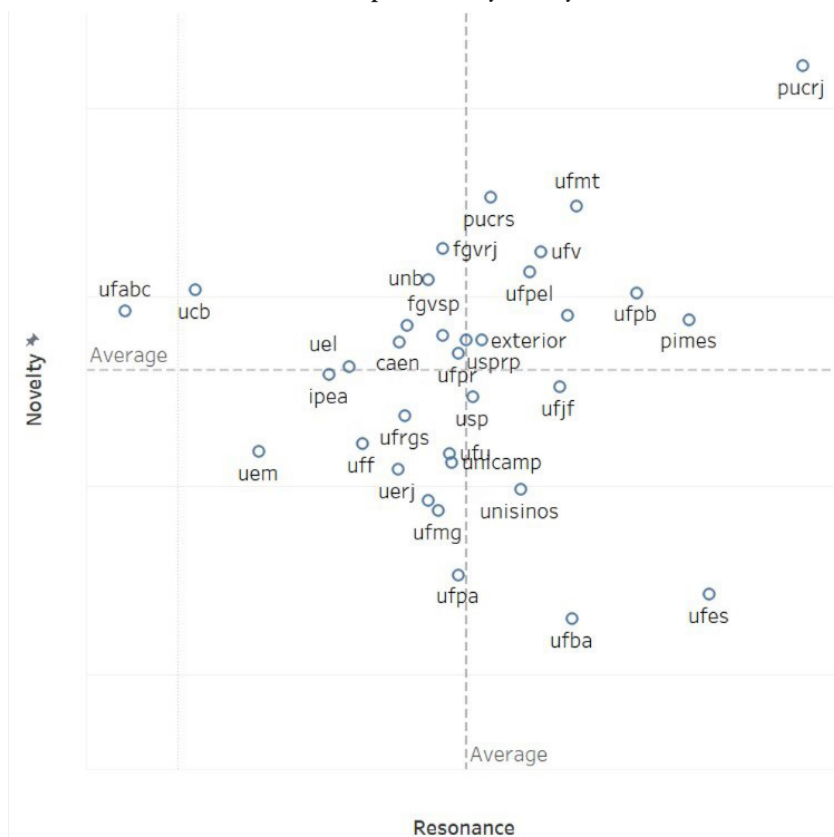
determinant of greater or lesser success. As shown by Salganik, Dodds and Watts (2006), in cultural and scientific markets “hit songs, books, and movies are many times more successful than average, suggesting that ‘the best’ alternatives are qualitatively different from ‘the rest’; yet experts routinely fail to predict which products will succeed.” Similarly, Barabási (2018) surveys research in a wide variety of fields, including art, sports, wine tasting, universities, academic productivity, Nobel prizes, Kickstarter campaigns, and others, and shows that “success, as it turns out, is not a direct result of our achievements, but instead an indirect reaction to how those achievements are perceived and valued by those around us.” It is not that performance is irrelevant for success, but that where performance is difficult to measure, it is networks that drive success. In academic markets the networks include the whole hierarchies of universities, departments, societies, research groups, journals, WhatsApp groups, etc.

Before closing this section, we use our data to explore the variation of novelty and resonance across departments and across ANPEC areas. The objective is to see if some departments or some areas are more prone to introduce novelty and/or to have greater resonance than others. We can imagine classifying departments and areas in a novelty-resonance space that can be divided into quadrants. Research that has higher than average novelty and higher than average resonance would fall in the top right quadrant. This is work that introduces new patterns in the meetings and these patterns persist. In the high-novelty, low-resonance quadrant new patterns are introduced but they fail to change the conversation. In the low-novelty, high resonance quadrant the research is not novel, that is, it uses the same patterns as before, but those are solid patterns that are maintained in the future. And in the low novelty, low resonance quadrant, few new patterns are introduced and those do not tend to persist.

Figure 6 shows the plots in novelty-resonance space by departments and Figure 7 for ANPEC areas. It is important to consider several caveats when analyzing the results. The first is that the comparison is limited to the context of the National ANPEC meetings. It says

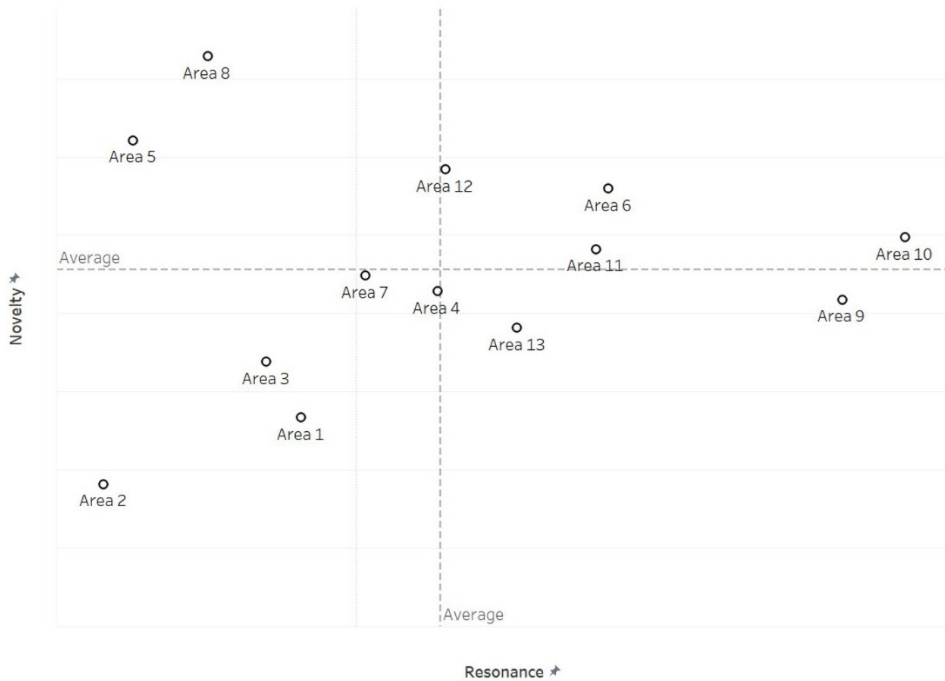
nothing about how those papers compare to papers in other meetings, journals or to papers in general. The results should not be understood as an evaluation of the work done in any department, only of the work that each department presented at National ANPEC. The number of researchers in a given department that consider ANPEC as an outlet for their work varies greatly. And those that do, may not be representative of the whole. Some departments are fully engaged with National ANPEC meetings and others might focus instead in other conferences, such as the SBE (*Sociedade Brasileira de Econometria*) meetings that are held in conjunction with ANPEC. It is possible that some researcher might choose to send their best work to foreign

