

Dynamics of senses of new physics discourse: co-keywords analysis

Yurij L. Katchanov^{a,*}, Yulia V. Markova^b

*^aInstitute for Statistical Studies and Economics of Knowledge, National Research
University Higher School of Economics 20 Myasnitskaya Ulitsa, Moscow 101000,
Russian Federation*

*^bAmerican Association for the Advancement of Science 1200 New York Ave NW, 20005,
Washington, DC, USA*

Abstract

The paper presents a longitudinal analysis of the evolution of new physics keywords co-occurrence patterns. For that, we explore the documents indexed in the INSPIRE database from 1989 to 2018. Our purpose is to quantify the knowledge structure of the fast-growing subfield of new physics. The development of a novel approach to keywords co-occurrence analysis is the main point of the paper. In contrast to traditional co-keyword network analysis, we investigate structures that unite physics concepts in different documents and bind different documents with the same physics concepts. We consider the structures that reveal relationships among concepts as topological and call them “physics senses”. Based on the notion of trajectory mutual information, the paper offers clustering of physics senses, determines their period of life, and constructs a classification of senses’ “authority”.

Keywords: Co-occurrence analysis, Keywords, New physics, INSPIRE

1. Introduction

Throughout the modern history of physics, high energy physics (or HEP for convenience) has been and remains one of the main driving forces of conceptual and methodological innovations. It is also a front-runner in the

*Corresponding author

Email addresses: yurij.katchanov@gmail.com (Yurij L. Katchanov),
yulia.markova@gmail.com (Yulia V. Markova)

development of scientific tools for publication and communication. E.g., the INSPIRE¹ database is the leading publication platform indexing HEP publications (Hecker, 2017). All these factors make HEP an interesting subject of study inside the science of science approach (Fortunato et al., 2018) (hereafter, abbreviated as SciSci). One of the principal objectives of SciSci is to comprehend the emergence of new knowledge, and HEP can provide a first-choice object of research in this regard.

The somewhat bureaucratic term of “the Standard Model” traditionally defines the monumental achievement of the twentieth-century HEP—the product of efforts by thousands of researchers spanning over 50 years (Langacker, 2017). The Standard Model constitutes a contemporary view of particle physics that combines theoretical and experimental results of research into the properties and interactions of particles. With no deviations from the Standard Model discovered to date, it remains on the cutting edge of our experimentally proven understanding of the fundamental laws of the microworld. However, there has to be a new physics beyond the Standard Model—something that researchers have been working to uncover for decades now. This smart term is used to describe the fundamental layer of the physics reality at the core of the structure of the universe, destined to replace the Standard Model as the leading theory.

New physics is not a specific theory, but rather a collective term for everything that is at variance with the Standard Model (Ghosh et al., 2020). Judging by abundant circumstantial evidence, new physics does exist but remains hard to define. For that reason, a multitude of various methods have been applied in search of its manifestations. Spanning decades, these concentrated and diverse experimental and theoretical quests for new physics make it into a unique testing ground for the development of new methods of the quantitative study of science.

This paper presents a longitudinal analysis of the evolution of new physics keywords co-occurrence patterns based on the documents indexed in the INSPIRE database from 1989 to 2018. We consider keywords co-occurrence structure as an operationalization of the new physics knowledge structure. These patterns reveal the relationships between new physics knowledge entities.

Our purpose is to quantify the knowledge structure of the fast-growing

¹<https://inspirehep.net>

new physics' subfield. The issue addressed in the paper is the development of a novel approach to keywords co-occurrence analysis.

The remainder of this paper is organized as follows. Section 2 refers to related studies on a given topic. Section 3 presents the methodological framework. Section 4 explains the data used in our analyses. Section 5 reports our main results and generalizes the findings of the paper. Finally, in Section 6, we draw the conclusions.

2. Background

Quantitative research on physics, physicists and the discourse of physics has by now become a well-established tradition (Sinatra et al., 2015; Colavizza and Franceschet, 2016; Battiston et al., 2019; Sun and Latora, 2020). It allows for a shift from descriptive analysis and general arguments to more keen and detailed research on the evolution of physics as a scientific field (Liu et al., 2019). The availability of extensive publication databases and advanced data processing tools open the door to new interdisciplinary research opportunities. Take, e.g., the study of the evolution of research interests or physics subject areas (Jia et al., 2017; Aleta et al., 2019; Zeng et al., 2019; Coccia, 2020), physics research institutions (Katchanov et al., 2016), careers pathways (Deville et al., 2014; Petersen, 2018; Xing et al., 2019), or strategies and talent of physicists (Pluchino et al., 2019; Tripodi et al., 2020). Due to its nature, physics has its own information environment, and researchers have access not only to general commercial databases like Web of Science or Scopus but also to specific physics publications repositories like arXiv, INSPIRE, or APS. Thanks to the openness of these specialized data sources, researchers may use the full range of quantitative methods to study physics.

Quantitative analysis of physics concepts provides a good opportunity to explore the relations between external classifications of publications and the outcomes of community detection algorithms (Palchykov et al., 2016). The authors demonstrated that classifications made by authors had some discrepancies from automatically retrieved communities. These differences may carry important information about relations between physics fields that the external observer does not realize. M. Krenn and A. Zeilinger used a large corpus of publications covering about a hundred years to train a neural network and try to predict future trends in physics (Krenn and Zeilinger, 2020). A large body of literature opens the way to interesting studies of cognitive aspects of science. Thus, S. Milojević compared the growth of

scientific literature and the lexical diversity of research articles titles. She found that periods of fast cognitive growth did not coincide with the rise of publication growth (Milojević, 2015).

Recent advances in text analysis methods have facilitated detailed quantitative research on physics concepts. Thus, M. Herrera, using publications and PACS categories, built physics fields and traced their evolution for a period of 20 years (Herrera et al., 2010). M. Fontana et al. (Fontana et al., 2020) analyzed more than 230 thousand articles published in 8 journals of the American Physical Society and 2.4 million citations to measure novelty and interdisciplinarity. R.K. Pan et al. studied the evolution of interdisciplinary research in physics (Pan et al., 2012). They constructed a network of publications and showed that interactions between physics fields were increasing, and the area of interdisciplinary research was growing. Using publications data spanning more than a century, M. Perc explored patterns and trends of physics terminology. The analysis revealed that the evolution of scientific paradigms in physics was governed by principles of self-organization (Perc, 2013).

Machine learning methods gained popularity in SciSci in recent years (Xu et al., 2020a). Thus, K. McKeown et al. (McKeown et al., 2016) presented a system that predicts the impact of a physics concept; M. Chinazzi et al. (Chinazzi et al., 2019) mapped research space in physics using the American Physical Society publications database; A. Palmucci et al. (Palmucci et al., 2020) suggested a representation for the relative movements of physics' domains in a multi-dimensional space based on the PACS codes.

Network analysis has a crucial part to play in the SciSci (Zeng et al., 2017) these days. Given its universal scope, network analysis allows the study of various aspects of physics as a scientific field. In recent years, a wealth of research has been produced on co-authorship networks in physics (Heiberger and Wieczorek, 2016; Silva et al., 2019), co-citation networks (Zhang et al., 2013; Renoust et al., 2017; de Arruda et al., 2018), and keyword co-occurrence networks (Palchykov et al., 2021; You et al., 2021; Karimi et al., 2021). In the past two decades, the application of co-keyword analysis, first used by Callon et al. (Callon et al., 1983), was both far-ranging and far-reaching, covering relevant aspects of research topic networks (Cobo et al., 2011; Yi and Choi, 2012; Leydesdorff and Nerghes, 2017; Radhakrishnan et al., 2017; Wen et al., 2017; Behrouzi et al., 2020; Wang et al., 2021).

When studying co-authorship networks, co-citation or co-keyword networks, graph theory is often perceived as a mathematical framework under-

lying the SciSci (Mingers and Leydesdorff, 2015). At the same time, quantitative science studies are evolving, and so is their mathematical apparatus. In a number of papers, a depiction produced with customary network tools is supplemented with that created with tools from applied topology. E.g., (Salnikov et al., 2018; Christianson et al., 2020) use a simplicial complex approach to word co-occurrences in order to explore the conceptual landscape of mathematical publications. A. Patania et al. (Patania et al., 2017) develop simplicial descriptions of publications and persistent homology to track the growth and development of collaborations based on co-authorship data.

3. Methods

3.1. Problem formulation

We apply the co-word analysis technique in order to map the structural dynamics of the “new physics” subfield. This analysis has proven effective for the study of relationships and interrelationships among scientific concepts (van Raan, 2019). The basic principle of co-word analysis is to characterize the underlying texts through the use of co-occurrence networks consisting of nodes that represent the keywords and edges between them (Callon et al., 1991).

Physics knowledge is a complex system, at its basic level consisting of physics concepts and their interactions. From the SciSci point of view, the complexity of the physics knowledge system largely results from the variety of interactions between concepts. The network approach focuses on the analysis of pairwise relationships between concepts (terms, keywords), invariably reducing multidimensional structures and losing information. Nevertheless, the recent decades saw a great number of intellectual structures described successfully as networks where co-occurring pairs of concepts are connected by links (Li et al., 2016; Wen et al., 2017; Radhakrishnan et al., 2017; Lozano et al., 2019). These achievements are prompting the quantitative science studies to develop new methods of analysis (Cheng et al., 2020; Joslyn et al., 2021).

The scientometrics community is in agreement that system of well-separated and well-connected clusters representing keywords in a co-occurrence network can ensure an adequate description of a research discourse (van Eck and Waltman, 2014; Hosseini et al., 2021). When such network representation of scientific discourse is assumed, a strong assumption is made: the overall interplay among the concepts is suggested to be completely described by pairwise

co-occurrence data. In reality, however, the constraints in a co-occurrence network may prove to be fairly weak, and so the identification of the true structure of research themes may require that more complex interactions are taken into account that help reveal more interplay among physics concepts. The problem of the context in which a keyword co-occurrence network is generated was first described in (Bornmann et al., 2018) and was studied in more detail in (Cheng et al., 2020). However, the proposed heuristic solutions concern the use a citation network and are beyond the scope of keyword co-occurrence analysis.

Research communication has a distinguishing feature, which is that all components of the system influence each other in a direct or indirect way. Each scientific publication features several concepts simultaneously, and one should not assume *a priori* that their interactions can be explained in terms of dyads (see details in (Battiston et al., 2020)).

3.2. *Physics sense: definition*

The sense of a physics concept cannot be fully explained by its formal definition. Its understanding is greatly improved by a series of examples illustrating its use in documents. For working physicists, these examples represent both motivation and the sense of a concept. It is for this reason that exploring the co-occurrence of physics concepts can prove to be a significant element in the study of the cognitive structure of physics. The sense helps to reveal the essence of a physics concept in the research context. A physics sense determines the place of the concept in scientific communication. The sense introduces the “whole—part” relationship, making the concept essential as part of a whole which is the system of a physics discourse.

