Predicting research trends with semantic and neural networks with an application in quantum physics

Mario Krenn[a,b,c,d,e,1] **and Anton Zeilinger**[a,b,1]

[a]Faculty of Physics, Vienna Center for Quantum Science & Technology, University of Vienna, 1090 Vienna, Austria; [b]Institute for Quantum Optics and Quantum Information, Austrian Academy of Sciences, 1090 Vienna, Austria; [c]Department of Chemistry, University of Toronto, Toronto, ON M5S 3H6, Canada; [d]Department of Computer Science, University of Toronto, Toronto, ON M5T 3A1, Canada; and [e]Vector Institute for Artificial Intelligence, Toronto, ON M5G 1M1, Canada

Contributed by Anton Zeilinger, October 24, 2019 (sent for review August 19, 2019; reviewed by Ebrahim Karimi and Terry Rudolph)

The vast and growing number of publications in all disciplines of science cannot be comprehended by a single human researcher. As a consequence, researchers have to specialize in narrow subdisciplines, which makes it challenging to uncover scientific connections beyond the own field of research. Thus, access to structured knowledge from a large corpus of publications could help push the frontiers of science. Here, we demonstrate a method to build a semantic network from published scientific literature, which we call SemNet. We use SemNet to predict future trends in research and to inspire personalized and surprising seeds of ideas in science. We apply it in the discipline of quantum physics, which has seen an unprecedented growth of activity in recent years. In SemNet, scientific knowledge is represented as an evolving network using the content of 750,000 scientific papers published since 1919. The nodes of the network correspond to physical concepts, and links between two nodes are drawn when two concepts are concurrently studied in research articles. We identify influential and prize-winning research topics from the past inside SemNet, thus confirming that it stores useful semantic knowledge. We train a neural network using states of SemNet of the past to predict future developments in quantum physics and confirm high-quality predictions using historic data. Using network theoretical tools, we can suggest personalized, out-of-the-box ideas by identifying pairs of concepts, which have unique and extremal semantic network properties. Finally, we consider possible future developments and implications of our findings.

semantic network | machine learning | quantum physics | metascience | computer-inspired science

A computer algorithm with access to a large corpus of published scientific research could potentially make genuinely new contributions to science. With such a body of knowledge, the algorithm could derive new scientific insights that are unknown to human researchers and note contradictions within existing scientific knowledge (1, 2). This level of automation of science is more in the realm of science fiction than reality at present. However, algorithms with access to and the capability of extracting semantic knowledge from the scientific literature can be employed in manifold ways to assist scientists and thereby, augment scientific progress. As an example, the evaluation of whether an idea is novel or surprising depends crucially on already-existing knowledge. Thus, a computer algorithm with the capability to propose new useful ideas or potential avenues of research will necessarily require access to published scientific literature—which forms at least partially the body of human knowledge in a scientific field.

Knowledge can be portrayed using semantic networks that represent semantic relations between concepts in a network (3). Over the last few years, significant results have been obtained by automatically analyzing the large corpus of scientific literature (4–6), including the development of semantic networks in several scientific disciplines.

In biochemistry, a semantic network has been built using a well-defined list of molecule names (which correspond to the nodes of the network) and forming edges when two components coappear in the abstract of a scientific paper. The network was derived from millions of papers published over 30 y, and the authors identify a more efficient collective strategy to explore the knowledge network of biochemistry (7, 8). In ref. 9, a semantic network was created using 100,000 papers from astronomy, ecology, economy, and mathematics. The nodes represent ideas or concepts [generated through automated generation of key concepts in large bodies of texts (10)]. The authors used the network to draw connections between human innovation process and random walks. In the field of neuroscience, semantic networks have been used to map the landscape of the field (11, 12). Papers from the interdisciplinary journal PNAS have been used to investigate sociological properties, such as interdisciplinary research (13).

Here, we show how to build and use a semantic network for quantum physics, which we call SemNet. It is built from 750,000 scientific papers in physics published since 1919. In the network, we identify a number of historic award-winning concepts, indicating that SemNet carries useful semantic knowledge. The evolution of such a large network allows us to use an artificial neural network for predicting research concepts that scientists will investigate in the next 5 y. Finally, we demonstrate the power of SemNet to suggest personalized and unique directions for future research.

Significance

The corpus of scientific literature grows at an ever increasing speed. While this poses a severe challenge for human researchers, computer algorithms with access to a large body of knowledge could help make important contributions to science. Here, we demonstrate the development of a semantic network for quantum physics, denoted SemNet, using 750,000 scientific papers and knowledge from books and Wikipedia. We use it in conjunction with an artificial neural network for predicting future research trends. Individual scientists can use SemNet for suggesting and inspiring personalized, out-of-the-box ideas. Computer-inspired scientific ideas will play a significant role in accelerating scientific progress, and we hope that our work directly contributes to that important goal.

Author contributions: M.K. and A.Z. designed research; M.K. performed research; M.K. and A.Z. analyzed data; and M.K. wrote the paper.

Reviewers: E.K., University of Ottawa; and T.R., Imperial College, London.

The authors declare no competing interest.

Data deposition: The semantic network as well as the source code for deriving personalized predictions and suggestions is published at GitHub (https://github.com/MarioKrenn6240/SEMNET).

[1]To whom correspondence may be addressed. Email: mario.krenn@univie.ac.at or anton.zeilinger@univie.ac.at.

This article contains supporting information online at https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.1914370116/-/DCSupplemental.

