Title: Inspiring new research directions with AI

Speakers: Mario Krenn

Series: Machine Learning Initiative

Date: December 10, 2021 - 11:00 AM

URL: https://pirsa.org/21120001

Abstract: 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.

In my talk, I will present a possible solution: I demonstrate the development of a semantic network for quantum physics (SemNet), using 750,000 scientific papers and knowledge from books and Wikipedia. I use it in conjunction with an artificial neural network for predicting future research directions. Finally, I show first indications how individual scientists can use SemNet for suggesting and inspiring personalized, out-of-the-box ideas.

I believe that computer-inspired scientific ideas will play a significant role in accelerating scientific progress, and am looking forward hearing your thoughts and ideas about this crucial question.

References

[1] Mario Krenn, Anton Zeilinger, Predicting research trends with semantic and neural networks with an application in quantum physics, PNAS 117(4) 1910-1916 (2020).
[2] IEEE BigData 2021 competition: Science4Cast: https://github.com/iarai/science4cast

Zoom Link: https://pitp.zoom.us/j/92240839439?pwd=LytUTHlMWE9ycjlsUXJkdHRta2c1UT09

Perimeter ML seminar, 10.12.2021

Predicting and inspiring new research directions with AI

Krenn, Zeilinger, *PNAS* **117**(4), 1910 (2020).

Mario Krenn

Artificial Scientist Lab, Theory Division

@mariokrenn6240

http://mariokrenn.wordpress.com/

MAX PLANCK INSTITUTE
FOR THE SCIENCE OF LIGHT

On Computer-Inspired Science

Goal:

Make a computer program which
inspires surprising and interesting ideas.

What does it mean that something is ***surprising*** *or* ***interesting***?

Very subjective; depends on one‘s knowledge.

How to get this knowledge into a computer??

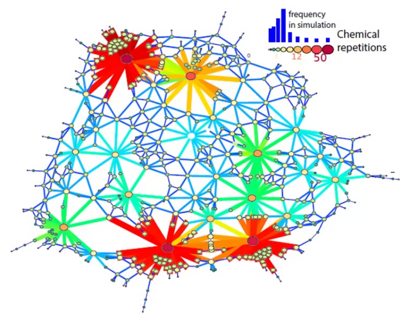
On Computer-Inspired Science

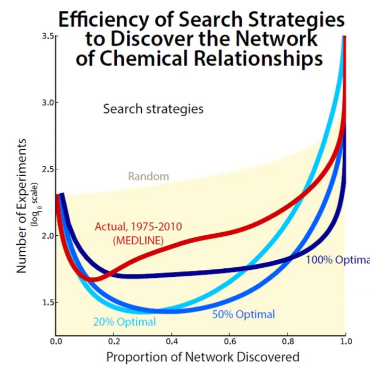
Choosing experiments to accelerate collective discovery

Andrey Rzhetsky[a,b,c,1], Jacob G. Foster[d], Ian T. Foster[b,e], and James A. Evans[b,f,1]

Edited by Yu Xie, University of Michigan, Ann Arbor, MI, and approved September 8, 2015 (received for review May 18, 2015)

A scientist's choice of research problem affects his or her personal career trajectory. Scientists' combined choices affect the direction and efficiency of scientific discovery as a whole. In this paper, we infer preferences that shape problem selection from patterns of published findings and then quantify their efficiency. We represent research problems as links between scientific entities in a knowledge network. We then build a generative model of discovery informed by qualitative research on scientific problem selection. We map salient features from this literature to key network properties: an entity's importance corresponds to its degree centrality, and a problem's difficulty corresponds to the network distance it spans. Drawing on millions of papers and patents published over 30 years, we use this model to infer the typical research strategy used to explore chemical relationships in biomedicine. This strategy generates conservative research choices focused on building up knowledge around important molecules. These choices become more conservative over time. The observed strategy is efficient for initial exploration of the network and supports scientific careers that require steady output, but is inefficient for science as a whole. Through supercomputer experiments on a sample of the network, we study thousands of alternatives and identify strategies much more efficient at exploring mature knowledge networks. We find that increased risk-taking and the publication of experimental failures would substantially improve the speed of discovery. We consider institutional shifts in grant making, evaluation, and publication that would help realize these efficiencies.