Specification of documents that mention a physics concept is imperative for the understanding of its physics sense. We consider a physics sense in the framework of an “information field”. The concept of the field emerges naturally as one analyzes the forming of a physics sense within a system of relationships between physics concepts. Information field is a holistic structure that describes both direct and indirect mutual influences of physics concepts in documents. In this paper, the intuitive viewpoint is that an information field that determines a physics sense is a topological space of physics concepts and their properties.

Following the generally accepted definition (see, e.g., (Klement, 2001; Deleuze, 1993)), one understands physics sense as a proposition including a set of physics concepts. The intellectual value of a concept is the result of its

position in a complex network of semantic relations. The study of physics discourse in its entirety involves a structural comparison of the concepts. Our methodology supposes that to study physics concepts, one needs to construct statistical relations between them. This type of relations we call “sense”.

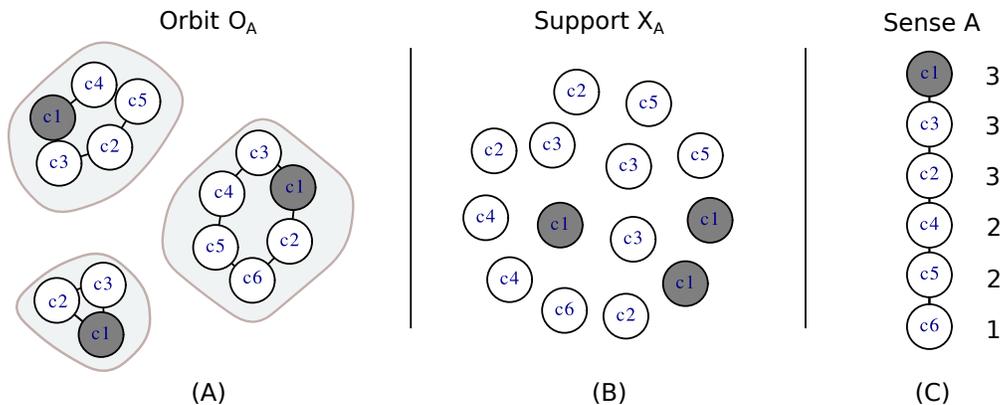


Figure 1: Construction of physics sense. Part A shows how the documents are related together in the orbit O_A of the concept $x_{\alpha 0} \equiv c_1$ (marked with a dark circle). Part B depicts the support X_A of a sense A . Part C displays the sense A .

For our further reasoning we will need auxiliary concepts. Let us refer to the set of all documents that mention the physics concept $x_{\alpha 0}$ as the orbit O_A of that concept. We will understand the spectrum S_A of physics concept $x_{\alpha 0}$ as a set of physics concepts $x_{\alpha 0}, x_{\alpha 1}, \dots, x_{\alpha k}$ corresponding to the orbit O_A . (Clearly, $x_{\beta 0} \in S_A \Leftrightarrow x_{\alpha 0} \in S_B$, etc.) Let X_A (with or without indices) be a sample of M physics concepts that occur in the orbit O_A with frequencies $m_{\alpha 0}, m_{\alpha 1}, \dots, m_{\alpha k}$, such that $\sum_{\alpha} m_{\alpha} = M$.

Let σ_A be a configuration of the set $X_A = \{x_{\alpha}\}_{\alpha \in A}$. The configuration σ_A can be represented in the form

$$(\forall x_{\alpha} \in S_A \subset O_A): \sigma_A = \begin{pmatrix} x_{\alpha 0} & x_{\alpha 1} & \dots & x_{\alpha k} \\ m_{\alpha 0} & m_{\alpha 1} & \dots & m_{\alpha k} \end{pmatrix}. \quad (1)$$

The set X_A will be called hereafter the support of a physics sense A (see Fig. 1). A topological structure on the support X_A is a collection τ_A of subsets of X_A which includes the empty set \emptyset and the whole support X_A and which is such that: a) the intersection of any number of elements of τ_A belongs to τ_A and b) the union of any set of elements of τ_A belongs to τ_A (Singh, 2019). In our case, the collection of all subsets of X_A is a topology on X_A , and τ_A is called the discrete topology.

One can think of S_A as having been obtained by “identifying” each pair of equivalent elements of the support X_A . Hence, the surjective map $\pi: X_A \rightarrow S_A$ carries each point of the support X_A to the equivalence class ω containing it. In the quotient topology induced by π , the set Ω such that $\Omega = \cup_{\alpha \in X} \omega_\alpha$ is called a quotient space of X_A (Singh, 2019). For further details, we refer to Appendix A.

In this paper, a *physics sense* A of concept $x_{\alpha 0}$ is a finite topological space (X_A, τ_A) —a support X_A with a discrete topological structure τ_A . By putting it this way, we symbolically equate the space (X_A, τ_A) with its focal point—the concept $x_{\alpha 0}$. The main point here is to realize that the topological structure τ_A is the elementary property of physics sense as a whole. It means this structure refers to the physics concepts collection X_A , but not to its parts or connections between these parts.

Knowing the configuration σ_A Eq.(1) of the physics sense A enables us to define a state function $h(A) = -\sum p(\omega) \ln p(\omega) = \sum_\alpha p(\alpha) \ln p(\alpha)$ (see Eq. (A.4))—the topological entropy, where $p(\alpha)$ is the probability of the realization of the physics concept x_α in the orbit O_A . The interpretation of entropy $h(A)$ is that it is a measure of the number of arrangements the physics concepts could be in.

3.3. Basic quantities

3.3.1. Dimension

The quantity of dimension $\dim(A)$ Eq. (A.1) is the number of the physics concepts in the spectrum S_A . It is possible to interpret a dimension as a number of degrees of freedom of the information field of a physics sense. A more straightforward interpretation of a dimension is that \dim constitutes the number of information channels that connect a single sense with others.

3.3.2. Trajectory mutual information

Thus, we attempted to describe the sense of a physics concept through a configuration of the information field where the concept functions in the discourse. In order to characterize the thematic structure of physics discourse itself, we need to examine the configuration of the field induced by all relevant physics senses. Since we have conceptualized a physics sense as a topological space of concepts within a single orbit, it is only natural to consider the interaction of physics senses as a topological space constructed on the intersection of respective orbits. If the orbits of physics concepts in-

tersect, it indicates that the concepts share a common context and we can evaluate its significance using the concept of topological entropy.

Let us consider physics senses A and B , i.e., topological spaces (X_A, τ_A) and (X_B, τ_B) with orbits O_A and O_B and spectra S_A and S_B so that $x_{\alpha 0} \in S_B$ (and, respectively, $x_{\beta 0} \in S_A$). On the intersection of orbits O_A and O_B as well as spectra S_A and S_B , a topological space (X_{AB}, τ_{AB}) is defined that describes the interplay between the physics senses A and B . One can easily define the support $X_{AB} = X_A \cap X_B$

$$(O_A \cap O_B) (\forall x_{\alpha\beta} \in S_A \cap S_B) : X_{AB} = \{x_{\alpha\beta}\}_{\alpha\beta \in AB}, \quad (2)$$

the discrete topology τ_{AB} on X_{AB} , and the quotient space $(\Omega_{AB}, \varpi_{AB})$ as shown in Appendix A.

We can relate the trajectory of the physics sense $A^T = \{A^1, A^2, \dots, A^l\}$ in time $T = \{t\}$ with the trajectories of all other physics senses $\{B_\gamma^t\}$ from the totality $\Gamma \setminus \{A^t\}$. Our goal here is to characterize information transmission between senses' trajectories via their path mutual information

$$\mu(A^T) = \sum_{t \in T} \sum_{\gamma=1}^{\dim(A)} p_{(A^t, B_\gamma^t)}(\alpha, \beta) \ln \frac{p_{(A^t, B_\gamma^t)}(\alpha, \beta)}{p_{A^t}(\alpha) p_{B_\gamma^t}(\beta)}, \quad (3)$$

where $p_{(A,B)}(\cdot)$ is the joint probability mass function of A and B , $p_A(\cdot)$ and $p_B(\cdot)$ are the marginal probability mass function of A and B , respectively. The quantity $\mu(A^T)$ will be called the trajectory mutual information (TMI, from now on) of physics sense A .

To derive a distribution of probabilities $\mu(A^T)$, we must characterize the principle whereby out of the set of all possible quotient spaces $\{(\Omega_A, \varpi_A)\}$ (see Appendix A), we select those that are realized. The comparison of quotient topological spaces we have drawn here allows for formulation of an extreme principle for the changeability of a physics sense: from the given state (Ω_A, ϖ_A) the sense tends towards a state (Ω_B, ϖ_B) with the maximum value of $\mathbf{P}\{B\}$ (Eq. (A.3)). We now admit that the actual trajectory of the physics sense represents a most probable trajectory. This is equivalent to the transition to a physics sense with the maximum value of the TMI. It is easy to derive the Fréchet extreme distribution for TMI.

The challenge facing us at this point is to approximate the probability distribution function of $\mu(A^T)$. The quantity of TMI represents a sum total of the information activity of a physics sense, understood as a sum of its own

activity and that of the senses whose orbits intersect its own. We are dealing here with values of the sums of the form

$$Z_\gamma = \sum_{\gamma=1}^{\dim(A)} \zeta_\gamma = \sum_{\gamma=1}^{\dim(A)} \sum_{t \in T} p_{(A^t, B_\gamma^t)}(\alpha, \beta) \ln \frac{p_{(A^t, B_\gamma^t)}(\alpha, \beta)}{p_{A^t}(\alpha) p_{B_\gamma^t}(\beta)},$$

such that $\mu(A^T) = Z_\gamma$. Since we postulated that the true trajectory of the physics sense A is corresponding to the maximum of $\mu(A^T)$, we must consider the limit behavior of extremes

$$(d \rightarrow \infty) (\dim(A) \rightarrow \infty) : \hat{Z}_{d\gamma} = \max_{1 \leq \delta \leq d} \sum_{\gamma} \zeta_{\delta\gamma}. \quad (4)$$

Let the variables $(\delta, \gamma \geq 1)$: $\zeta_{\delta\gamma}$ be mutually independent and have a common distribution F . We assume that F belongs to the class of subexponential distributions (Foss et al., 2013): the convolution tail $\overline{F * F}(u)$ is equivalent to $2\overline{F}(u)$ as $u \rightarrow \infty$, where $\overline{F}(u) = 1 - F(u)$. Assume also that F belongs to the class of distributions with regularly varying tails, i.e., $\overline{F}(u) \sim u^{-a} L(u)$, $u \rightarrow \infty$, where $a > 0$ and $L(u)$ is a slowly varying function (Bingham et al., 1987). This is not necessarily true, but it represents a reasonably good approximation in a large number of subsystems of science (Zeng et al., 2017). Under the conditions considered here, can be obtained that the family of extremes similar to $\hat{Z}_{d\gamma}$ has well-known Fréchet extreme distribution $\mathbf{P}\{Z' \leq z\} \rightarrow \exp(-z^{-a})$, $0 < a < 2$ (Lebedev, 2005). It means, among other things, that bigger TMI values are normally a result of individual senses having higher information activity rather than a combined activity of a large number of senses.