FIGURE 6
Economics departments by novelty and resonance



Notes: Departments set according to the first author. Only departments with more than 10 papers at the meetings from 2013 to 2019 were included.

FIGURE 7
ANPEC areas by novelty and resonance



Notes: Area 1 - History of economic thought; Area 2 - Political economy; Area 3 - Economic history; Area 4 - Macro, monetary, finance; Area 5 - Public sector econ.; Area 6 - Growth, develop., institutions; Area 7 - International econ.; Area 8 - Micro. Quant. methods, finance; Area 9 - Industrial econ. & technology; Area 10 - Regional and urban; Area 11 - Agricultural & Environmental; Area 12 - Social & demographic; Area 13 - Labor economics.

meetings instead. Furthermore, it may not even make much sense to think of a department as a unit, since in Economics at least, research is often done individually or with colleagues from other institutions. Another important caveat is that for these plots, we classify each paper as belonging to the department of the first author. Many papers, however, have multiple authors and authors from different universities. Also, we considered only departments that had more than 10 observations in our sample, so many departments are excluded.

At first glance Figures 6 and 7 seem to suggest interesting patterns in the average novelty and resonance of research done across Brazilian economics departments and across ANPEC areas. But if we take into consideration the variation around the averages it turns out that the difference is not that large. Figures A2 and A3 in the Appendix

1 show the values one-standard deviation above and below the mean. In practically all cases the department and area averages are within this interval for other departments and areas. This is not due to a lack of variation across papers. Figures 4 and 5 show much heterogeneity at the level of individual papers.¹⁶ What these results show is that there are no systematic differences across departments or areas. In the same department or area there is high and low novelty/transience. This is probably due to the nature of research in economics, which is, more so than in many other areas, an individual pursuit or one in which the collaborations are done across universities.¹⁷ Brazilian federal research agencies (CNPq and CAPES) often try to arrange their programs and grants around the concept of research groups.¹⁸ Our results suggest that research in economics is not structured in this way. Therefore, it does not make much sense to say that a given department or area has a certain characteristic, at least in terms of novelty and resonance.

8. Conclusions

Most studies that seek to identify and evaluate the impact of new ideas, use citations or some other measure external to the actual papers' content. In this paper we used a technique based on Barron et al. (2018) that uses the actual corpus of text in which the ideas were formulated. We used topic modeling and Kullback-Leibler divergence to create measures of novelty, transience, and resonance for all the paper accepted to the ANPEC meetings of the Brazilian Association for Graduate Economics from 2013 to 2019. Our results confirm the "law" that most novel ideas are quickly forgotten. We showed, however, that

¹⁶ In the Appendix 1 we list the 15 papers with highest resonance values (Table A5) and highest novelty values (Table A4).

¹⁷ For example, Meadows (1998) found that while 83% of papers in Economics were sole-authored, the numbers for other disciplines were Biochemistry 19%, Psychology 45%, Sociology 75% and Social Work 75%.

¹⁸ The CNPq (National Council for Scientific Development) classifies research done in the country through a Directory of Research Groups (BRASIL, 2021).

those ideas that have greater impact also tend to have higher levels of novelty. We also showed that there is a positive correlation between our measure of resonance and the papers' subsequent citation record, so that at least in part they seem to be measuring related aspects of impact.

In a sense, our measure can be thought of as a different form of citation. When a paper introduces a new pattern, subsequent research can cite that paper in the conventional way, which is picked up in citation statistics such as those in Google Scholar. However, novel ideas can also be "cited" when the new patterns are repeated in subsequent literature, sometimes even unconsciously and without standard citation procedures. One might argue that this is actually a more sincere form of citation as it shapes the new paper more profoundly than a conventional citation, which is often perfunctory.

While our tests are based on a very specific sample of papers – those presented at the National ANPEC meetings 2013 to 2019 – we believe that the results would hold for most other corpora of paper in Economics, given that what is produced in this area in Brazil is broadly similar to economic literature produced elsewhere. Although we believe that the same results would hold for other disciplines in the social sciences and the humanities, as well as more distant fields, such as, the biological and physical sciences, any such claim can only be made after the research has been extended to these areas. Also, expanding the coverage of the sample to include years before 2013, as well as to include other economic research in Brazil, such as the regional ANPEC Meetings and other economic societies, would improve our knowledge of the determinants of impact in Brazilian economic research.

References

ANUATTI NETO, F. Competição e complementaridade dos centros de pós-graduação em economia. In: LOUREIRO, M. R. (Ed.). 50 anos de ciência econômica no Brasil. Petrópolis: Vozes, 1997.

