Predicting Research Trends with Semantic and Neural Networks with an application in Quantum Physics

Mario Krenn[1, 2, 3, 4] and Anton Zeilinger[1, 2]

[1] *Vienna Center for Quantum Science & Technology (VCQ), Faculty of Physics, University of Vienna, Austria.*
[2] *Institute for Quantum Optics and Quantum Information (IQOQI), Austrian Academy of Sciences, Vienna, Austria.*
[3] *Department of Chemistry & Computer Science, University of Toronto, Canada.*
[4] *Vector Institute for Artificial Intelligence, Toronto, Canada.*

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 humans and note contradictions within existing scientific knowledge. 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 [1, 2]. As an example, the evaluation of whether an idea is novel or surprising depends crucially on already-existing knowledge. Thus a 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.

Here we demonstrate a method to build a semantic network from published scientific literature, which we call SemNet [3]. We use SemNet to predict future trends in research and to inspire new, personalized and surprising seeds of ideas in science, specifically in quantum physics.

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, which are derived from Wikipedia, quantum physics books and natural language processing techniques. Links between two nodes are drawn when two physical concepts are concurrently studied in research articles.

We identify influential and prize-winning research topics from the past inside SemNet thus confirm that it stores useful semantic knowledge. We train a deep neural network using states of SemNet of the past, to predict future developments in quantum physics research, and confirm high quality predictions using historic data. With the neural network and theoretical network tools we are able to suggest new, personalized, out-of-the-box ideas, by identifying pairs of concepts which have unique and extremal semantic network properties.

Our approach can be interpreted as one potential road towards computer-inspired science, in the following sense: We imagine cases (which we believe is possible) where SemNet produces seeds or inspirations 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 remains the task of a creative, human scientist.

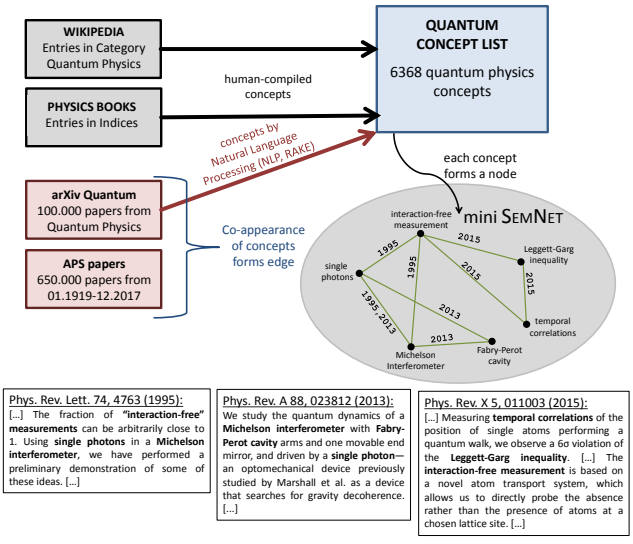

Figure 1. **Construction of SemNet**, using data from physics papers and books and Wikipedia.

[1] J.A. Evans and J.G. Foster, Metaknowledge. *Science* **331**, 721 (2011).
[2] S. Fortunato, et.al, Science of Science. *Science* **359**, 1007 (2018).
[3] M. Krenn and A. Zeilinger, Predicting Research Trends with Semantic and Neural Networks with an application in Quantum Physics. *Proceedings of the National Academy of Sciences* **117**, 1910–1916 (2020).