3.4. Assistant quantities

3.4.1. Similarity

To derive similarity between (X_A, τ_A) and (X_B, τ_B) , one should keep in mind that the value Eq. (A.3) characterizes the probability of a particular state of a physics sense. It follows that the similarity between physics senses A and B will be given by

$$s_{AB} = \ln \frac{p_{(A,B)}(\alpha, \beta)}{p_A(\alpha) p_B(\beta)}.$$

This expression is nothing but the point-wise mutual information.

3.4.2. Normalized entropy

To suppress the dependence of the entropy $h(A)$ on k in Eq. (A.4) and keeping in the mind Eq. (A.1), we consider the normalized entropy $h_n(A)$ defined as followed:

$$h_n(A) = \frac{h(A)}{\ln \dim(A)}. \quad (5)$$

3.4.3. Complexity

Following (López-Ruiz et al., 1995; Martin et al., 2006; Rosso et al., 2007), we computed the values of the so-called statistical complexity

$$c(A) = h_n(A)d(A), \quad (6)$$

where the “disequilibrium” $d(A)$, which measures the deviation of the probability distribution $\mathbb{P}\{A\}$ from uniform probability distribution, is defined as

$$d(A) = \sum_{\alpha=1}^{\dim(A)} \left(p(\alpha) - \frac{1}{\dim(A)} \right)^2.$$

By construction, the quantity $d(A)$ vanishes for probability distributions $\mathbb{P}\{A\}$ that correspond to absolute order and maximal randomness.

3.4.4. Transitivity

The fraction of pairs of physics senses that are themselves intersected defines the sense’s transitivity $tr(A)$ (Newman, 2018):

$$tr(A) = \frac{\# \text{ pairs of senses whose orbits intersect with } O_A}{\# \text{ pairs of senses}}. \quad (7)$$

The transitivity is an important quantity, which informs on the density of a sense’s information field.

3.4.5. Citation rate

To construct the citation rate of the concepts, we used a three-year citation window and fractional counting. It means the citation rate of the concept x of document y is a share of all other concepts of this document. Thus the total concept citation rate is the sum of all its fractional citations in all documents where it is used.

4. Data

The publication data for this study comes from the INSPIRE database. INSPIRE is a leading information system in HEP (Xu et al., 2020b) that provides open access to more than 1 million scientific publications. It continues the SPIRES physics literature information system. The INSPIRE database is the result of close collaboration between such institutes as CERN, DESY, Fermilab, and SLAC. It brings together information from arXiv.org, NASA-ADS, PDG, and other HEP publishers. In addition to documents, INSPIRE integrates information about authors, institutes, journals, conferences, and jobs in the field of HEP. This complex integrated system provides opportunities to analyze information processes in HEP from different points of view. The INSPIRE system is of interest not only to physicists but also to researchers in quantitative science studies (Zingg et al., 2020). Researchers who have worked with INSPIRE datasets pointed out that the data is well-curated and easily accessible (Jang and Ko, 2019; Strumia and Torre, 2019).

Another significant aspect of the INSPIRE information system is the keywords that accompany documents. The keywords are automatically extracted with the help of regular expressions from full-text papers by the BibClassify extraction algorithm². This algorithm draws upon a controlled vocabulary resulting from the HEP Index (HEPI). The HEPI thesaurus is maintained by the German High Energy Physics Laboratory DESY since 1963. It is prepared by the subject specialists and updated occasionally. The HEPI thesaurus is actively used by the principal HEP institutes and information databases. The BibClassify algorithm extracts two types of keywords: single keywords and composite keywords. Composite keywords are combined with several single keywords. If single keywords are placed next to each other, they form a composite keyword. The words of composite keywords are separated by a colon e.g. “Higgs particle: mass” or “CP: violation”. The final keyword list consists of top 20 first best single and composite keywords. These keywords serve as an information base for the study.

Our dataset counts documents with the term “new physics” in titles, abstracts, or keyword lists published in 1989—2018³. To make the sample

²For details, <http://cds.cern.ch/help/hacking/bibclassify-extraction-algorithm>, last accessed on 08 December 2020.

³Data was retrieved from the INSPIRE database in October 2019. Processing of the XML dump, data preparation, and statistical analysis were performed with the R pro-

more stable and consistent, we excluded some irrelevant documents and concepts. First of all, we removed the documents featuring only one physics concept or no concepts at all. Secondly, we removed keywords having errors due to the indexing procedure and keywords, denoting document types like “book”, “thesis”, “report”, and so on. After these procedures, the final dataset counted 19,830 documents and 8,822 physics concepts. The collection of documents thusly obtained was subjected to further statistical analysis.

5. Results and discussion

5.1. Statistical features of new physics discourse

Fig. 2A illustrates the number of new physics unique concepts indexed in INSPIRE in the years 1989–2018. It can be seen that the figures gradually increase from between 1989 and 1995. Then they show some fluctuation to proceed with rapid growth after 2006. In 2006–2015, one can note a quick rise of numbers and even a spike in 2013 and 2015. In 2015, the number of concepts reached its maximum value of 4259 items, followed by the curve dropping sharply to 3557 concepts in 2018. For the whole 30-year period, the number of unique concepts rises from 373 to 4259 items, and the curve fits the exponential regression line (cf. (Milojević, 2015)).

Fig. 2B, Fig. 2C, and Fig. 2D show approximations of the probability density functions of dimension, citation, and TMI, respectively. One can spot at a glance the qualitative difference of the dimension and the TMI probability distributions from that of citation. Indeed, the probability density functions in Fig. 2B and Fig. 2D are of Fréchet type, whereas Fig. 2C shows a power-law distribution function. It seems only natural that the citation distribution can be fitted by a Pareto distribution (Zeng et al., 2017; Fortunato et al., 2018). The Fréchet distribution observed for dimension is untypical of scientometric quantities of this kind (cf. (Börner and Milojević, 2019)). A quantity homologous to dimension, the “node degree” is normally distributed according to a power law (Zeng et al., 2017; Palchykov et al., 2021). We have, in turn, predicted the Fréchet distribution for TMI.

Despite the obvious differences between the dimension and the TMI probability distributions, on the one hand, and that of citation, on the other, they

programming language. Citation data was downloaded from the INSPIRE database in August 2021.

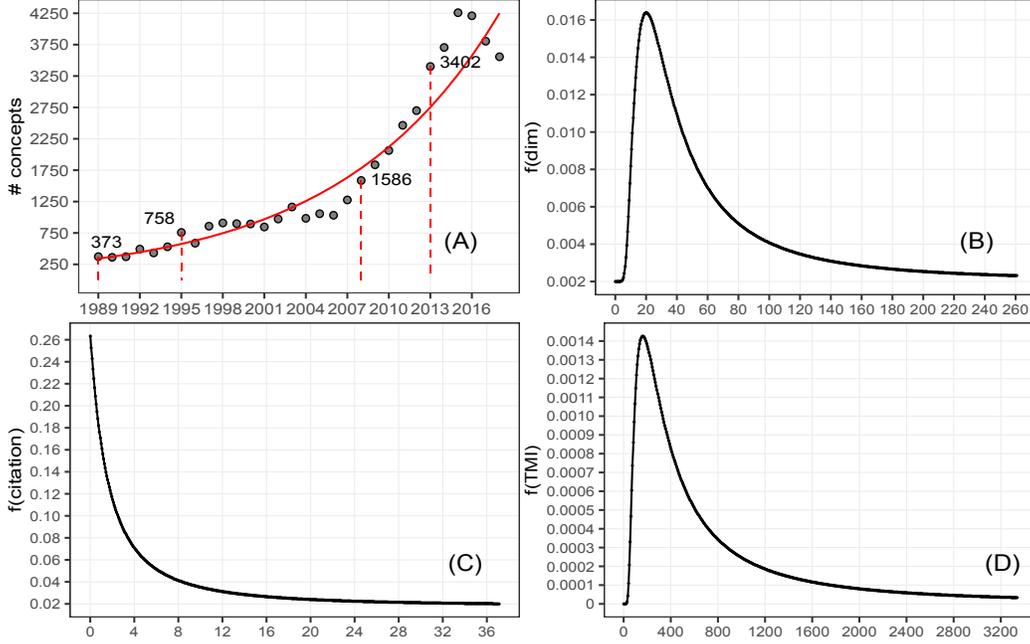


Figure 2: Features of new physics discourse.

(A) Dynamics of new physics concepts in 1989–2018. The equation of the regression curve: $\# \text{ concepts} = 312.612 \cdot \exp(0.087t)$ ($R^2 = 0.954, p = 0.000$). Dashed vertical lines indicate the generations of physics senses (see 5.3 “Generations of physics senses”).

(B) The approximation of the probability density function of the dimension: a Fréchet distribution with parameters $\alpha = 1.1978, \beta = 34.672$ (Kolmogorov–Smirnov statistic 0.0137.)

(C) The approximation of the probability density function of the citation: a Pareto Type II distribution with parameters $\alpha = 0.7986, \beta = 3.0318$ (Kolmogorov–Smirnov statistic 0.0145.)

(D) The approximation of the probability density function of the TMI: a Fréchet distribution with parameters $\alpha = 0.8947, \beta = 382.26$ (Kolmogorov–Smirnov statistic 0.0128.)

share an internal similarity. The fact is, the Pareto distribution belongs to the maximum domain of attraction of the Fréchet distribution (Foss et al., 2013). The asymptotic behavior of citation distribution is similar to that which defines the distribution of dimension and TMI. In particular, larger values of citation, dimension, and TMI are achieved not through a larger number of physics senses but due to individual physics senses with higher activity.

The Fréchet extreme distribution is a distribution of maximums; it expresses relatively enduring structures of the dimension and the TMI. These

structures are characterized via the probability of extreme events (exceedances of a high level). Both the dimension Eq. (A.1) and the TMI Eq. (3) are high outliers by nature, so it is only logical that they are described by the Fréchet extreme distribution. This distribution is a pattern or arrangement of physics senses; we can surmise, therefore, that it is the trends defined by record values that form the basis of a nascent new physics discourse.

The citation structure of physics senses is, in turn, described by the Pareto distribution. This distribution signifies the fact that very few physics senses possess an extremely high citation rate, while very many demonstrate an extremely low one.