First published January 14, 2020.

Our work differs in several aspects from previous semantic networks created from scientific literature. First, we use machine learning to draw conclusions from earlier states to SEMNET's future state, which enables us to make predictions about the future research trends of the discipline. Second, we use network theoretical tools and machine learning to identify pairs of concepts with exceptional network properties. Those concept combinations can be restricted to the research interest of a specific scientist. This ability allows us to not only predict but also, suggest uninvestigated concept pairs, which human scientists might not have identified because they are out of the own subfield but that have properties that indicate an exceptional relation. They could be a seed of a new, out-of-the-box idea. Third, we apply SEMNET to quantum physics, which has seen an enormous growth during the last decade due to the potential transformative technologies. The growth can be seen in the establishment of several high-quality journals for quantum research (such as *Quantum*, *npj Quantum Information*, and Institute of Physics's *Quantum Science and Technology*), multibillion dollar funding from governments, and strong involvement of private companies and startups worldwide. The growth rate leads to enormous increase in scientific results and publications, which are difficult to follow for individual researcher—thus, quantum physics is an ideal test bed for SEMNET.

Semantic Network of Quantum Physics

A semantic network, or knowledge network, represents relations between concepts in the form of a network. Now, we describe in more detail how the network is built, especially how the concept list is generated and how links are formed. A schematic illustration can be seen in Fig. 1; more details are in Fig. 2.

Creation of the Concept List. We generate the concept list via two independent methods. We use human-made lists of physical concepts. These concepts are compiled from the indices of 13 quantum physics books (which were available to us in a digital form) as well as titles of Wikipedia articles that are linked in a quantum physics category. This human-made collection contains ∼5,000 entries of physical concepts.

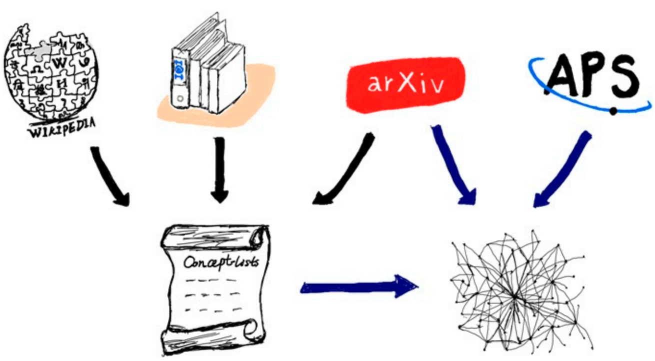

Fig. 1. Creating a semantic network for quantum physics (SEMNET). The nodes represent quantum physical concepts, and the edges (connections between nodes) indicate how frequently two concepts are investigated jointly in the scientific literature. The concept list is created using human-made lists (from Wikipedia categories and quantum physics books) and automatically generated lists using natural language processing tools on 100,000 quantum physics articles from the online preprint repository arXiv (this is indicated by black arrows). An edge between two concepts is drawn when both concepts appear in the abstract of a scientific paper (indicated by blue arrows). The scientific database consists of 750,000 physics papers: 100,000 from arXiv and 650,000 papers published by the American Physical Society (APS) since 1919. Illustrations courtesy of Xuemei Gu (Institute for Quantum Optics and Quantum Information, Vienna, Austria).

We extend the human-generated list with an automatically generated list of physical concepts. For this, we apply a natural language processing tool called Rapid Automatic Keyword Extraction (RAKE) (14) to the titles and abstracts of ∼100,000 articles published in quantum physics categories on the arXiv preprint server, which we chose to optimize the list for current research topics in quantum physics. RAKE is based on statistical text analysis and can automatically find relevant keywords in texts. We combine the human- and machine-generated lists of concepts and further optimize them to delete incorrectly identified concepts (which were introduced by imperfections of the statistical analysis of RAKE) and names of people (which are not concepts), merge synonyms, and normalize for the singular and plural of the same concept. Ultimately, this yields a list of 6,300 terms. As an example, five randomly chosen examples are "three-level system," "photon antibunching," "chemical shift," "neutron radiation," and "unconditionally secure quantum bit commitment." Each of these quantum physics concepts is a node in SEMNET.

Creation of the Network. To form connections between different quantum physics concepts, we use 100,000 articles of quantum physics categories on arXiv and the dataset of all 650,000 articles ever published by the American Physical Society (APS). We chose these two data sources because the APS database contains peer-reviewed physics papers from the last 100 y (allowing for investigation of long-term trends), while the arXiv database contains specific quantum physics papers, allowing for more precise coverage of the quantum physics research trends.

Whenever two concepts occur together in a title or an abstract of an article, we interpret that as a semantic connection between these concepts and add a unique link between the two corresponding nodes in the network. Relations between two concepts can take many forms. Concepts may be put together, for example, when a mathematical tool (such as "Schmidt rank") is used to investigate a specific quantum system (such as "vector beam" or "exciton polariton"), when insights from a specific technique (such as "lasing without inversion" or "rabi oscillation") lead to conclusions about another property (such as "transport property" or "atom transition frequency"), or when fundamental ideas (such as "quantum decoherence" or "quantum energy teleportation") are studied in the context of foundational experiments (such as "delayed choice experiment" or "Mermin inequality"). While this method clearly cannot represent all quantum physics knowledge, it represents elements of its semantic structure, which we demonstrate in what follows.

The resulting network SEMNET has 6,368 vertices with more than 1.7 million edges (drawn from more than 15 million concept pairs pulled from 750,000 physics articles) using physics articles from 1919 to December 2017.