- ARROW, K. J. The economics of information. Cambridge: Harvard University Press, 1984. v. 4.
- ASKIN, N.; MAUSKAPF, M. What makes popular culture popular? Product features and optimal differentiation in music. *American Sociological Review*, Menasha, v. 82, n. 5, p. 910-944, 2017.
- ASSOCIAÇÃO NACIONAL DOS CENTROS DE PÓS-GRADUAÇÃO EM ECONOMIA – ANPEC. Brasília, 2021a. Available from: <<http://www.anpec.org.br/novosite/br>>. Access in: 28 Oct 2021.
- ASSOCIAÇÃO NACIONAL DOS CENTROS DE PÓS-GRADUAÇÃO EM ECONOMIA – ANPEC. 49º Encontro Nacional de Economia. Brasília, 2021b. Available from: <<https://en.anpec.org.br/previous-editions.php>>. Access in: 28 Oct 2021.
- ASUNCION, A. et al. On smoothing and inference for topic models. In: UAI '09: CONFERENCE ON UNCERTAINTY IN ARTIFICIAL INTELLIGENCE, 25., Montreal, Canada. Proceedings... Arlington: AUAI Press, 2012. p. 27-34.
- BARABÁSI, A.-L. The formula: the five laws behind why people succeed. London: Pan Macmillan, 2018.
- BARRON, A. T. et al. Individuals, institutions, and innovation in the debates of the French revolution. *Proceedings of the National Academy of Sciences of the United States of America*, Washington, v. 115, n. 18, p. 4607-4612, 2018.
- BARRON, A. T. NTRExample_FRevNCA. [S.l.]: GitHub, Inc., 2021. Available from: <https://github.com/CogentMentat/NTRExample_FRevNCA>. Access in: 28 Oct 2021.
- BARTO, A.; MIROLI, M.; BALDASSARRE, G. Novelty or surprise? *Frontiers in Psychology*, Pully, v. 4, p. 907, 2013.
- BIANCHI, A. M. Do encontro de Itaipava ao encontro da USP: comentários à margem da história da ANPEC. In: LOUREIRO, M. R. (Ed.). 50 anos de ciência econômica no Brasil. Petrópolis: Vozes, 1997.

- BLEI, D. M.; NG, A. Y.; JORDAN, M. I. Latent Dirichlet allocation. *Journal of Machine Learning Research*, Brookline, v. 3, p. 993-1022, 2003.
- BOGDANOV, P.; MOHR, J. W. Topic models: what they are and why they matter. *Poetics*, The Hague, v. 31, p. 545-569, 2013.
- BOIANOVSKY, M. Economists, scientific communities, and pandemics: an exploratory study of Brazil (1918-2020). *Economia*, Brasília, v. 2, n. 1, p. 1-18, 2021.
- BOUDREAU, K. J. et al. Looking across and looking beyond the knowledge frontier: intellectual distance, novelty, and resource allocation in science. *Management Science*, Providence, v. 62, n. 10, p. 2765-2783, 2016.
- BOURDIEU, P. *Science of science and reflexivity*. United Kingdom: Polity Press, 2004.
- BRASIL. Conselho Nacional de Desenvolvimento Científico e Tecnológico – CNPQ. Plataforma Lattes. Diretório dos Grupos de Pesquisa no Brasil. Brasília, 2021. Available from: <<http://www.lattes.cnpq.br/web/dgp/objetivos/>>. Access in: 28 Oct 2021.
- CASTILLA, J. P. To kill a black swan: the credibility revolution at CEDE, 2000-2018. [S.l.: s.n.], 2020. Available from: <<http://hdl.handle.net/1992/45870>>. Access in: 28 Oct 2021.
- CHANG, K. K.; DEDEO, S. Divergence and the complexity of difference in text and culture. *Journal of Cultural Analytics*, Montréal, v. 4, p. 1-36, 2020.
- CHECCHI, D.; DE FRAJA, G.; VERZILLO, S. Incentives and careers in academia: theory and empirical analysis. *The Review of Economics and Statistics*, Boston, v. 103, n. 4, p. 786-802, 2021.
- CORREIA, M. P. R. Innovation on Brazilian economic research. [S.l.]: GitHub, Inc., 2021. Available from: <<https://github.com/correia-marcos/Innovation-on-brazilian-economic-research>>. Access in: 28 Oct 2021.

COVER, T. M.; THOMAS, J. A. Elements of information theory. USA: Wiley-Interscience, 1991.

DEGAETANO-ORTLIEB, S.; TEICH, E. Using relative entropy for detection and analysis of periods of diachronic linguistic change. In: JOINT SIGHUM WORKSHOP ON COMPUTATIONAL LINGUISTICS FOR CULTURAL HERITAGE, SOCIAL SCIENCES, HUMANITIES AND LITERATURE, 2., Santa Fe, New Mexico. Proceedings... Stroudsburg: Association for Computational Linguistics, 2018. p. 22-33.

ECONOMIA. Brasília: ANPEC, 2021. Available from: <<https://www.sciencedirect.com/journal/economia>>. Access in: 28 Oct 2021.

EVREN, A.; TUNA, E. On some properties of goodness of fit measures based on statistical entropy. International Journal of Research and Reviews in Applied Sciences, Islamabad, v. 13, p. 192-205, 2012.

FARIA, J. R. Is there a trade-off between domestic and international publications? Journal of Socio-Economics, Amsterdam, v. 34, n. 2, p. 269-280, 2005.

FARIA, J. R. Some reflections on incentives for publication: the case of the CAPES list of economic journals. Economia Aplicada, São Paulo, v. 8, n. 4, p. 791-816, 2004.

FARIA, J. R.; ARAUJO JUNIOR, A. F. D.; SHIKIDA, C. D. The international research of academic economists in Brazil: 1999-2006. Economia Aplicada, São Paulo, v. 11, n. 3, p. 387-406, 2007.