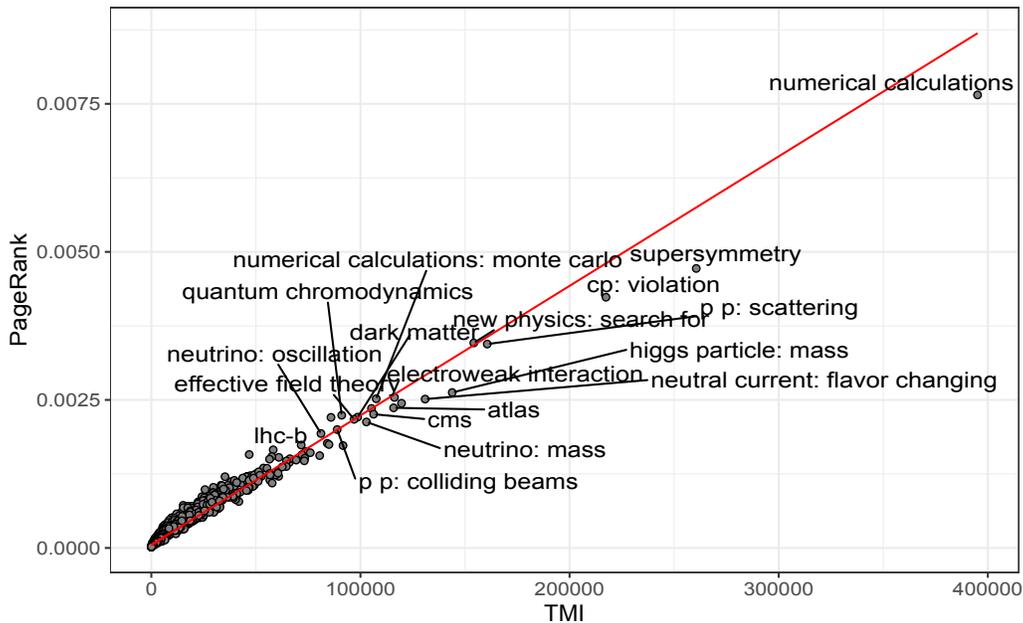


Figure 3: Scatterplot of PageRank against the TMI.

5.2. PageRank and TMI

In order to study the dynamics of the “importance” and “authority” of physics senses in the years 1989–2018, the PageRank value of physics senses was calculated with the use of the *igraph* software package. The Top-30 physics senses of the PageRank value are listed for illustration and comparison, as shown in Table 1.

Table 1: The Top-30 physics senses of the PageRank value

PageRank	Physics sense	dim
1	numerical calculations	5762
2	supersymmetry	3962
3	cp: violation	3523
4	new physics: search for	3040
5	p p: scattering	3007
6	higgs particle: mass	2506
7	electroweak interaction	2398
8	numerical calculations: monte carlo	2351
9	neutral current: flavor changing	2324
10	minimal supersymmetric standard model	2287
11	atlas	2238
12	electron positron: annihilation	2119
13	cms	2138
14	quantum chromodynamics	2052
15	dark matter	2099
16	effective lagrangian	2013
17	effective field theory	2099
18	neutrino: mass	2055
19	p p: colliding beams	1902
20	neutrino: oscillation	1823
21	higgs particle: coupling	1781
22	lepton: flavor: violation	1725
23	lhcb	1644
24	coupling: yukawa	1784
25	numerical calculations: interpretation of experiments	1512
26	angular distribution: asymmetry	1600
27	electroweak interaction: symmetry breaking	1643
28	interpretation of experiments	1504
29	batavia tevatron coll	1608
30	top: pair production	1592

Fig. 3 shows PageRank as a linear function of the TMI. The relationship between the two quantities is described by the regression equation $\text{PageRank} = b \cdot \mu(A^T) + a$, where $b = 2.187 \cdot 10^{-8}$, $a = 5.157 \cdot 10^{-5}$ ($R^2 = 0.958$, $p = 0.000$). The importance of physics senses is directly proportional to their measure of the mutual dependence: the stronger the sense is connected with others, the higher its authority. This means that the authority of a sense is determined by its interconnections with a totality of physics senses. This interrelationship between PageRank and the TMI illustrates the structure of the informational process of new physics.

Table 2: The characteristics of generations

Generation	Years	Total number of senses	Number of new senses	Percent of new senses
I	1989—1994	1363	1363	15.4
II	1995—2007	4071	3030	34.3
III	2008—2012	4952	2274	25.8
IV	2013—2018	7183	2156	24.4

5.3. Generations of physics senses

“Generations” offer a convenient method of classifying physics senses. We chose the doubling of the number of senses in 1995, 2008, and 2013 as the criterion for identifying generations. Following this way, we got four generations (see Table 2 and Fig. 2A). A physics sense belongs to a particular generation if its first occurrence happens in the period corresponding to that generation. Thus, all senses that occur in the years 1989–1994 are generation I senses. Generation II senses are those that occurred for the first time in doc-

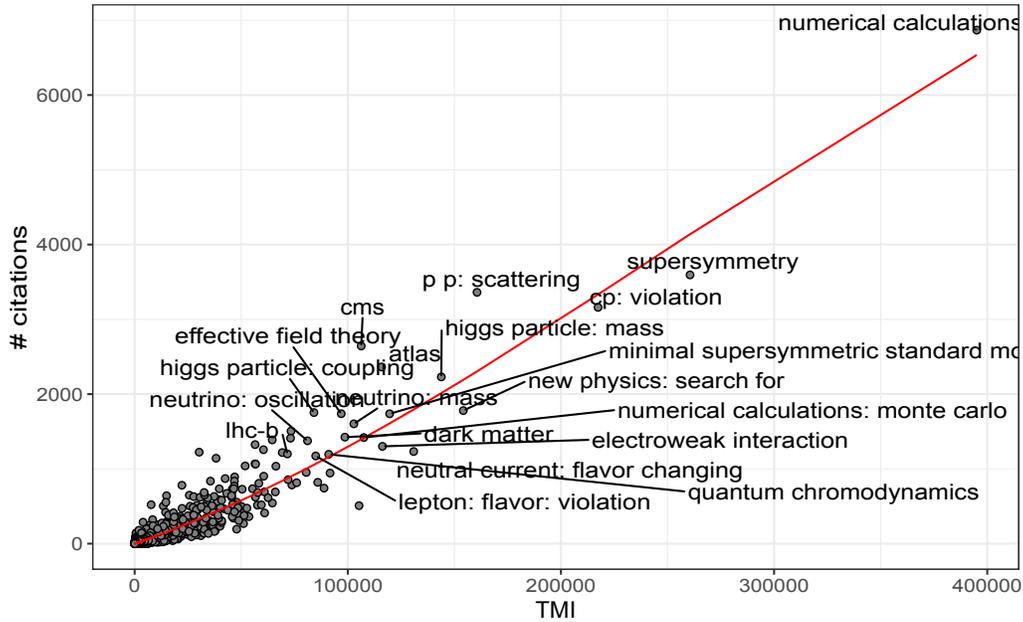


Figure 4: Scatterplot of # citations against the TMI. The equation of the regression curve of # citations is the following: $\# \text{ citations} = 1.547 \cdot 10^{-8} \cdot \mu^2 + 0.11 \cdot \mu - 3.278$ ($R^2 = 0.903$, $p = 0.000$).

uments published in 1995–2007, etc. Table 2 shows that the total number of senses increases from generation to generation, although new senses become less common over time. Of the total dataset, 15.4% are generation I, 34.3% generation II, 25.8% generation III, and 24.4% generation IV senses. Analysis of the generations helped to identify the “core” of physics senses—senses that occur in the discourse of new physics throughout all four generations. Such senses amount to 7.3% of the total dataset. They comprise the most frequently occurring physics senses with connections to many other senses. Their average dim is 471. For reference, an average dim for senses that occur for the duration of three generations is 204, whereas senses occurring for the period of one or two generations have an average dim of 34 and 74, respectively.

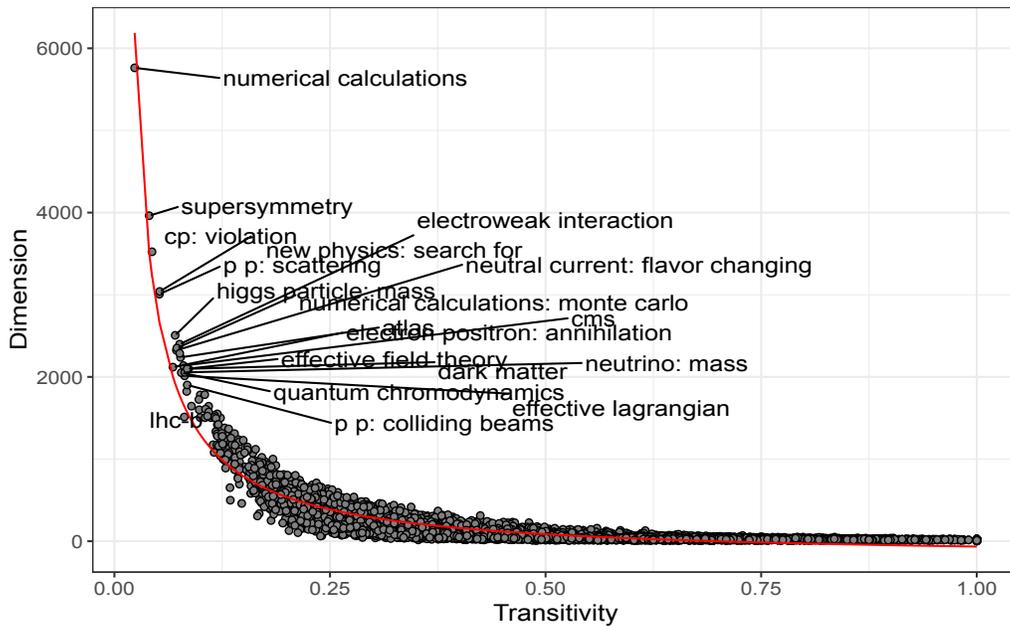


Figure 5: Scatterplot of the dimension against the transitivity. The equation of the regression curve of dim is the following: $\text{dim} = 150.418 \cdot \text{tr}^{-1} - 215.254$ ($R^2 = 0.895$, $p = 0.000$).

5.4. TMI, dimension, number of citations, and number of documents

Fig. 4 through Fig. 7 show interconnections between quantities essential for understanding the structure of physics senses.