We first use the evolution of the semantic network to identify impactful emerging fields of research in the past. We define emerging fields as either concepts or concept pairs that have grown significantly after they had been introduced or connected for the first time over periods of 5 y.

Fig. 3*A* shows the quantum physics topics that have grown the fastest (in terms of numbers of papers in which they have been mentioned) after their emergence from the years 1987 to 2017. Fig. 3*B* shows, for each year, which two-concept combinations have grown the fastest in the first 5 y after they had been first connected. In Fig. 3, many of the emerging fields clearly correspond to important discoveries, advances in understanding, and shifts of thought within quantum science research. One of the fastest growing concepts is qubit, which emerged in 1995 [first in April in a *Physical Review A* paper by Schumacher (15) and then, in a paper by Chuang and

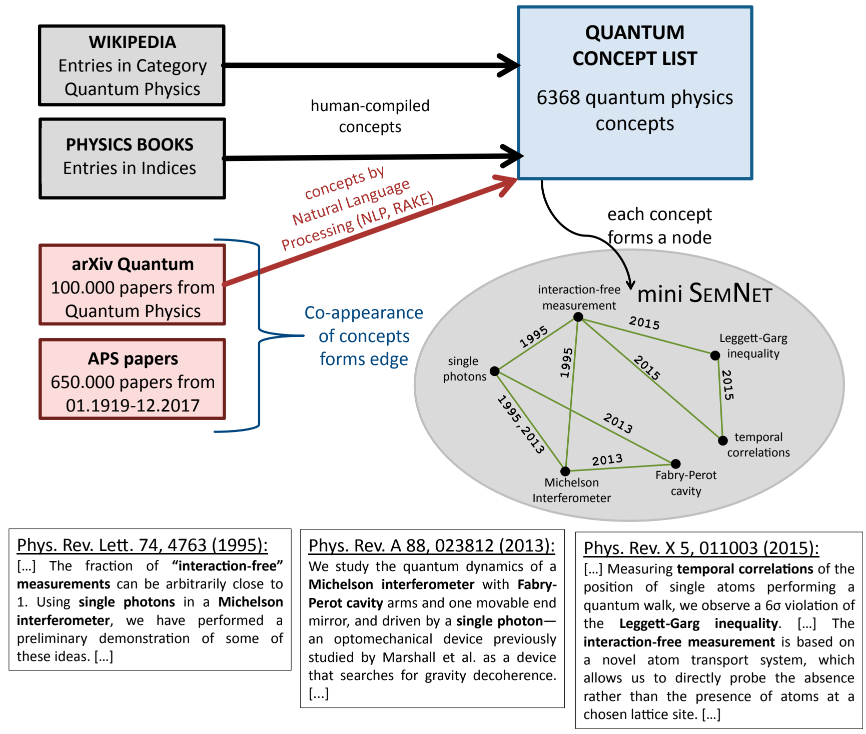

Fig. 2. Diagrammatic inner working of SEMNET. Human-generated concept lists (from Wikipedia and books) are combined with automatically generated lists (with natural language processing, using RAKE on 100,000 arXiv articles) to generate a list of quantum physics concepts. Each concept forms a link in a semantic network. The edges are formed when two concepts coappear in a title or abstract of any of the 750,000 papers (from arXiv and APS). A mini-version of SEMNET is shown, using parts of three articles from APS. Edges carry temporal information of their formation year, which leads to an evolution of the semantic network SEMNET over time.

Yamamoto (16) and arXiv preprints by Knill (17, 18)]. Qubits are the basic units of quantum information—generalizing classical bits to coherent quantum superpositions and connecting quantum mechanics and information science. The emergence of the qubit can be interpreted as the start of the discipline of quantum information science. Enormous growth is seen for topics connected to graphene starting in 2005, the discoverers of which were awarded the 2010 Nobel Prize in Physics. Interestingly, graphene itself was mentioned (in our data collection) already back in the early 1990s in *Physical Review B* papers (19–21) when it was not a strongly emergent concept itself. Strong growth in research into topological materials can be observed from ∼2008; the Nobel Prize in Physics was subsequently awarded in this area in 2016. The approach of Aaronson