FOSTER, J. G.; RZHETSKY, A.; EVANS, J. A. Tradition and innovation in scientists' research strategies. American Sociological Review, Menasha, v. 80, n. 5, p. 875-908, 2015.

GARROUSTE, P. What economics borrows from the statistical theory of information? Boston: Springer, 2001.

GUIMARÃES, B. Qualis as a measuring stick for research output in economics. Brazilian Review of Econometrics, Rio de Janeiro, v. 31, n. 1, p. 3-18, 2011.

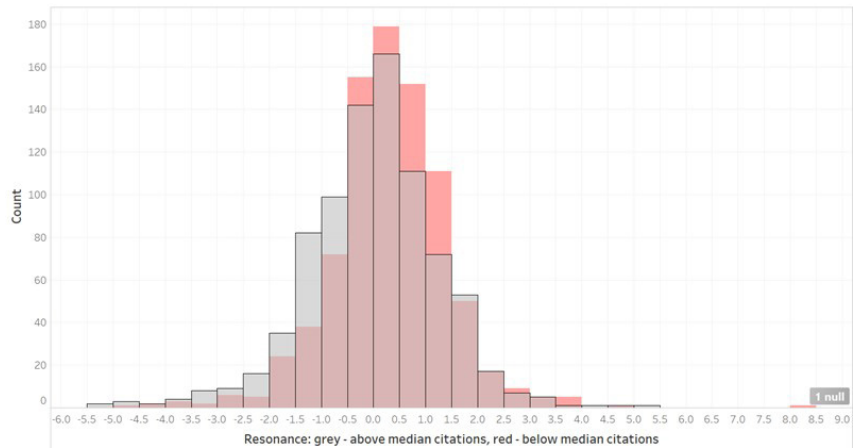
- HADDAD, E. A.; MENA-CHALCO, J. P.; SIDONE, O. Produção científica e redes de colaboração dos docentes vinculados aos programas de pós-graduação em economia no Brasil. *Estudos Econômicos*, São Paulo, v. 47, n. 4, p. 617-679, 2017.
- HOFFMAN, M.; BACH, F. R.; BLEI, D. M. Online learning for latent Dirichlet allocation. In: LAFFERTY, J. et al. (Eds.). *Advances in neural information processing systems*. La Jolla: Neural Information Processing Systems, 2010. p. 856-864.
- IDEAS. Top 10% Economic Institutions, as of June 2022. St. Louis: Federal Reserve Bank of St. Louis, 2022. Available from: <<https://ideas.repec.org/top/top.inst.all.html>>. Access in: 2 Aug 2022.
- ISSLER, J. V.; FERREIRA, R. C. Avaliando pesquisadores e departamentos de economia no Brasil a partir de citações internacionais. *Pesquisa e Planejamento Econômico*, Brasília, v. 34, n. 3, p. 491-538, 2004.
- JING, E., DEDEO, S.; AHN, Y.-Y. Sameness attracts, novelty disturbs, but outliers flourish in fanfiction online. arXiv, Ithaca, p. 1904.07741, 2019. In press.
- KEYNES, J. M. *The general theory of employment, interest, and money*. Switzerland: Springer, 1937.
- KNIGHT FRANK, H. *Risk, uncertainty and profit*. Boston: Houghton Mifflin, 1921.
- KUHN, T. S. *The structure of scientific revolutions*. 3rd ed. Chicago: University of Chicago Press, 1996.
- LOUREIRO, M. R. *Anos de ciência econômica no Brasil (1946-1996): pensamento, instituições, depoimentos*. Petrópolis: Vozes, 1997.
- LOUREIRO, M. R.; LIMA, G. T. A internacionalização da ciência econômica no Brasil. *Brazilian Journal of Political Economy*, São Paulo, v. 14, n. 3, p. 31-50, 1994.
- MAASOUMI, E. A compendium to information theory in economics and econometrics. *Econometric Reviews*, New York, v. 12, n. 2, p. 137-181, 1993.

- MEADOWS, A. J. Communicating research. San Diego: Academic Press, 1998.
- MERTON, R. K. The Matthew effect in science: the reward and communication systems of science are considered. *Science*, Washington, v. 159, n. 3810, p. 56-63, 1968.
- MUELLER, B. Where'd you get that idea? determinants of creativity and impact in popular music. *EconomiA*, Brasília, v. 22, n. 1, p. 38-52, 2021.
- MURDOCK, J.; ALLEN, C.; DEDEO, S. Exploration and exploitation of Victorian science in Darwin's reading notebooks. *Cognition*, Amsterdam, v. 159, p. 117-126, 2017.
- NOVAES, W. A pesquisa em economia no Brasil: uma avaliação empírica dos conflitos entre quantidade e qualidade. *Revista Brasileira de Economia*, Rio de Janeiro, v. 62, n. 4, p. 467-495, 2008.
- PETTERINI, F. C. Brazilian academic economics: a picture from the ANPEC exam microdata. *EconomiA*, Brasília, v. 21, n. 3, p. 325-339, 2020.
- SALGANIK, M. J.; DODDS, P. S.; WATTS, D. J. Experimental study of inequality and unpredictability in an artificial cultural market. *Science*, Washington, v. 311, n. 5762, p. 854-856, 2006.
- SHANNON, C. E. A mathematical theory of communication. *The Bell System Technical Journal*, New York, v. 27, n. 3, p. 379-423, 1948.
- UNIVERSAL DEPENDENCIES. CoNLL 2018 Shared Task. [S.l.: s.n.], 2018. Available from: <<https://universaldependencies.org/conll18/results-lemmas.html>>. Access in: 28 Oct 2021.
- UZZI, B. et al. Atypical combinations and scientific impact. *Science*, Washington, v. 342, n. 6157, p. 468-472, 2013.
- WALLACH, H. M.; MIMNO, D. M.; MCCALLUM, A. Rethinking LDA: why priors matter. In: BENGIO, Y. et al. (Eds.). *Advances in neural information processing systems*. La Jolla: Neural Information Processing Systems, 2009.