- In Fig. 4, we see TMI as a quadratic function of the number of citations. This is the way in which the Fréchet extreme distribution manifests itself: maximums of different values belong to the same physics senses. Fig. 4 indicates that, for the orbit of the sense, its measure of intersection with other orbits directly depends on the citation rate of the sense. This is how the difference between the Fréchet distribution of the TMI and the Pareto distribution of the citation expresses itself: the quadratic dependence stems from the fact that the citation is distributed more unevenly than the TMI.
- Fig. 5 shows a dimension as an inverse function of transitivity. This is in line with results already published in the literature (Newman, 2018). This can be easily interpreted: the higher the value of the dimension, the bigger the number of pairs of physics senses whose orbits intersect with the orbit of the given sense and, therefore, the lower the transitivity value (see Eq. (7)). Since physics senses with a high dimension value are few, and those with a low one are many (the fact that determines the Fréchet extreme distribution of the dimension), the chart in Fig. 5 resembles a hyperbola.
- A TMI—dim linear graph in double logarithmic coordinates is shown in Fig. 6; and Fig. 7 has a TMI—# documents linear graph in double logarithmic coordinates. It is worth emphasizing that the straight line in double logarithmic coordinates indicates a power function. The level of TMI changes in proportion to the dimension to the power of 1.327 and the number of documents to the power of 1.059. This fact indicates heavy tails of TMI distributions. Moreover, the power functions $\text{TMI}(\text{dim})$ and $\text{TMI}(\# \text{ documents})$ are stable under changes in the scales of TMI, dim, and number of documents.
- Fig. 6 shows the influence of the topology of physics senses on their informational value in the new physics subfield. It appears that knowledge structure is a function of the dimension. To put it simply, the principal topological characteristic of a physics sense determines its position in the information structure of new physics. Notably, the TMI grows faster than the linear function of the dimension. It follows that physics senses with high TMI values play a disproportionately more significant role, lending a hierarchical character to the knowledge structure of the new physics subfield.

- Importantly, extreme values in all four charts belong to the same physics senses as listed in Table 1. These Top-30 physics senses determine, to a certain extent, the properties of the scholarly discourse of new physics.

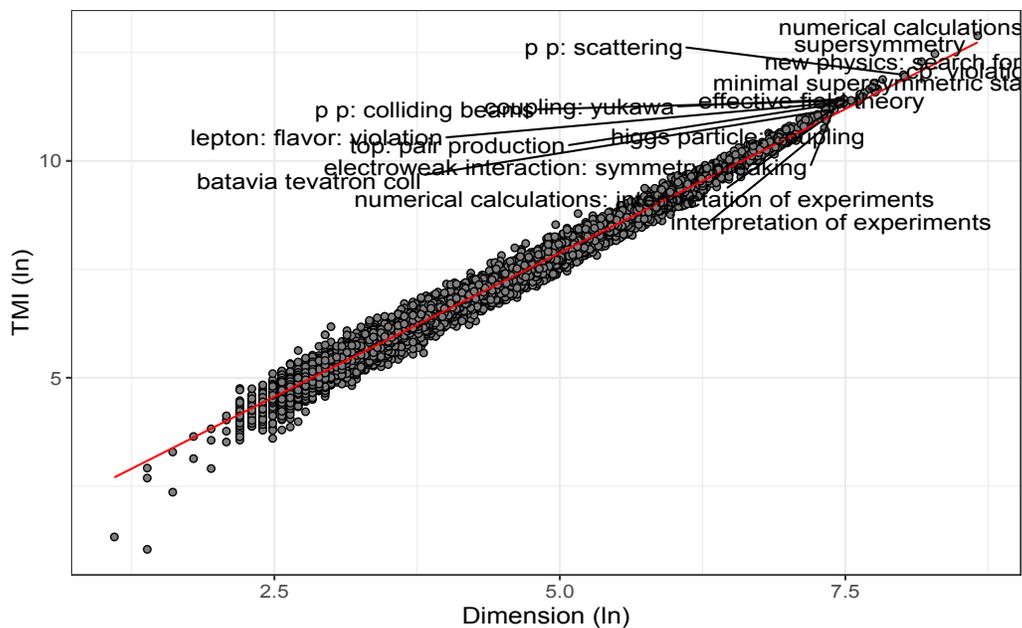


Figure 6: Scatterplot of the TMI (ln) against the dimension (ln). The equation of the regression curve of $\ln \mu$ is the following: $\ln \mu = 1.327 \cdot \ln \text{dim} + 1.245$ ($R^2 = 0.981$, $p = 0.000$).

5.5. Clusters of physics senses

We calculated the values of s_{AB} for all pairs of physics senses and carried out the procedure of sense clustering with the help of the function `cluster.walktrap`⁴. This function uses a random walk algorithm to find dense subgraphs in complex networks (see details in (Masuda et al., 2017)). The algorithm assumes that short random walks are in the same community. Based on this, the algorithm tries to find objects that form subgraphs within one network.

⁴See `igraph` software package.

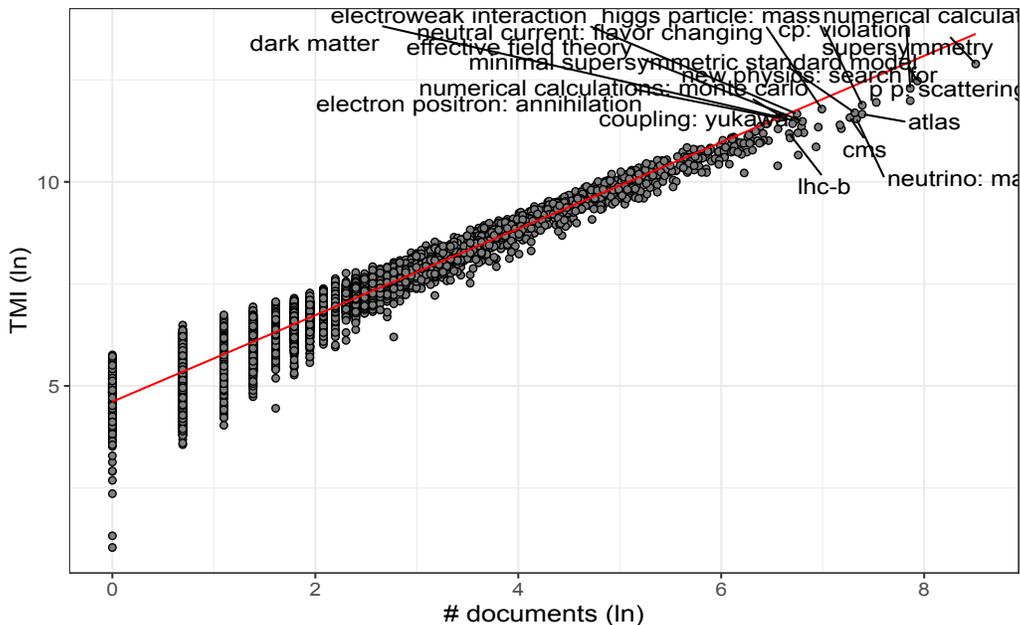


Figure 7: Scatterplot of the TMI (ln) against $\ln \#$ documents. The equation of the regression curve of $\ln \mu$ is the following: $\ln \mu = 1.059 \cdot \ln \# \text{ documents} + 4.615$ ($R^2 = 0.936$, $p = 0.000$).

Obtained were 12 clusters Fig. 8. We interpret the clusters on the basis of average values calculated for physics senses that form the clusters. Analysis of the behavior of the senses with regard to TMI and dim allows us to characterize their information interaction with other senses (see Fig. 8A). The closer senses get in terms of their information interaction, the bigger their chance of ending up in the same cluster and sharing other similarities. Fig. 8B shows the interdependence between the citation rate of the senses in a cluster and the number of documents where they occur. The figure shows that the first three clusters are comprised of senses with the highest citation rate and the most frequently occurring in the studied documents. The same three clusters contain senses with the most authority. It is these clusters that boast a concentration of the PageRank's Top-30 senses Table 1. Of the latter, the majority can be found in the second cluster: it includes ten of the Top-30 senses. For reference, clusters from 8 to 12 do not feature a single sense with a high PageRank value. Clusters 1 through 3 comprise senses with the highest TMI and dim values.

As becomes clear from Fig. 8A and Fig. 8B, the set of clusters splits into three groups. The first group contains basic physics senses. Due to their high dim and TMI values, they consolidate the scholarly discourse about new physics. The first group of clusters forms a key part of the HEP subfield that relates to new physics. The second can be interpreted as the mainstream of the discourse; it consists of the main physics senses. Main discourse events can occur within the group’s perimeter. Finally, the third group of clusters appears largely unstructured and chaotic. The most likely reason is that it is formed by physics senses that are either peripheral or emerging or declining (cf. (Hosseini et al., 2021)).

5.6. Complexity and entropy

The statistical complexity $c(A)$ and the normalized entropy $h_n(A)$ allow us to study the structural properties of a physics sense. The quantity $c(A)$ indicates the “state of disorder” such that senses with extremely ordered or disordered structures will receive lower scores, while senses with non-trivial structures—higher ones. At the same time, $h_n(A)$ measures the degree

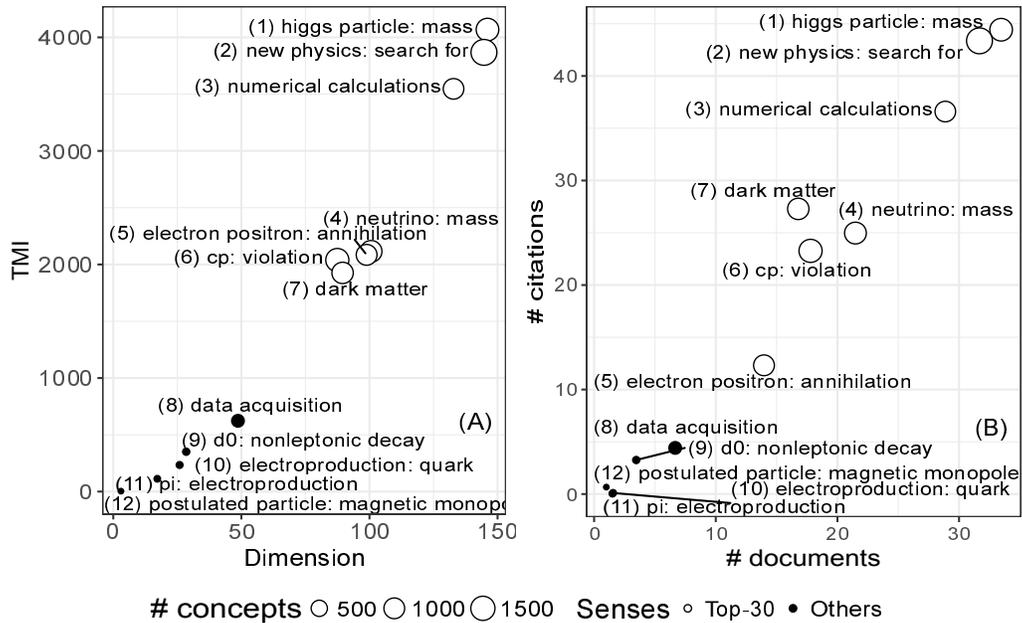


Figure 8: Space of physics sense clusters.