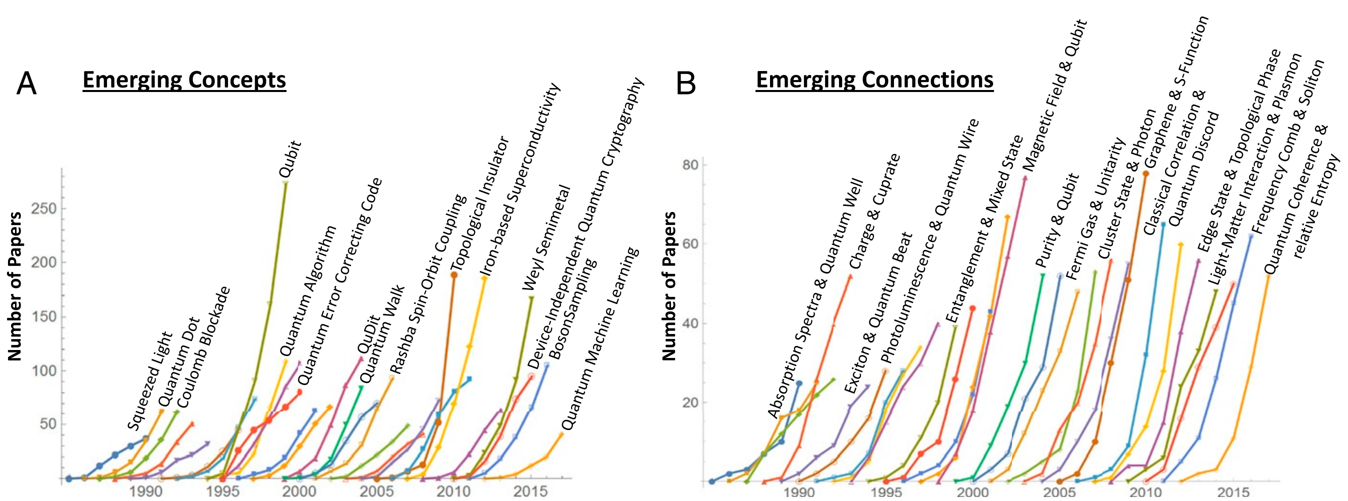

Fig. 3. The evolution of quantum physics research observed using SEMNET reflected in the change in number of articles that contain a concept or concept pair per year from 1987 to 2017. (*A*) Newly emerged concepts and their growth in popularity over a 5-y period after emergence. Shown are the strongest growing concepts of a 5-y period, which had not been mentioned before that period. (*B*) Newly connected pairs of concepts that become strongly influential in the scientific community in a 5-y period. Shown are the strongest growing connections of concept pairs that already existed before the connection was drawn, which had not been connected before that period. Many emergent concepts and connections can be related to important discoveries and understandings in quantum science.

and Arkhipov to achieving quantum supremacy (22) using linear photonic networks, termed BosonSampling (23), achieved considerable attention (with more than 600 citations since its introduction in 2011 and considerable experimental efforts in this direction). Since 2012, the application of machine learning to quantum physics has become a prominent and diverse topic of research that falls under the umbrella of quantum machine learning (recently summarized in two prominent reviews [24, 25] and also observable by the foundation of a novel high-quality journal for this topic, *Quantum Machine Intelligence*). These findings confirm that SEMNET contains useful semantic information.

Results

Past Quantum Physics Trends.

Predictive ability of the SEMNET. Having used SEMNET to study past quantum trends, we investigate its ability to provide projections of knowledge developments in the future. This essential question in network science is called link-prediction problem and asks which new link will be formed between unconnected vertices of the network in the future given the current state of the network (a detailed investigation of the link-prediction problem in network theory is in ref. 26). We apply this problem in the context of semantic networks, which are generated from published scientific literature. In this case of looking at the field of quantum physics, we ask which two concepts that have not yet been studied together might be investigated together in a scientific article over the next 5 y. To answer this question, we use an artificial neural network with four fully connected layers (two hidden layers). The structure of the neural network and its training are shown in Fig. 4. Its task is to rank all unconnected pairs of concepts (roughly 5% of all edges have been drawn by the end of 2017) starting with the pair that is most likely to be connected in 5 y up to the pair that most likely stays unconnected. Ultimately, we want to apply the neural network to the current SEMNET and predict the future trends. To validate its quality, we first input to the neural network past states of SEMNET (for example, containing data only up to 2002) and train it to predict new links by 2007. After the training, we apply this network to 2007 data and validate its quality for data of the year 2012 (which it has never seen before).

The semantic network is very large [consisting of $6{,}368 \times 6{,}368$ entries for each year, which are the number of possible connections between the 6,368 quantum physics concepts, compared with 28×28 pixels for the famous MNIST dataset of handwritten images and 256×256 pixels for ImageNet (27)] and involves combinatorial, graph-based information that is more structured than images (28). For that reason, it is an unsuitable direct input to the neural network. Instead, we compute semantic network properties for each pair of concepts. For each pair of concepts (c_i, c_j) that are unconnected in SEMNET, we calculate 17 network properties $p_{i,j} = (p^1_{i,j}, p^2_{i,j}, \ldots, p^{17}_{i,j})$ where $p^k_{i,j} \in \mathbb{R}$. Here, $p^1_{i,j}$ and $p^2_{i,j}$ are the degrees of concept c_i and c_j, and $p^3_{i,j}$ and $p^4_{i,j}$ are the numbers of papers in which they are mentioned. While these four properties are purely local, $p^5_{i,j}$ is the cosine similarity between the two concepts, which corresponds to the number of common neighbors. A cosine similarity of 1 indicates that the terms might be synonyms. The next nine properties indicate the number of paths with lengths of 2, 3, and 4 between the physics concepts in the current and previous 2 y. These properties allow us to draw conclusions from the evolution over time of various topics as tracked by SEMNET. The choice to use large path lengths as one of the properties is strengthened by a very recent observation that the paths of length 3 are crucial for link-prediction tasks in a network for protein interactions (29). Finally, the last three properties correspond to three different measures of distance between the two concepts. More details can be seen in *SI Appendix*.