- WANG, J.; VEUGELERS, R.; STEPHAN, P. Bias against novelty in science: a cautionary tale for users of bibliometric indicators. *Research Policy*, Amsterdam, v. 46, n. 8, p. 1416-1436, 2017.
- WUCHTY, S.; JONES, B. F.; UZZI, B. The increasing dominance of teams in production of knowledge. *Science*, Washington, v. 316, n. 5827, p. 1036-1039, 2017.
- YOUN, H. et al. Invention as a combinatorial process: evidence from US patents. *Journal of the Royal Society, Interface*, London, v. 12, n. 106, 20150272, 2015.

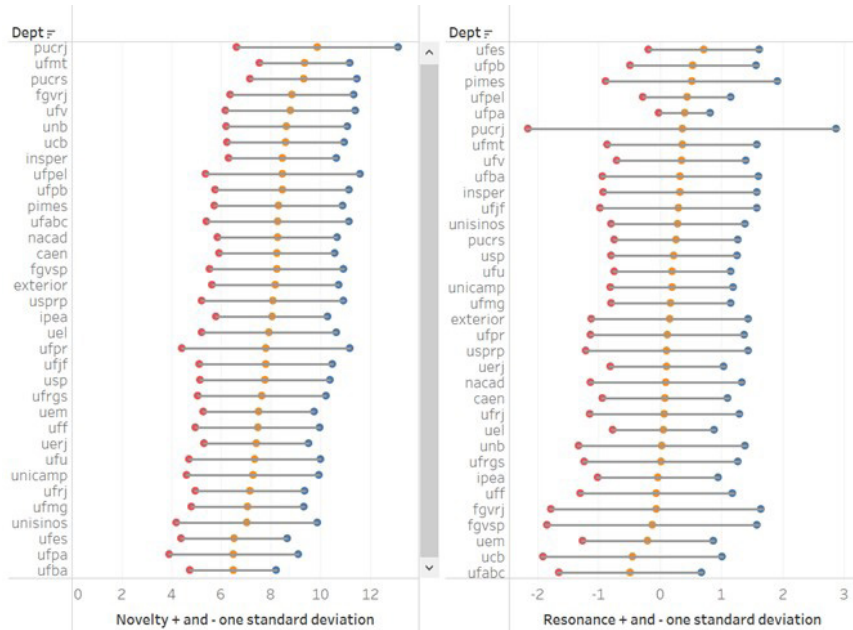
APPENDIX 1
Supplementary material

FIGURE A1
Histogram of resonance above and below median citations



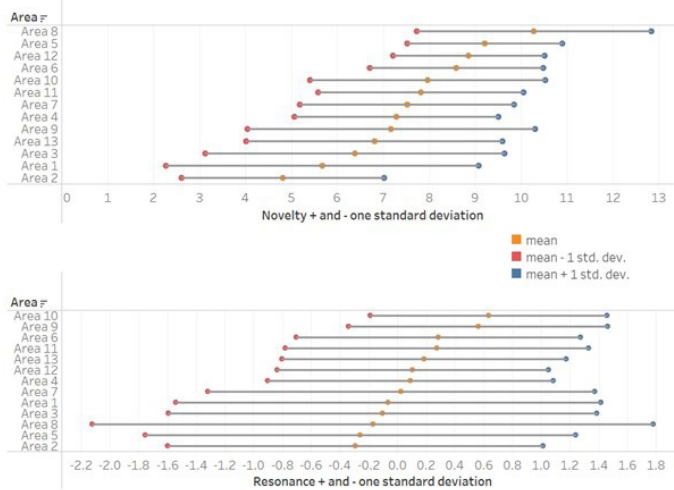
Source: Figure created by the authors.

FIGURE A2
Novelty and resonance variation by departments



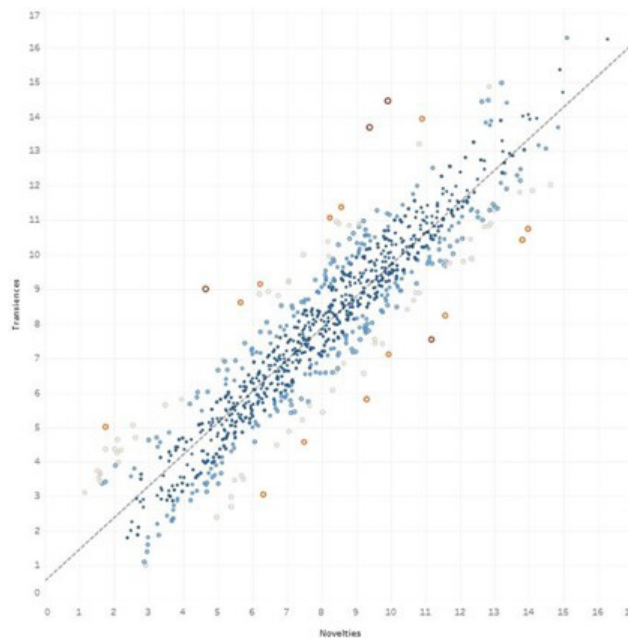
Source: Created by the authors.

FIGURE A3
Novelty and resonance variation by ANPEC area



Source: Created by the authors.