A. The clusters of physics senses in the TMI—the dimension plane.

B. The clusters of physics senses in the # citations—the # documents plane.

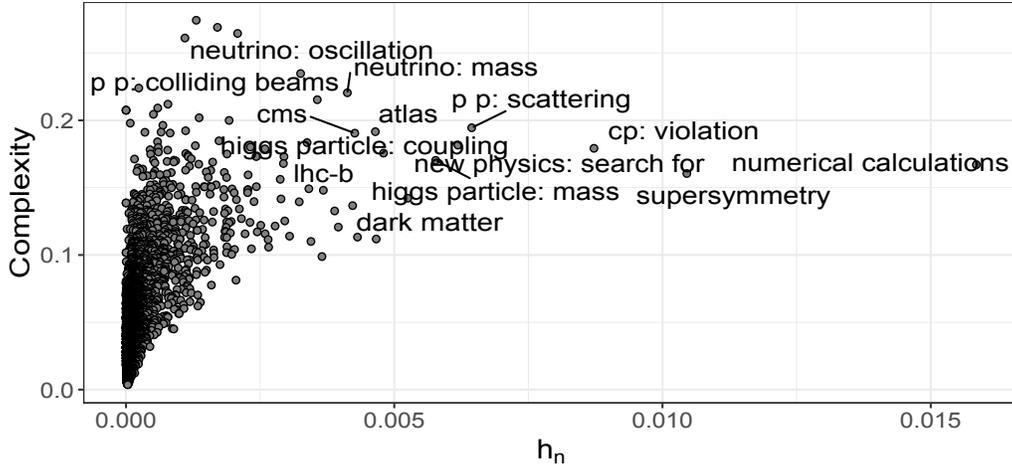


Figure 9: Scatterplot of the statistical complexity $c(A)$ against the normalized entropy $h_n(A)$.

of diversity of the support $X(A)$ of a sense and stochasticity. Thus, the statistical complexity $c(A)$ —the normalized entropy $h_n(A)$ plane (see Fig. 9) partially reflects the relationship of the notions of disorder/complexity versus order/simplicity with regard to physics senses. Although these quantities are loosely connected statistically (the Pearson correlation coefficient is equal to

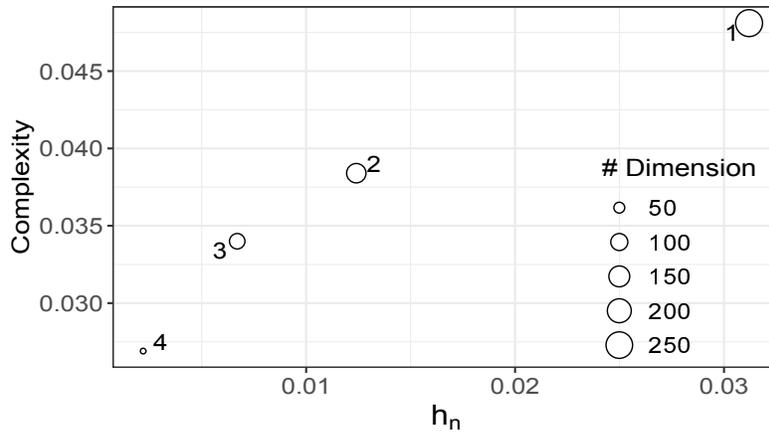


Figure 10: The generations of physics senses in the statistical complexity $c(A)$ —the normalized entropy $h_n(A)$ plane.

0.557 between $c(A)$ and $h_n(A)$), their relationship partially encodes structural features of each physics sense. For example, smaller values of $h_n(A)$ and larger values of $c(A)$ correspond to senses that consist of regular sets of recurring ordinal concepts. By contrast, senses that feature a significant diversity of concepts grouped into stochastic sets will generate larger values of $h_n(A)$ and smaller values of $c(A)$.

75% of physics senses have the value $c(A) \leq 0.045$ and $h_n(A) \leq 6.67 \cdot 10^{-5}$, i.e., they occur in that domain of the statistical complexity $c(A)$ —the normalized entropy $h_n(A)$ plane that can be identified through the paired notion of order/simplicity. As illustrated in Fig. 9, the Top-30 physics senses possess the highest values of normalized entropy: from 0.0159 for “numerical calculations” to 0,0019 for “interpretation of experiments”. However, the values of statistical complexity observed for the leaders of the rating are far from unprecedented: they range from 0.2348 (“neutrino oscillation”) to 0.0988 (“quantum chromodynamics”). We may argue that physics senses in the Top-30 rated positions are characterized by the highest level of disorder and a fairly high complexity level. It is these structural properties that, on the qualitative level, explain the position of the central senses in the scholarly discourse of new physics.

We studied the evolution of physics senses in the statistical complexity $c(A)$ —normalized entropy $h_n(A)$ plane. For each physics sense, we calculated the values of $c(A)$ and $h_n(A)$ by years and then averaged the results by generation. The results are presented for each generation in Fig. 10. We can see that each successive generation demonstrates lower values of statistical complexity and normalized entropy than their predecessors. This observation conforms to the overall picture as shown in Fig. 9. As we know, the vast majority of physics senses occur in the domain of low values of $c(A)$ and $h_n(A)$. At the same time, it follows from Fig. 2A that the number of physics senses was growing exponentially in time. Therefore, physics senses with lower values of $c(A)$ and $h_n(A)$ belong mainly to younger generations. In other words, generation I constitutes the semantic core of the scholarly discourse of new physics.

6. Conclusion

Our paper contributes to the literature by addressing the question of quantitative research on the new physics knowledge structure.

1. We develop a new method of keyword co-occurrence analysis that unites physics concepts of different documents and binds different documents with the same physics concepts. It is a more general approach than the network one since it offers a natural framework for quantifying knowledge structure of the new physics subfield. We consider topological structures that reveal relationships among concepts and call them physics senses. These senses represent complex systems of direct and indirect interactions between physics concepts.
2. The proposed phenomenological model has the advantage of providing a relevant quantitative evaluation of the distribution of physics senses without recourse to a relatively limited scientometric framework. In case of the new physics' subfield, the Fréchet extreme distribution indicates a high level of competition between physics concepts.
3. Physics senses form certain network clusters. The position of physics sense in the new physics' subfield is completely conditioned by the value of dimension and the belonging to a generation. Our findings indicate that the research interests of physicists are focused on a limited number of physics senses which, due to their higher dimension, act as a kind of glue that holds together a heterogeneous discourse of new physics.

However, some limitations need to be considered. First of all, our results are limited to the content of the INSPIRE database. However, given its state as one of the most comprehensive databases in high energy physics, we believe it provides the most possibly complete picture of new physics discourse. A second potential drawback of our study is that it depends on the BibClassify algorithm that extracts keywords from documents. Nevertheless, BibClassify draws upon the HEP Index maintained by the physics community. This index manifests the state of physics discourse as the physicists see it. Therefore, despite the above shortcomings, we believe our study gives a relevant view of the historical dynamics of new physics discourse.

That various potential studies can be conducted in the future using the proposed methodology. It would be interesting to carry out a comparison of several HEP fields: for example, new physics vs. the standard model. Also, assessment and improvement of the suggested methodology would benefit from the involvement of new data sources and other text analysis tools.

Appendix A. Framework

Recall that a partition (or a cluster expansion) of X is a closed disjoint cover of X , and each partition of a set X determines the equivalence relation on physics concepts: Two elements of X are said to be equivalent if they belong to the same cluster of the partition. In what follows, we shall denote by Ω a fixed cluster expansion of the finite T_1 topological space (X, τ) . There is defined a canonical projection $\pi: X \rightarrow \Omega$ of the discrete topological space (X, τ) onto the set Ω , i.e. the mapping π assign to the physics concept $x \in X$ the equivalence class $[x] \in \Omega$. We can now assign to Ω a quotient topology on Ω by calling a set $\omega \subset \Omega$ to open if and only if its preimage $\pi^{-1}(\omega)$ under the canonical projection is open in X (Singh, 2019). The collection of these sets $\{\omega\}$ is a topological structure ϖ in the quotient set Ω , and $\Omega = \cup_{\alpha \in X} \omega_\alpha$, $\omega_\alpha \cap \omega_\beta = \emptyset$ when $\alpha \notin \beta$. We stress that by construction, each cluster ω_α of the cluster expansion Ω corresponds uniquely to one physics concept x_α according to the configuration σ (see Eq (1)), and each cluster ω_α is indiscrete subspace of (X, τ) .

Here $|\cdot|: X \rightarrow \mathbb{N}$ denotes the counting measure, so $|X| = M$ is the number of physics concepts in the support X , $|\omega_\alpha| = m_\alpha$ is the number of elements in the cluster ω_α , and $\sum_\alpha m_\alpha = M$. We can introduce the corresponding non-strict partial order on (Ω, ϖ) : $\omega_1 \preceq \omega_2 \preceq \dots \preceq \omega_k$. Here, the non-strict partial order \preceq on the set $\{m_\alpha\}$ gives the non-strict partial order \preceq on (Ω, ϖ) — on the maximal T_0 -quotient of (X, τ) . Hence, a finite topological space (X, τ) is fully described by a non-strict partially ordered set (Ω, \preceq) whose clusters $\omega_1, \omega_2, \dots, \omega_k$ are equipped with multiplicities m_1, m_2, \dots, m_k .

The nonempty set $C_\Omega(\omega_\alpha) = \{\omega_\alpha \in \Omega: \omega_\alpha \preceq \omega_\beta\}$ is said to be a cone of ω_α over (Ω, \preceq) . We can look at this in another way: If $\omega_\beta \in C_\Omega(\omega_\alpha)$, then ω_α precedes or equal to ω_β , i.e. $\omega_\alpha \preceq \omega_\beta$. For elements $\omega_\alpha, \omega_\beta$ of (Ω, \preceq) , $C_\Omega(\omega_\alpha) = C_\Omega(\omega_\beta)$ implies $\omega_\alpha = \omega_\beta$, and $C_\Omega(\omega_1) \subset C_\Omega(\omega_2) \subset \dots \subset C_\Omega(\omega_k)$ implies $\omega_1 \preceq \omega_2 \preceq \dots \preceq \omega_k$.