We explain these properties on a concrete pair of concepts (Fig. 2): "interaction-free measurement" and "Leggett–Garg inequality." (We chose the example randomly from unconnected concepts that had been mentioned individually more than 30 times.) The concept c_{2526} represents interaction-free measurement, which is mentioned in 60 abstracts and has 135 connections to other concepts by 2012. The concept c_{2819} represents the Leggett–Garg inequality, which occurs in 33 abstracts and has 141 connections to other concepts by the end of 2012. These two concepts were not connected in SEMNET as of 2012; therefore, the 15th property, their network distance, is $p^{15}_{2526,2819} = 2$ (neighbors have a distance of 1; in other words, there is a direct path connecting them of length 1). In 2012, the two concepts have a cosine similarity $p^5_{2526,2819} = 0.228$, meaning that 22.8% of their neighbors are shared. Two years later, in 2014 an article on arXiv mentioned both of these concepts in the abstract, and the work was later published (31) and featured (32) in the high-impact journal *Physical Review X*, achieving $\sim$100 citations within 4 y. This example indicates that drawing first connections between concepts can lead to significant scientific insights.

The 17 properties for each unconnected concept pair in SEMNET are used by the neural network to estimate which pairs of quantum physics concepts are likely to be connected within 5 y and which are not.

To quantify the quality of the predictions, we employ a commonly used technique called the receiver operating characteristic (ROC) curve (30). For this, the neural network is used to classify unconnected nodes into two sets: one set that is connected after 5 y and one set that is nonconnected. Fig. 5 shows a significant ability to predict connections between pairs of topics—even though we restrict ourselves to pairs that share less than 20% of their neighbors (to prevent predictions of concepts that have similar meaning). This indicates that even research that draws new connections between concepts can be predicted with high quality.

Proposing Future Research Topics

Next, we attempt to use SEMNET and the artificial neural network to suggest potentially fruitful research directions in quantum physics. While it is interesting and useful to understand

PHYSICS

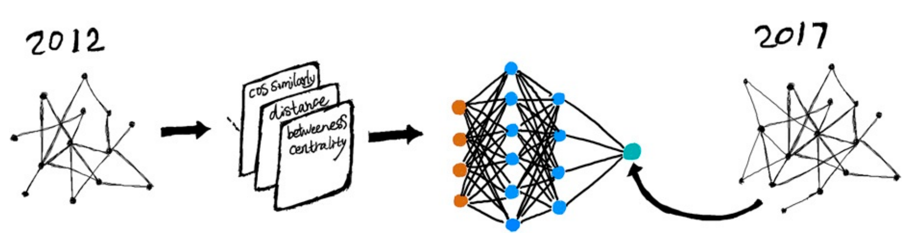

Fig. 4. Artificial neural network for predicting the future of quantum physics research using the evolution of the semantic network SEMNET. For each unconnected pair of concepts at a specific year, we derive a vector of 17 network properties (such as distance or cosine similarity). In the training phase, we input these network properties into an artificial neural network and ask the question of whether they will be connected 5 y later. SEMNET of 2017 is used for supervision. After training, we can apply the neural network to SEMNET of 2017 and ask what will have happened until the year 2022. Illustrations courtesy of Xuemei Gu (Institute for Quantum Optics and Quantum Information, Vienna, Austria).

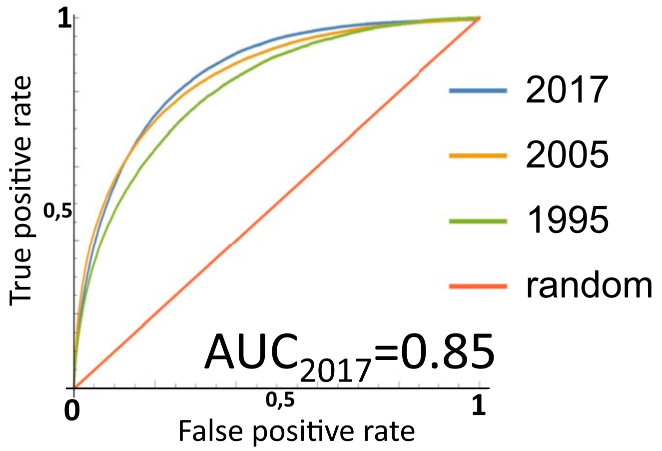

Fig. 5. Quantifying the prediction quality of the neural network regarding whether unconnected pairs will be connected within 5 y using an ROC curve. The *y* axis shows the true-positive (TP) rate (rate of pairs that have been correctly identified to be connected within 5 y). The *x* axis shows the false-positive (FP) rate of predictions—concept pairs that have falsely been predicted to be connected. We restrict ourselves to concept pairs that share less than 20% of their neighbors to prevent predictions of terms with similar semantical meaning. A perfect neural network would have TP = 1 while FP = 0. A network that classifies 50% of true instances correctly and misclassifies 10% of false instance as true would have TP = 0.5 and FP = 0.1. A random classifier is incorrect half the time and thus, lies along the diagonal. The area under the curve (AUC) for a perfect neural network is 1, while for random predictions, it is AUC = 0.5. The AUC can be interpreted as the probability that the neural network will rank a randomly chosen true instance higher than a randomly chosen negative instance (30). The ROC validation curves for 1995, 2005, and 2017 (trained with SemNet using data from only 1990, 2000, and 2012 and earlier, respectively) are consistently and significantly nonrandom, with AUC$_{2017}$ = 0.85. These results show that the neural network can learn to predict future research interests in quantum physics based on historical information to a high accuracy.

future trends, it potentially cannot by itself lead to surprising or out-of-the-box ideas (otherwise, they would not be predictable). Therefore, we extend our previous approach with network theoretic tools to identify concept pairs with exceptional network-theoretic properties. Furthermore, since science is conducted by (groups of) individual scientists, suggestions for proposed new research directions need to be personalized (otherwise, we would obtain suggestions for topics in which nobody is an expert in—which may be potentially interesting but limited in applicability).