FIGURE A4
Novelty vs. Transience at National ANPEC meetings, without 2013 and 2019



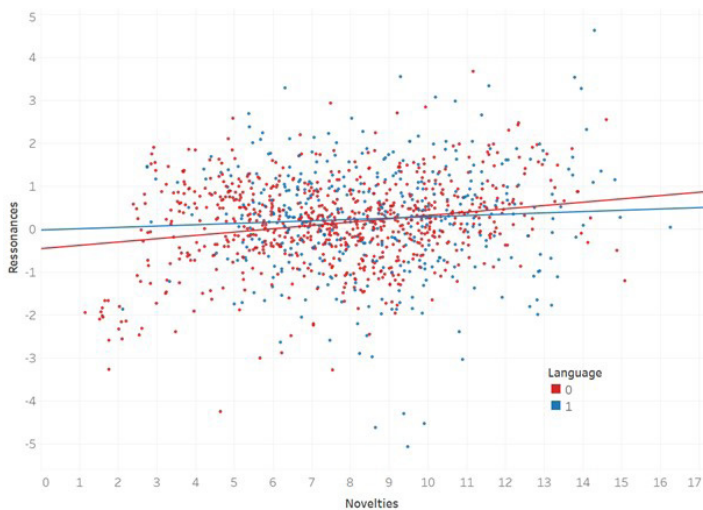
Source: Source: Authors' elaboration using data from Associação Nacional dos Centros de Pós-Graduação em Economia (2021) and code from Barron et al. (2018).

FIGURE A5
Novelty vs. Transience at National ANPEC meetings colored by year, without 2013 and 2019



Source: Authors' elaboration using data from Associação Nacional dos Centros de Pós-Graduação em Economia (2021) and code from Barron et al. (2018).

FIGURE A6
Novelty vs. Resonance at National ANPEC meetings, without 2013 and 2019



Source: Authors' elaboration using data from Associação Nacional dos Centros de Pós-Graduação em Economia (2021) and code from Barron et al. (2018).

TABLE A1
Topics in the ANPEC Annals in Portuguese

Tópico 1	Tópico 2	Tópico 3	Tópico 4	Tópico 5
Comércio	Inovação	Trabalhador	Espacial	Choque
Exportação	Tecnologia	Salário	Cidade	Inflação
Industria	Firma	Mulher	Regional	Regime
Comercial	Crescimento	Ocupação	Município	Ciclo
Importação	Técnico	Homem	Urbano	Previsão
Industrial	Patentes	Idade	Distância	Expectativa
China	Industrial	Renda	Centro	Monetário
Exportador	Interação	Hora	Indústria	IPCA
Insumo	Industria	Trabalhar	Trabalhador	Hiato
Trade	Inovativo	Casa	Densidade	Curva
Doméstico	Intensidade	Salarial	Mobilidade	Estrutural
Exportar	Rede	Diferencial	Concentração	Matriz
Setorial	Innovation	Branco	Aglomeração	Quebra
Matriz	Esforço	Chefe	Localização	Phillips
Mundial	Inovativo	Benefício	Transporte	Meta
Área 7 Economia Internacional	Área 9 Industrial e Tecnologia	Área 13 Economia do Trabalho	Área 10 Regional & Urbana	Área 4 Macro., Monetária & Finança
Tópico 6	Tópico 11	Tópico 12	Tópico 13	Tópico 14
Moeda	Emissão	Ciência	Marx	Desigualdade
Ativo	Energia	Economista	Capitalista	Pobreza
Crise	Cenário	Instituição	Dinheiro	Regional
Risco	Família	Conhecimento	Mercadoria	Pobre
Monetário	Ambiental	Conceito	Lucro	Nordeste
Carteira	Etanol	Científico	Capitalista	Região
Retorno	Simulação	Veblen	Acumulação	Decomposição
Liquidez	Energia	Humano	Força	Rural
Fluxo	Transporte	Mundo	Categoria	Educação
Global	Choque	Visão	Riqueza	Rendimento
Ação	Combustível	Pensamento	Classe	Sudeste
Juros	Equilíbrio	Hayek	Expansão	Norte
Crédito	Petróleo	Crítico	Troca	Família
Título	Climático	Prática	Mundial	Urbano
Câmbio	Insumo	Ação	Infraestrutura	Gini
Área 4 Macro., Monetária & Finança	Área 11 Agricultura & Meio Ambiente	Tópico 12 Historia do Pensamento	Área 2 Economia Política	Área 12 Economia Social & Demografica

Source: Authors' elaboration using data from ANPEC Meeting annals (ASSOCIAÇÃO NACIONAL DOS CENTROS DE PÓS-GRADUAÇÃO EM ECONOMIA, 2021a).