The order topologies on the quotient space (Ω, ϖ) are in bijective correspondence with the relation \preceq . The collection of all cones $C_\Omega(\omega_\alpha)$ forms a minimal base for the topology ϖ (Singh, 2019). Further, the dimension $\dim(A)$ is the maximum of all non-negative integers k such that exists a chain $\omega_1 \preceq \omega_2 \preceq \dots \preceq \omega_k$ in Ω :

$$(\omega_\alpha \in \Omega)(k \in \mathbb{N}): \dim(A) = \max \{k: \exists \omega_1 \preceq \omega_2 \preceq \dots \preceq \omega_k\}. \quad (\text{A.1})$$

Since our key intuition is to treat the order topology ϖ as an internal property of Ω , our next task is to find an easy way of computing some topological quantity that characterizes (Ω, ϖ) . We can describe the difference between the physics sense A and physics sense B by mapping quotient space (Ω_A, ϖ_A) into quotient space (Ω_B, ϖ_B) . The topological structures on Ω_A and Ω_B are completely identical if there exists a homeomorphism of (Ω_A, ϖ_A) onto (Ω_B, ϖ_B) (in topological terms,

these spaces are homeomorphic: $(\Omega_A, \varpi_A) \cong (\Omega_B, \varpi_B)$). Let us remember that a homeomorphism is an invertible continuous map $\varphi: (\Omega_A, \varpi_A) \rightarrow (\Omega_B, \varpi_B)$ such that its inverse is continuous (Singh, 2019). Physics senses A and B are homeomorphic if and only if there exists a continuous bijective map between (Ω_A, ϖ_A) and (Ω_B, ϖ_B) that preserves the multiplicities m_1, m_2, \dots, m_k .

To give a precise meaning to the statements about distinctions or similarities of physics senses, we need to investigate the mappings of $\varphi: (\Omega_A, \varpi_A) \rightarrow (\Omega_B, \varpi_B)$. It is clear that, if φ is the homeomorphism, physics senses A and B are equivalent from topological point of view. Let $\nu = |\{\varphi\}|$ be the cardinality of the set of homeomorphisms. Suppose the equality $\nu_A = \nu_B$ holds, then $(\Omega_A, \varpi_A) \cong (\Omega_B, \varpi_B)$. In this case, φ is the a one-to-one correspondence between each class $\omega_\alpha \in \Omega$ and itself, and we get

$$\left(\sum_{\alpha=1}^{\dim(A)} m_\alpha = M \right) : \nu \{(\Omega_A, \varpi_A)\} = \prod_{\alpha=1}^{\dim(A)} m_\alpha^{m_\alpha}. \quad (\text{A.2})$$

Let $\mathbb{P}\{A\}$ be the ratio of the number of homeomorphisms ν_A to the number of all possible mappings from Ω to Ω

$$\mathbb{P}\{A\} = \frac{\nu_A}{M^M}. \quad (\text{A.3})$$

Thus, the number $\mathbb{P}\{A\}$ is the probability of a partition Ω_A of a support X_A with a fixed topological structure ϖ_A . It is obvious that $\mathbb{P}\{A\}$ is proportional to M . We now introduce a normalized logarithmic representation of $\mathbb{P}\{A\}$

$$h(A) = -\frac{1}{M} \ln \frac{\nu_A}{M^M} = -\left(\sum_{\alpha=1}^{\dim(A)} \frac{m_\alpha}{M} \ln \frac{m_\alpha}{M} \right). \quad (\text{A.4})$$

It is readily seen that the right-hand side (A.4) is nothing other than the entropy, as defined by Shannon.

Funding

The article was prepared in the framework of a research grant funded by the Ministry of Science and Higher Education of the Russian Federation (grant ID: 075-15-2020-928).

References

- Aleta, A., Meloni, S., Perra, N., Moreno, Y., 2019. Explore with caution: Mapping the evolution of scientific interest in physics. *EPJ Data Science* 8. doi:10.1140/epjds/s13688-019-0205-9.
- Battiston, F., Cencetti, G., Iacopini, I., Latora, V., Lucas, M., Patania, A., Young, J.G., Petri, G., 2020. Networks beyond pairwise interactions: Structure and dynamics. *Physics Reports* 874, 1–92. doi:10.1016/j.physrep.2020.05.004.
- Battiston, F., Musciotto, F., Wang, D., Barabási, A.L., Szell, M., Sinatra, R., 2019. Taking census of physics. *Nature Review Physics* 1, 89–97. doi:10.1038/s42254-018-0005-3.
- Behrouzi, S., Sarmoor, Z.S., Khosrow, H., Kavousi, K., 2020. Predicting scientific research trends based on link prediction in keyword networks. *Journal of Informetrics* 14, 101079. doi:10.1016/j.joi.2020.101079.
- Bingham, N.H., Goldie, C.M., Teugels, J.L., 1987. *Regular Variation*. Cambridge University Press, Cambridge. doi:10.1017/CB09780511721434.
- Börner, K., Milojević, S., 2019. *Springer Handbook of Science and Technology Indicators*. Springer International Publishing, Cham. chapter Science Forecasts: Modeling and Communicating Developments in Science, Technology, and Innovation. pp. 145–157. doi:10.1007/978-3-030-02511-3_6.
- Bornmann, L., Haunschild, R., Hug, S.E., 2018. Visualizing the context of citations referencing papers published by Eugene Garfield: a new type of keyword co-occurrence analysis. *Scientometrics* 114, 427–437. doi:10.1007/s11192-017-2591-8.
- Callon, M., Courtial, J., Laville, F., 1991. Co-word analysis as a tool for describing the network of interactions between basic and technological research: The case of polymer chemistry. *Scientometrics* 22, 155–205. doi:10.1007/BF02019280.
- Callon, M., Courtial, J.P., Turner, W.A., Bauin, S., 1983. From translations to problematic networks: An introduction to co-word analysis. *Social Science Information* 22, 191–235. doi:10.1177/053901883022002003.

- Cheng, Q., Wang, J., Lu, W., Huang, Y., Bu, Y., 2020. Keyword-citation-keyword network: a new perspective of discipline knowledge structure analysis. *Scientometrics* 124, 1923–1943. doi:10.1007/s11192-020-03576-5.
- Chinazzi, M., Gonçalves, B., Zhang, Q., Vespignani, A., 2019. Mapping the physics research space: A machine learning approach. *EPJ Data Science* 8. doi:10.1140/epjds/s13688-019-0210-z.
- Christianson, N.H., Sizemore, B.A., Bassett, D.S., 2020. Architecture and evolution of semantic networks in mathematics texts. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences* 476, 20190741. doi:10.1098/rspa.2019.0741.
- Cobo, M., López-Herrera, A., Herrera-Viedma, E., Herrera, F., 2011. An approach for detecting, quantifying, and visualizing the evolution of a research field: A practical application to the Fuzzy Sets Theory field. *Journal of Informetrics* 5, 146–166. doi:10.1016/j.joi.2010.10.002.
- Coccia, M., 2020. The evolution of scientific disciplines in applied sciences: dynamics and empirical properties of experimental physics. *Scientometrics* 124, 451–487. doi:10.1007/s11192-020-03464-y.
- Colavizza, G., Franceschet, M., 2016. Clustering citation histories in the Physical Review. *Journal of Informetrics* 10, 1037–1051. doi:10.1016/j.joi.2016.07.009.
- de Arruda, H.F., Comin, C.H., Costa, L.d.F., 2018. How integrated are theoretical and applied physics? *Scientometrics* 116, 1113–1121. doi:10.1007/s11192-018-2777-8.
- Deleuze, G., 1993. *Logic of Sense*. Columbia University Press, New York, NY. URL: <http://cup.columbia.edu/book/the-logic-of-sense/9780231059831>.
- Deville, P., Wang, D., Sinatra, R., Song, C., Blondel, V.D., Barabási, A.L., 2014. Career on the move: Geography, stratification, and scientific impact. *Scientific Reports* 4, 4770. doi:10.1038/srep04770.
- Fontana, M., Iori, M., Montobbio, F., Sinatra, R., 2020. New and atypical combinations: An assessment of novelty and interdisciplinarity. *Research Policy* 49, 104063. doi:10.1016/j.respol.2020.104063.

- Fortunato, S., Bergstrom, C.T., Börner, K., Evans, J.A., Helbing, D., Milojević, S., Petersen, A.M., Radicchi, F., Sinatra, R., Uzzi, B., Vespignani, A., Waltman, L., Wang, D., Barabási, A.L., 2018. Science of science. *Science* 359, eaao0185. doi:10.1126/science.aao0185.
- Foss, S., Korshunov, D., Zachary, S., 2013. *An Introduction to Heavy-Tailed and Subexponential Distributions*. Springer-Verlag, New York, NY; Heidelberg; Dordrecht; London. doi:10.1007/978-1-4614-7101-1.
- Ghosh, S.S., Hebbeker, T., Meyer, A., Pook, T., 2020. *General Model Independent Searches for Physics Beyond the Standard Model*. Springer International Publishing, Cham. doi:10.1007/978-3-030-53783-8.
- Hecker, B.L., 2017. Four decades of open science. *Nature Physics* 13, 523–525. doi:10.1038/nphys4160.
- Heiberger, R.H., Wiecek, O.J., 2016. Choosing collaboration partners. How scientific success in physics depends on network positions. arXiv [physics.soc-ph] , 1608.03251 URL: <http://arxiv.org/abs/1608.03251>.
- Herrera, M., Roberts, D.C., Gulbahce, N., 2010. Mapping the evolution of scientific fields. *PLoS ONE* 5, e10355. doi:10.1371/journal.pone.0010355.
- Hosseini, S., Baziyad, H., Norouzi, R., Jabbedari Khiabani, S., Gidòfalvi, G., Albadvi, A., Alimohammadi, A., Seyedabrishami, S., 2021. Mapping the intellectual structure of GIS-T field (2008–2019): a dynamic co-word analysis. *Scientometrics* 126, 2667–2688. doi:10.1007/s11192-020-03840-8.
- Jang, Y.S., Ko, Y.J., 2019. How latecomers catch up to leaders in high-energy physics as Big Science: transition from national system to international collaboration. *Scientometrics* 119, 437–480. doi:10.1007/s11192-019-03030-1.
- Jia, T., Wang, D., Szymanski, B.K., 2017. Quantifying patterns of research-interest evolution. *Nature Human Behaviour* 1, 0078. doi:10.1038/s41562-017-0078.
- Joslyn, C.A., Aksoy, S.G., Callahan, T.J., Hunter, L.E., Jefferson, B., Pragastis, B., Purvine, E., Tripodi, I.J., 2021. Hypernetwork science: From multidimensional networks to computational topology, in: Braha, D.,

- de Aguiar, M.A.M., Gershenson, C., Morales, A.J., Kaufman, L., Naumova, E.N., Minai, A.A., Bar-Yam, Y. (Eds.), Unifying Themes in Complex Systems X. ICCS 2020. Springer, Cham. Springer Proceedings in Complexity, pp. 377–392. doi:10.1007/978-3-030-67318-5_25.
- Karimi, F., Green, D., Matous, P., Varvarigos, M., Khalilpour, K.R., 2021. Network of networks: A bibliometric analysis. *Physica D: Nonlinear Phenomena* 421, 132889. doi:10.1016/j.physd.2021.132889.
- Katchanov, Y.L., Markova, Y.V., Shmatko, N.A., 2016. How physics works: scientific capital in the space of physics institutions. *Scientometrics* 108, 875–893. doi:10.1007/s11192-016-2005-3.
- Klement, K.C., 2001. *Frege and the Logic of Sense and Reference*. Routledge, New York, NY; London.
- Krenn, M., Zeilinger, A., 2020. Predicting research trends with semantic and neural networks with an application in quantum physics. *Proceedings of the National Academy of Sciences* 117, 1910–1916. doi:10.1073/pnas.1914370116.
- Langacker, P., 2017. *The Standard Model and Beyond*. CRC Press, Boca Raton, FL. doi:10.1201/b22175.
- Lebedev, A.V., 2005. Maxima of independent sums in the presence of heavy tails. *Theory of Probability & Its Applications* 49, 700–703. doi:10.1137/S0040585X9798138X.
- Leydesdorff, L., Nerges, A., 2017. Co-word maps and topic modeling: A comparison using small and medium-sized corpora ($n < 1,000$). *Journal of the Association for Information Science and Technology* 68, 1024–1035. doi:10.1002/asi.23740.
- Li, H., An, H., Wang, Y., Huang, J., Gao, X., 2016. Evolutionary features of academic articles co-keyword network and keywords co-occurrence network: Based on two-mode affiliation network. *Physica A: Statistical Mechanics and its Applications* 450, 657–669. doi:10.1016/j.physa.2016.01.017.