How do we obtain suggestions for an individual scientist? What we find interesting and surprising strongly depends on what we already know. To gauge that, we need to investigate a given scientist's previously published body of research papers and extract a list of concepts (from the concept list generated before) that define that person's personal research agenda(s). We define key concepts as concepts investigated overproportionally often by the scientist compared with the relative frequency of that concept in all 750,000 papers. Each concept c_i in the papers authored by the scientist has a probability $p_{\text{scientist}}(c_i)$ that we calculate by the number of occurrences of the concept $N(c_i)$ divided by the sum of occurrences of all concepts, which is $p_{\text{scientist}}(c_i) = \frac{N(c_i)}{\sum_j N(c_j)}$. Each concept also has a probability of occurring in all 750,000 papers that we use written as $p_{\text{total}}(c_i) = \frac{M(c_i)}{\sum_j M(c_j)}$, where $M(c_i)$ is the number of occurrences of the concept c_i in all 750,000 articles. The ratio $r_{\text{scientist}}(c_i) = \frac{p_{\text{scientist}}(c_i)}{p_{\text{total}}(c_i)}$ indicates the research agenda of the scientist. A value of $r_{\text{scientist}}(c_i) > 1$ shows that the scientist investigates the concept c_i overproportionally often.

Our approach is to identify personalized suggestions of pairs of concepts that have never been connected. The concepts with $r_{\text{scientist}}(c_i) > 1$ value are paired with all of the other 6,368 concepts. This translates to a list of potentially 100,000s of possible topic pairs. For further usability, we introduce a way to sort the candidate suggestions. Suggestions can be sorted by identifying concept pairs with unique and unusual properties. For each pair of concepts, we have already calculated 18 different network properties: 17 properties that have been used by the neural network for generating predictions and the prediction value itself. Together, these properties define a multidimensional space in which the location of each concept pair depends on its network properties.

To identify unusual and unique concept pairs, we search for outliers in this high-dimensional space. An outlier indicates a pair of concepts that is uniquely located in the space and thus, has unique properties in the semantic SemNet network. We can visualize, for an anonymous example scientist, a three-dimensional projection of the high-dimensional space in Fig. 6. There, every dot corresponds to a concept pair, which is located according to its network properties. Outliers can be identified by the darkness of their color.

A few suggestions from SemNet for the example scientist are as follows. Some of the highest predicted pairs (from top 10) are "orbital angular momentum" and "magnetic skyrmion,"

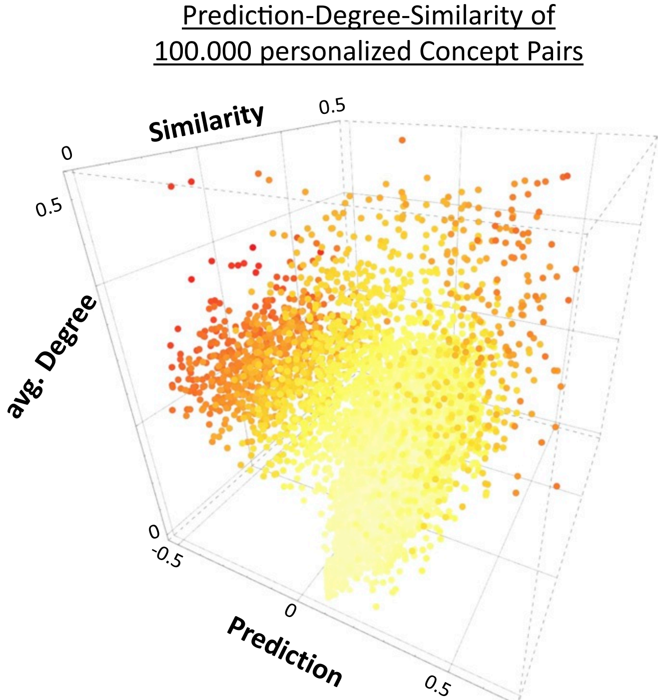

Fig. 6. Personalized prediction of topic pairs that could form future research directions for a given scientist. Each dot represents one unconnected pair of physical concepts. The concepts in use are filtered by a scientist's previous research agenda (in the text). The dot is placed in a three-dimensional space, which is proscribed by the properties of SemNet and the predictions of the neural network. One axis is the neural network predictions of whether two unconnected points will be connected in 2022 (the prediction −0.5 stands for very unlikely, and the prediction 0.5 is very likely). The *y* axis represents the average (normalized) degree of the pair (the concept with the highest degree in the complete network has a degree of 1). The *z* axis is the cos similarity, which is the ratio of shared neighbors in the networks of the two concepts. The color of the dots represents the distance from the most common, average point in this space—darker dots are farther away from the average. Outliers represent pairs of concepts with a unique network property, which make them ideal candidate suggestions.

"spin orbit coupling" and "quantum sensing," or "dicke model" and "cloning." For highly predicted, uncommon pairs (cosine similarity <0.03; from top 10): "topos theory" and "cyclic operation," "critical exponent" and "reed muller code," and "quantum key distribution" and "adhm construction." Unrestricted concept lists are (normalized concept degree <0.1; from top 10) "atom cavity system" and "mode volume," "entanglement of formation" and "multiqubit state," and "neutrino oscillation" and "dark photon." More examples are in *SI Appendix*.