TABLE A2
Descriptive statistics by year

Variable	Observations	Mean	Srd. Dev.	Min	Max
Novelty 2013-2019	1,676	7.848	2.697	1.154	16.269
Transience 2013-2019	1,676	7.703	2.728	1.010	16.274
Resonance 2013-2019	1,676	0.145	1.219	-5.455	8.1634
Citations 2013-2019	1,676	2.554	7.750	0	149
Novelty 2013	242	7.382	2.988	1.292	14.624
Transience 2013	242	7.657	2.642	2.650	14.933
Resonance 2013	242	-0.295	1.598	-5.455	5.153
Citations 2013	242	5.289	9.661	0	94
Novelty 2014	238	7.841	2.738	1.154	14.833
Transience 2014	238	7.874	2.632	2.443	14.395
Resonance 2014	238	-0.033	1.113	-4.299	3.539
Citations 2014	238	4.874	14.212	0	149
Novelty 2015	239	7.700	2.716	1.520	14.967
Transience 2015	239	7.692	2.593	2.018	15.352
Resonance 2015	239	0.007	0.969	-3.008	3.287
Citations 2015	239	3.456	7.881	0	73
Novelty 2016	240	8.230	2.669	1.756	14.008
Transience 2016	240	8.110	2.720	1.607	14.864
Resonance 2016	240	0.121	1.061	-2.970	3.273
Citations 2016	240	1.896	4.487	0	51
Novelty 2017	237	8.092	2,815	1.756	16.269
Transience 2017	237	7.693	2.876	1.010	16.2738
Resonance 2017	237	0.405	1.051	-4.524	3.668
Citations 2017	237	1.155	4.309	0	50
Novelty 2018	240	8.055	2.352	2.601	14.310
Transience 2018	240	7.704	2.712	1.307	14.428
Resonance 2018	240	0.354	1.122	-5.069	4.629
Citations 2018	240	0.658	2.502	0	25
Novelty 2019	240	7.643	2.486	2.387	14.085
Transience 2019	239	7.192	2.860	1.565	14.275
Resonance 2019	239	0.463	1.318	-3.604	8.164
Citations 2019	240	0.179	0.554	0	4

Source: Authors' elaboration using data from ANPEC (ASSOCIAÇÃO NACIONAL DOS CENTROS DE PÓS-GRADUAÇÃO EM ECONOMIA, 2021a) and Google Scholar.

TABLE A3
Robustness tests for results in Table 2

Dep. Var.	I OLS 2013-2019 Citations	II Neg. Binomial 2013-2019 Citations	III Neg. Binomial 2014-2018 Citations
Resonance	0.421* (1.73)	0.019 (0.56)	0.123*** (3.08)
Brazilian journal	4.558*** (8.13)	1.434*** (15.23)	0.864*** (8.72)
Foreign journal	8.949*** (5.26)	2.243*** (15.77)	1.277*** (9.37)
English	0.509** (2.03)	-0.102 (-0.95)	-0.037 (-0.38)
Number authors	0.307* (1.95)	0.178** (3.92)	0.073* (1.67)
2014	-0.928 (-1.02)	-0.398*** (-3.27)	
2015	-2.361*** (-3.10)	-0.572*** (-4.46)	-0.214* (-1.85)
2016	-3.836*** (-5.27)	-1.208*** (-8.95)	-0.508*** (-4.29)
2017	-4.578*** (-5.92)	-1.611*** (-11.15)	-0.692*** (-5.23)
2018	-4.759*** (-6.92)	-2.300*** (-13.71)	-0.967*** (-6.67)
2019	-5.134*** (-7.39)	-3.284*** (-13.72)	
Constant	3.122*** (3.39)	0.085 (0.31)	0.642*** (2.82)
Observations Pseudo R-squared	1,676 0.23	1,675 0.15	529 0.13
Wald chi ² (55), p-value		1101.0, 0.000	416.6, 0.000

Source: Authors' elaboration using data from ANPEC (ASSOCIAÇÃO NACIONAL DOS CENTROS DE PÓS-GRADUAÇÃO EM ECONOMIA, 2021a) and Google Scholar. Robust errors. *10%, **5%, ***1%. Columns I and II include only papers with no citations. Base year is 2013 or 2014 in column III. Foreign and Brazilian publications compared to papers only published in the ANPEC annals. ANPEC area effects and department effects were included in the estimation but are not shown in the table.

TABLE A4
Fifteen papers with highest resonance values

Num	Resonance	Citations	Area	Year	Title / Authors
1	8.16	0	Area 5	2019	DO PROTESTS REACH THE BALLOTS. THE ELECTORAL DIVIDEND OF THE BRAZILIAN SPRING - Holanda, C. and Lima, R.C.
2	5.15	1	Area 10	2013	URBAN SPRAWL AND SPATIAL SEGREGATION IN SÃO PAULO METROPOLITAN REGION - Ramos, F.R. and Biderman, C.
3	4.80	1	Area 3	2013	HISTORICAL ORIGINS OF BRAZILIAN RELATIVE BACKWARDNESS – Barros, A.R.
4	4.63	0	Area 7	2018	IMPACTO DAS MEDIDAS NÃO-TARIFÁRIAS SOBRE O COMÉRCIO DE VALOR ADICIONADO - Araujo Jr, I.F, Perobelli, F.S. and Faria, W.R.
5	4.01	2	Area 11	2013	AMAZON MONITORING AND DEFORESTATION SLOWDOWN: THE PRIORITY MUNICIPALITIES - Rocha, R., Assunção, J. and Gandou, C.
6	3.90	0	Area 8	2019	ROTATIVIDADE DE TREINADORES E O DESEMPENHO DAS EQUIPES DE FUTEBOL NO BRASIL - Azevedo, C.O., Almeida, A.T.C. and Ramalho, H.M.B.
7	3.67	0	Area 6	2017	CARACTERÍSTICAS QUE INFLUENCIAM A PERCEPÇÃO DE CONFIANÇA NAS INSTITUIÇÕES E CORRUPÇÃO NO BRASIL - Monteiro, V.S., Justo, W.R., Rocha, R.M. and Castanheira, L.F.

Table created by the authors.

TABLE A4
Continued...