- Liu, W., Saganowski, S., Kazienko, P., Cheong, S.A., 2019. Predicting the evolution of physics research from a complex network perspective. *Entropy* 21, 1152. doi:10.3390/e21121152.
- López-Ruiz, R., Mancini, H.L., Calbet, X., 1995. A statistical measure of complexity. *Physics Letters A* 209, 321–326. doi:10.1016/0375-9601(95)00867-5.
- Lozano, S., Calzada-Infante, L., Adenso-Díaz, B., García, S., 2019. Complex network analysis of keywords co-occurrence in the recent efficiency analysis literature. *Scientometrics* 120, 609–629. doi:10.1007/s11192-019-03132-w.
- Martin, M.T., Plastino, A., Rosso, O.A., 2006. Generalized statistical complexity measures: Geometrical and analytical properties. *Physica A: Statistical Mechanics and its Applications* 369, 439–462. doi:10.1016/j.physa.2005.11.053.
- Masuda, N., Porter, M.A., Lambiotte, R., 2017. Random walks and diffusion on networks. *Physics Reports* 716–717, 1–58. doi:10.1016/j.physrep.2017.07.007.
- McKeown, K., Daume, H., Chaturvedi, S., Paparrizos, J., Thadani, K., Barrio, P., Biran, O., Bothe, S., Collins, M., Fleischmann, K.R., Gravano, L., Jha, R., King, B., McInerney, K., Moon, T., Neelakantan, A., O’Seaghdha, D., Radev, D., Templeton, C., Teufel, S., 2016. Predicting the impact of scientific concepts using full-text features. *Journal of the Association for Information Science and Technology* 67, 2684–2696. doi:10.1002/asi.23612.
- Milojević, S., 2015. Quantifying the cognitive extent of science. *Journal of Informetrics* 9, 962–973. doi:10.1016/j.joi.2015.10.005.
- Mingers, J., Leydesdorff, L., 2015. A review of theory and practice in scientometrics. *European Journal of Operational Research* 246, 1–19. doi:10.1016/j.ejor.2015.04.002.
- Newman, M., 2018. *Networks*. Oxford University Press, Oxford. doi:10.1093/oso/9780198805090.001.0001.

- Palchykov, V., Gemmetto, V., Boyarsky, A., Garlaschelli, D., 2016. Ground truth? Concept-based communities versus the external classification of physics manuscripts. *EPJ Data Science* 5. doi:10.1140/epjds/s13688-016-0090-4.
- Palchykov, V., Krasnytska, M., Mryglod, O., Holovatch, Y., 2021. Network of scientific concepts: Empirical analysis and modeling. *Advances in Complex Systems* 24, 24001. doi:10.1142/S0219525921400014.
- Palmucci, A., Liao, H., Napoletano, A., Zaccaria, A., 2020. Where is your field going? A machine learning approach to study the relative motion of the domains of physics. *PLoS ONE* 15, e0233997. doi:10.1371/journal.pone.0233997.
- Pan, R.K., Sinha, S., Kaski, K., Saramäki, J., 2012. The evolution of interdisciplinarity in physics research. *Scientific Reports* 2, 551. doi:10.1038/srep00551.
- Patania, A., Petri, G., Vaccarino, F., 2017. The shape of collaborations. *EPJ Data Science* 6, 18. doi:10.1140/epjds/s13688-017-0114-8.
- Perc, M., 2013. Self-organization of progress across the century of physics. *Scientific Reports* 3, 1720. doi:10.1038/srep01720.
- Petersen, A.M., 2018. Multiscale impact of researcher mobility. *Journal of The Royal Society Interface* 15, 20180580. doi:10.1098/rsif.2018.0580.
- Pluchino, A., Burgio, G., Rapisarda, A., Biondo, A.E., Pulvirenti, A., Ferro, A., Giorgino, T., 2019. Exploring the role of interdisciplinarity in physics: Success, talent and luck. *PLoS ONE* 14, e0218793. doi:10.1371/journal.pone.0218793.
- Radhakrishnan, S., Erbis, S., Isaacs, J.A., Kamarthi, S., 2017. Novel keyword co-occurrence network-based methods to foster systematic reviews of scientific literature. *PLoS ONE* 12, e0185771. doi:10.1371/journal.pone.0172778.
- Renoust, B., Claver, V., Baffier, J.F., 2017. Flows of Knowledge in Citation Networks, in: Cherifi, H., Gaito, S., Quattrociocchi, W., Sala, A. (Eds.), *Complex Networks & Their Applications V*, Springer, Cham. pp. 159–170. doi:10.1007/978-3-319-50901-3_13.

- Rosso, O.A., Larrondo, H.A., Martin, M.T., Plastino, A., Fuentes, M.A., 2007. Distinguishing noise from chaos. *Phys. Rev. Lett.* 99, 154102. doi:10.1103/PhysRevLett.99.154102.
- Salnikov, V., Cassese, D., Lambiotte, R., Jones, N.S., 2018. Co-occurrence simplicial complexes in mathematics: identifying the holes of knowledge. *Applied Network Science* 3, 37. doi:10.1007/s41109-018-0074-3.
- Silva, F.S.V., Schulz, P.A., Noyons, E.C.M., 2019. Co-authorship networks and research impact in large research facilities: Benchmarking internal reports and bibliometric databases. *Scientometrics* 118, 93–108. doi:10.1007/s11192-018-2967-4.
- Sinatra, R., Deville, P., Szell, M., Wang, D., Barabási, A.L., 2015. A century of physics. *Nature Physics* 11, 791–796. doi:10.1038/nphys3494.
- Singh, T.B., 2019. *Introduction to Topology*. Springer Singapore, Singapore. doi:10.1007/978-981-13-6954-4.
- Strumia, A., Torre, R., 2019. Biblioranking fundamental physics. *Journal of Informetrics* 13, 515–539. doi:10.1016/j.joi.2019.01.011.
- Sun, Y., Latora, V., 2020. The evolution of knowledge within and across fields in modern physics. *Scientific Reports* 10, 12097. doi:10.1038/s41598-020-68774-w.
- Tripodi, G., Chiaromonte, F., Lillo, F., 2020. Knowledge and social relatedness shape research portfolio diversification. *Scientific Reports* 10, 14232. doi:10.1038/s41598-020-71009-7.
- van Eck, N.J., Waltman, L., 2014. *Measuring Scholarly Impact: Methods and Practice*. Springer International Publishing, Cham. chapter Visualizing Bibliometric Networks. pp. 285–320. doi:10.1007/978-3-319-10377-8_13.
- van Raan, A., 2019. *Springer Handbook of Science and Technology Indicators*. Springer International Publishing, Cham. chapter Measuring Science: Basic Principles and Application of Advanced Bibliometrics. pp. 237–280. doi:10.1007/978-3-030-02511-3_10.

- Wang, X., Wang, H., Huang, H., 2021. Evolutionary exploration and comparative analysis of the research topic networks in information disciplines. *Scientometrics* 126, 4991–5017. doi:10.1007/s11192-021-03963-6.
- Wen, B., Horlings, E., van der Zouwen, M., van den Besselaar, P., 2017. Mapping science through bibliometric triangulation: An experimental approach applied to water research. *Journal of the Association for Information Science and Technology* 68, 724–738. doi:10.1002/asi.23696.
- Xing, Y., Zeng, A., Fan, Y., Di, Z., 2019. The strong non-linear effect in academic dropout. *Scientometrics* 120, 793–805. doi:10.1007/s11192-019-03135-7.
- Xu, S., Hao, L., An, X., Pang, H., Li, T., 2020a. Review on emerging research topics with key-route main path analysis. *Scientometrics* 122, 607–624. doi:10.1007/s11192-019-03288-5.
- Xu, S., Mariani, M.S., Lü, L., Medo, M., 2020b. Unbiased evaluation of ranking metrics reveals consistent performance in science and technology citation data. *Journal of Informetrics* 14, 101005. doi:10.1016/j.joi.2019.101005.
- Yi, S., Choi, J., 2012. The organization of scientific knowledge: the structural characteristics of keyword networks. *Scientometrics* 90, 1015–1026. doi:10.1007/s11192-011-0560-1.
- You, T., Yoon, J., Kwon, O., Jung, W., 2021. Tracing the evolution of physics with a keyword co-occurrence network. *Journal of the Korean Physical Society* 78, 236–243. doi:10.1007/s40042-020-00051-5.
- Zeng, A., Shen, Z., Zhou, J., Fan, Y., Di, Z., Wang, Y., Stanley, H.E., Havlin, S., 2019. Increasing trend of scientists to switch between topics. *Nature Communications* 10, 3439. doi:10.1038/s41467-019-11401-8.
- Zeng, A., Shen, Z., Zhou, J., Wu, J., Fan, Y., Wang, Y., Stanley, H.E., 2017. The science of science: From the perspective of complex systems. *Physics Reports* 714-715, 1–73. doi:10.1016/j.physrep.2017.10.001.
- Zhang, Q., Perra, N., Gonçalves, B., Ciulla, F., Vespignani, A., 2013. Characterizing scientific production and consumption in Physics. *Scientific Reports* 3, 1640. doi:10.1038/srep01640.

Zingg, C., Nanumyan, V., Schweitzer, F., 2020. Citations driven by social connections? A multi-layer representation of coauthorship networks. *Quantitative Science Studies* 1, 1493–1509. doi:10.1162/qss_a_00092.