Outlook

Machine Learning. Graph-based machine learning models, which have been studied in recent years, could improve prediction qualities in the link-prediction task (28, 33, 34). Furthermore, as SEMNET represents a time evolution of quantum physics' semantic network, applying efficient tools for handling time-dependent data, such as a long short-term memory (35), might further significantly improve the prediction quality. Application of techniques from machine translation could be beneficial to introduce multiple classes of connections within semantic networks (36). Additionally, combining our approach with unsupervised embedding of scientific literature, as shown in ref. 37, could lead to interesting, dynamic networks.

Network Theory and Science of Science. Currently, SEMNET represents connections between concepts that appear in the scientific literature. This is of course a vast simplification of scientific knowledge, as concepts in natural languages can have manifold relations (38). An extension could employ more complex structures for knowledge representation, such as hypergraphs (39). The concept list, which represents the nodes of SEMNET, can be improved by various different, sophisticated ways for generating lists of concepts and categories (10, 40). The extension to combinations of more than pairs of concepts will lead to more complex knowledge representations. Furthermore, it would be insightful to fold into the semantic network numbers of article citations, which are, at least in the field of science, frequently used as a proxy for scientific impact (41–43). This may enable the prediction of future research directions to be made taking into consideration the highest potential impact, potentially accelerating the evolution of individual scientific knowledge (44, 45).

Surprisingness. In this work, we place pairs of concepts in an abstract high-dimensional space and identify outliers that have unique and potentially valuable properties. It would be interesting to apply more and different measures of surprisingness. An interesting example is the information-based Bayesian surprise function, which has been introduced in the context of human attention (46) and successfully applied to the subfield of computational creativity (47, 48). In order to achieve further progress, it would be important to further explore and genuinely understand what human scientists consider as surprising and creative.

Discussion

We show how to create a semantic network in the field of quantum physics, demonstrate its usage to predict future trends in the field, and show how it can be used to suggest pairs of concepts, which are not yet investigated jointly but have distinct network properties. We show how to filter the suggestions for the research agendas of an individual scientist. The approach presented here is independent of the discipline of science. As such, it can be applied to other fields of research.

This can be interpreted as one potential road toward computer-inspired science in the following sense. We imagine cases (which we believe are possible) where SEMNET produces seeds or inspiration of unusual ideas or directions of thoughts that a researcher alone might not have thought of. The subsequent successful interpretation and scientific execution of the suggestions fully remain the task of a creative, human scientist.

ACKNOWLEDGMENTS. M.K. thanks James A. Evans and Sasha Belikov for exciting discussions of metaknowledge research and automation of science and Jacob G. Foster for a short but influential conversation at the International Symposium on Science of Science 2016. Furthermore, we acknowledge Nora Tischler, Armin Hochrainer, Robert Fickler, Radek Lapkiewicz, Manuel Erhard, and Philip Haslinger for many interesting discussions on related topics. We also thank the APS for providing access to the database of all published articles in APS journals. This work was supported by Austrian Academy of Sciences, University of Vienna Project Quantum Experiments at Space Scale and Austrian Science Fund Grant SFB F40 (Foundations and Applications of Quantum Science) and Erwin Schrödinger Fellowship J4309.