Num	Resonance	Citations	Area	Year	Title / Authors
8	3.54	0	Area 8	2019	LONG MEMORY AND TERM STRUCTURE OF INTEREST RATES - Valente, F. and Laurini, M.
9	3.54	0	Area 11	2014	CLIMATE CHANGE POLICY IN BRAZIL AND MEXICO: HOW SIMILAR ARE THE IMPACTS AND SOLUTIONS? - Gurgel, A.C., Octaviano, C. and Paltsev, S.
10	3.53	0	Area 13	2014	DOES MONEY MOVE TEACHERS? - Silva Filho, G.A., Pinto, G.C.X., and Vieira, M.T.
11	3.51	2	Area 6	2013	ENDOGENOUS LABOR EFFORT AND WAGE DIFFERENTIALS IN A DYNAMIC MODEL OF CAPACITY - Silveira, J.J. and Lima, G.T.
12	3.41	0	Area 1	2019	PATTERNS OF INTERDISCIPLINARY CITATIONS AND ASYMMETRY BETWEEN ECONOMICS - Silva, V.C. and Cavalieri, M.
13	3.38	0	Area 11	2019	IRRIGATION, TECHNICAL EFFICIENCY AND FARM SIZE IN BRAZIL - Morais, G.A.S., Silva, F.F., Freitas, C.O. and Braga, M.J.
14	3.34	0	Area 7	2019	VANTAGENS COMPARATIVAS AO NÍVEL DE FIRMAS: EVIDÊNCIAS INICIAIS PARA A INDÚSTRIA - Hidalgo, A.B., Casagrande, D.L. and Feistel, P.R.
15	3.34	110	Area 5	2014	TERM LIMITS AND POLITICAL BUDGET CYCLES AT THE LOCAL LEVEL: EVIDENCE FROM A YOUNG DEMOCRACY - Klein, F.A. and Sakurai, S.N.

Table created by the authors.

TABLE A5
Fifteen papers with highest novelty values

Num	Novelty	Citations	Area	Year	Title / Authors
1	16.27	0	Area 9	2017	LENIENCY AND DAMAGE LIABILITY IN BRAZIL: THE EFFECTS ON COLLUSIVE BEHAVIOR - Pinha, L.C. and Braga, M.J.
2	15.09	0	Area 9	2017	O USO DE FILTROS DE CARTÊIS: UMA APLICAÇÃO PARA O CASO DO VAREJO - Silva, A.S., Vasconcelos, S. and Vasconcelos, C.
3	14.97	0	Area 9	2015	LOST IN TIME AND SPACE: THE DETERRENCE EFFECT OF CARTEL BUSTS ON THE RETAIL - Grezzana, S.
4	14.88	0	Area 9	2015	DINÂMICA DE PRECIFICAÇÃO EM MERCADOS CARTELIZADOS: O CASO DA GASOLINA - Silva, A.S., Vasconcelos, S. and Vasconcelos, C.
5	14.83	10	Area 8	2014	THE 2D:4D RATIO AND MYOPIC LOSS AVERSION (MLA): AN EXPERIMENTAL INVESTIGATION - Teixeira, A.M., Tabak, B.M. and Cajueiro, D.O.
6	14.62	2	Area 3	2013	FOREIGN ELECTRICITY COMPANIES IN ARGENTINA & BRAZIL: THE CASE OF AMERICAN - Saes, A.M. and Lanciotti, N.
7	14.61	0	Area 8	2014	THIN SUBSIDIES NO BRASIL: UMA INVESTIGAÇÃO DOS SEUS EFEITOS SOBRE A DEMANDA DE FRUTAS - Silva, M.M.C. and Coelho, A.B.
8	14.48	3	Area 13	2014	LABOR MARKET EQUILIBRIUM EFFECTS OF CASH TRANSFERS - EVIDENCE FROM A STRUCTURAL MODEL - Lehmann, M.C.
9	14.46	3	Area 1	2013	LUCAS'S EARLY RESEARCH IN THE 1960'S - Silva, D.F.R.

Table created by the authors.

TABLE A5
Continued...

Num	Novelty	Citations	Area	Year	Title / Authors
10	14.16	1	Area 3	2013	HISTORICAL ORIGINS OF BRAZILIAN RELATIVE BACKWARDNESS - Barros, A.R.
11	14.31	0	Area 7	2018	IMPACTO DAS MEDIDAS NÃO-TARIFÁRIAS SOBRE O COMÉRCIO DE VALOR ADICIONADO - Araujo Jr., I.E.A., Perobelli, F.S. and Faria, W.R.
12	14.27	3	Area 8	2014	PROPAGATION OF SYSTEMIC RISK IN INTERBANK NETWORKS - Quadros, V.H., Gonzalez-Avell, J.C. and Iglesias, J.R.
13	14.21	4	Area 9	2015	IMPACTO DE FUSÕES E AQUISIÇÕES SOBRE A QUALIDADE DO ENSINO SUPERIOR - Garcia, C.P. and Azevedo, P.F.
14	14.13	0	Area 12	2018	RELAÇÃO ENTRE EXPOSIÇÃO À VIOLÊNCIA E HABILIDADES SOCIOEMOCIONAIS: O CASO DOS ... - Silva, W.P., Scorzave, L.G., Sarmiento, C.M. and Santos, D.
15	14.11	2	Area 11	2013	AMAZON MONITORING AND DEFORESTATION SLOWDOWN: THE PRIORITY MUNICIPALITIES - Rocha, R., Assunção, J.C., and Gandour, C.

Table created by the authors.

Author's contribution:

A. Literature review and problematization: Marcos Paulo R. Correia and Bernardo Mueller

B. Data collection and statistical analysis: Marcos Paulo R. Correia and Bernardo Mueller

C. Preparation of figures and tables: Marcos Paulo R. Correia and Bernardo Mueller

D. Manuscript development: Marcos Paulo R. Correia and Bernardo Mueller

E. Bibliography selection: Marcos Paulo R. Correia and Bernardo Mueller

Conflict of interest: The authors declare that there is no conflict of interest.

Funding: The authors declare there was no funding for this project.



This is an Open Access article distributed under the terms of the Creative Commons Attribution license, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.