PHYSICS

1. J. A. Evans, A. Rzhetsky, Advancing science through mining libraries, ontologies, and communities. *J. Biol. Chem.* **286**, 23659–23666 (2011).
2. J. You, Darpa sets out to automate research. *Science* **347**, 465 (2015).
3. F. Lehmann, *Semantic Networks in Artificial Intelligence* (Elsevier Science Inc., 1992).
4. J. A. Evans, J. G. Foster, Metaknowledge. *Science* **331**, 721–725 (2011).
5. A. Zeng *et al.*, The science of science: From the perspective of complex systems. *Phys. Rep.* **714**, 1–73 (2017).
6. S. Fortunato *et al.*, Science of science. *Science* **359**, eaao0185 (2018).
7. J. G. Foster, A. Rzhetsky, J. A. Evans, Tradition and innovation in scientists' research strategies. *Am. Sociol. Rev.* **80**, 875–908 (2015).
8. A. Rzhetsky, J. G. Foster, I. T. Foster, J. A. Evans, Choosing experiments to accelerate collective discovery. *Proc. Natl. Acad. Sci. U.S.A.* **112**, 14569–14574 (2015).
9. I. Iacopini, S. Milojević, V. Latora, Network dynamics of innovation processes. *Phys. Rev. Lett.* **120**, 048301 (2018).
10. S. Milojević, Quantifying the cognitive extent of science. *J. Informetrics* **9**, 962–973 (2015).
11. E. Beam, L. G. Appelbaum, J. Jack, J. Moody, S. A. Huettel, Mapping the semantic structure of cognitive neuroscience. *J. Cogn. Neurosci.* **26**, 1949–1965 (2014).
12. J. D. Dworkin, R. T. Shinohara, D. S. Bassett, The landscape of neuroimage-ing research. *Neuroimage* **183**, 872–883 (2018).
13. J. D. Dworkin, R. T. Shinohara, D. S. Bassett, The emergent integrated network structure of scientific research. *PLoS One* **14**, e0216146 (2019).
14. S. Rose, D. Engel, N. Cramer, W. Cowley, "Automatic keyword extraction from individual documents" in *Text Mining: Applications and Theory*, M. W. Berry, J. Kogan, Eds. (John Wiley & Sons, Ltd, Chichester, UK, 2010), pp. 1–20.
15. B. Schumacher, Quantum coding. *Phys. Rev. A* **51**, 2738–2747 (1995).
16. I. L. Chuang, Y. Yamamoto, Simple quantum computer. *Phys. Rev. A* **52**, 3489–3496 (1995).
17. E. Knill, Approximation by quantum circuits. https://arxiv.org/abs/quant-ph/9508006 (8 August 1995).
18. E. Knill, Bounds for approximation in total variation distance by quantum circuits. https://arxiv.org/abs/quant-ph/9508007 (8 August 1995).
19. V. Bayot *et al.*, Two-dimensional weak localization in partially graphitic carbons. *Phys. Rev. B* **41**, 11770–11779 (1990).
20. S. Di Vittorio, M. Dresselhaus, M. Endo, T. Nakajima, Magnetic-field dependence of the hole-hole interaction in fluorine-intercalated graphite fibers. *Phys. Rev. B* **43**, 1313–1321 (1991).
21. R. Moreh, N. Shnieg, H. Zabel, Effective and debye temperatures of alkali-metal atoms in graphite intercalation compounds. *Phys. Rev. B* **44**, 1311–1317 (1991).
22. A. W. Harrow, A. Montanaro, Quantum computational supremacy. *Nature* **549**, 203–209 (2017).
23. S. Aaronson, A. Arkhipov, "The computational complexity of linear optics" in *Proceedings of the Forty-Third Annual ACM Symposium on Theory of Computing* (ACM, New York, NY, 2011), pp. 333–342.
24. J. Biamonte *et al.*, Quantum machine learning. *Nature* **549**, 195–202 (2017).
25. V. Dunjko, H. J. Briegel, Machine learning & artificial intelligence in the quantum domain: A review of recent progress. *Rep. Prog. Phys.* **81**, 074001 (2018).
26. D. Liben-Nowell, J. Kleinberg, The link-prediction problem for social networks. *J. Am. Soc. Inf. Sci. Technol.* **58**, 1019–1031 (2007).
27. Y. LeCun, Y. Bengio, G. Hinton, Deep learning. *Nature* **521**, 436–444 (2015).
28. Z. Wu *et al.*, A comprehensive survey on graph neural networks. https://arxiv.org/abs/1901.00596 (3 January 2019).
29. I. A. Kovács *et al.*, Network-based prediction of protein interactions. *Nat. Commun.* **10**, 1240 (2019).
30. T. Fawcett, Roc graphs: Notes and practical considerations for researchers. *Mach. Learn.* **31**, 1–38 (2004).
31. C. Robens, W. Alt, D. Meschede, C. Emary, A. Alberti, Ideal negative measurements in quantum walks disprove theories based on classical trajectories. *Phys. Rev. X* **5**, 011003 (2015).
32. G. C. Knee, Do quantum superpositions have a size limit? *Physics* **8**, 6 (2015).
33. Y. Li, D. Tarlow, M. Brockschmidt, R. Zemel, Gated graph sequence neural networks. https://arxiv.org/abs/1511.05493 (17 November 2015).

34. M. Niepert, M. Ahmed, K. Kutzkov, "Learning convolutional neural networks for graphs" in *Proceedings of the International Conference on Machine Learning* (ACM, New York, NY, 2016), pp. 2014–2023.
35. S. Hochreiter, J. Schmidhuber, Long short-term memory. *Neural Comput.* **9**, 1735–1780 (1997).
36. A. Vaswani *et al.*, "Attention is all you need" in *Advances in Neural Information Processing Systems 30*, I. Guyon *et al.*, Eds. (Neural Information Processing Systems Foundation, 2017), pp. 5998–6008.
37. V. Tshitoyan *et al.*, Unsupervised word embeddings capture latent knowledge from materials science literature. *Nature* **571**, 95–98 (2019).
38. H. Helbig, *Knowledge Representation and the Semantics of Natural Language* (Springer, 2006).
39. F. Shi, J. G. Foster, J. A. Evans, Weaving the fabric of science: Dynamic network models of science's unfolding structure. *Soc. Netw.* **43**, 73–85 (2015).
40. S. Sreenivasan, Quantitative analysis of the evolution of novelty in cinema through crowdsourced keywords. *Sci. Rep.* **3**, 2758 (2013).
41. B. Uzzi, S. Mukherjee, M. Stringer, B. Jones, Atypical combinations and scientific impact. *Science* **342**, 468–472 (2013).
42. T. Martin, B. Ball, B. Karrer, M. Newman, Coauthorship and citation patterns in the physical review. *Phys. Rev. E* **88**, 012814 (2013).
43. T. Kuhn, M. Perc, D. Helbing, Inheritance patterns in citation networks reveal scientific memes. *Phys. Rev. X* **4**, 041036 (2014).
44. R. Sinatra, D. Wang, P. Deville, C. Song, A. L. Barabási, Quantifying the evolution of individual scientific impact. *Science* **354**, aaf5239 (2016).
45. A. L. Barabási, *The Formula: The Universal Laws of Success* (Hachette, 2018).
46. L. Itti, P. Baldi, Bayesian surprise attracts human attention. *Adv. Neural Inf. Process. Syst.* **18**, 547–554 (2006).
47. L. R. Varshney *et al.*, A big data approach to computational creativity. https://arxiv.org/pdf/1311.1213.pdf (5 November 2013).
48. F. Pinel, L. R. Varshney, D. Bhattacharjya, "A culinary computational creativity system" in *Computational Creativity Research: Towards Creative Machines*, T. R. Besold, M. Schorlemmer, A. Smaill, Eds. (Springer, 2015), pp. 327–346.