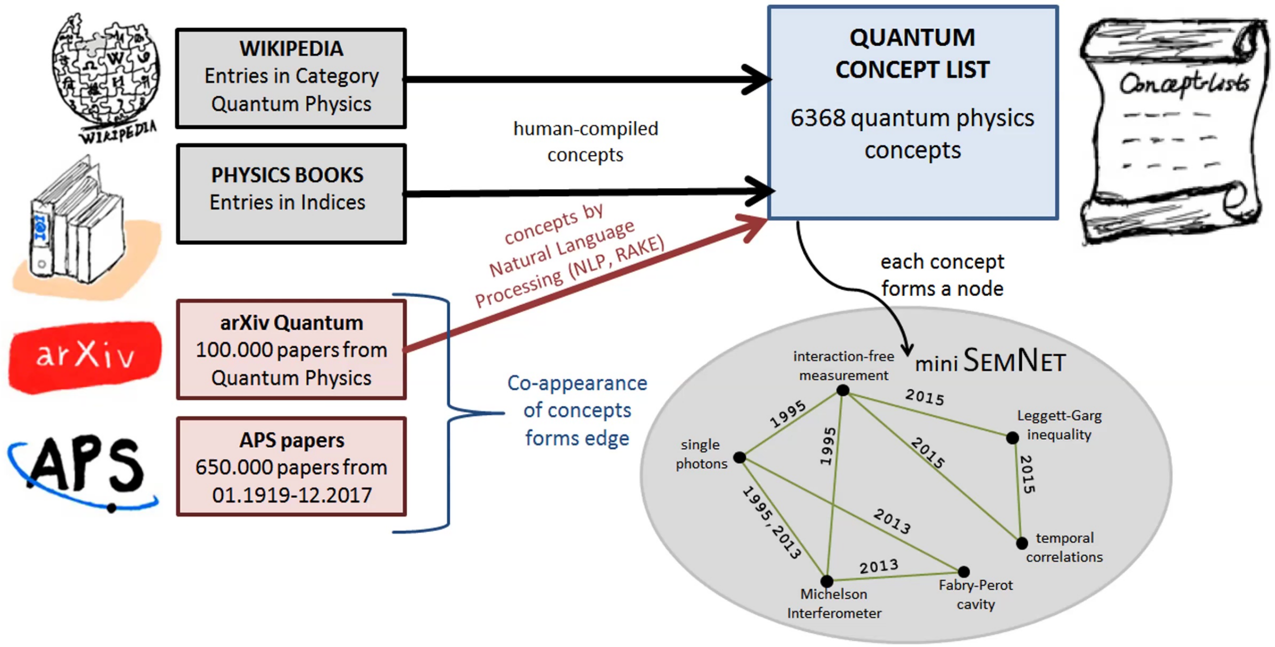




Concept List of Quantum Physics

Krenn, Zeilinger "Predicting Research Trends with Semantic and Neural Networks with an application in Q.Physics", *PNAS* **117**(4), 1910 (2020).

Concept List of Quantum Physics



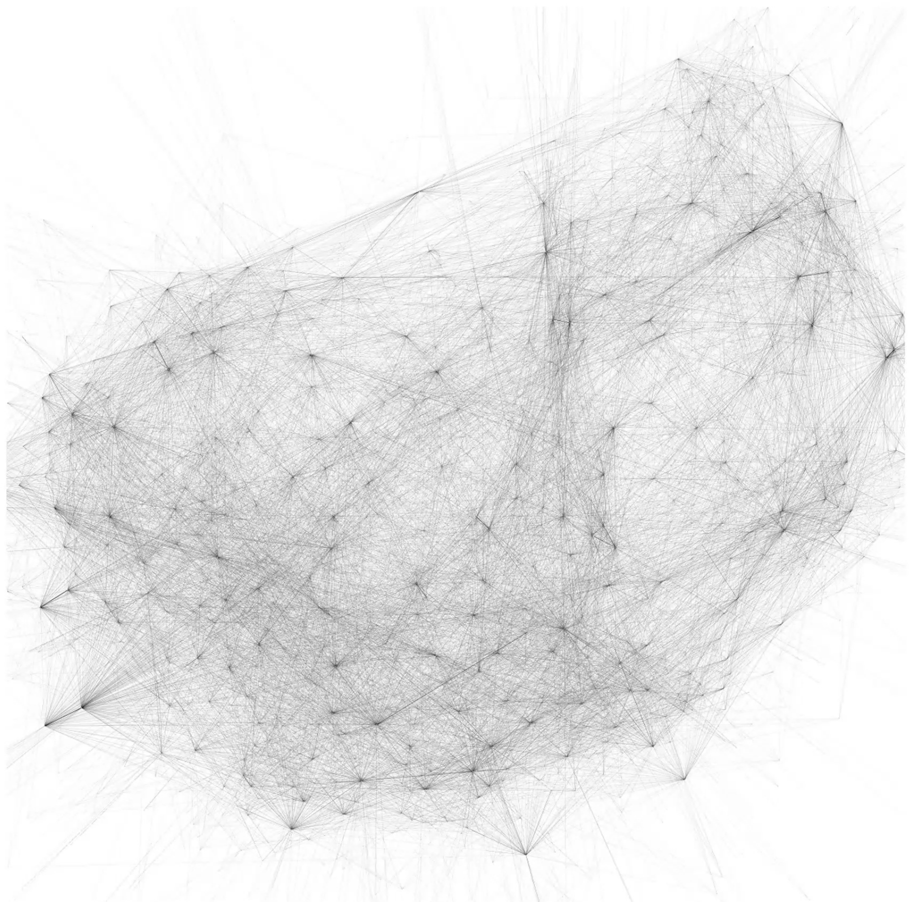
Phys. Rev. Lett. 74, 4763 (1995):
[...] The fraction of **"interaction-free" measurements** can be arbitrarily close to 1. Using **single photons** in a **Michelson interferometer**, we have performed a preliminary demonstration of some of these ideas. [...]

Phys. Rev. A 88, 023812 (2013):
We study the quantum dynamics of a **Michelson interferometer** with **Fabry-Perot cavity** arms and one movable end mirror, and driven by a **single photon**—an optomechanical device previously studied by Marshall et al. as a device that searches for gravity decoherence. [...]

Phys. Rev. X 5, 011003 (2015):
[...] Measuring **temporal correlations** of the position of single atoms performing a quantum walk, we observe a 6σ violation of the **Leggett-Garg inequality**. [...] The **interaction-free measurement** is based on a novel atom transport system, which allows us to directly probe the absence rather than the presence of atoms at a chosen lattice site. [...]

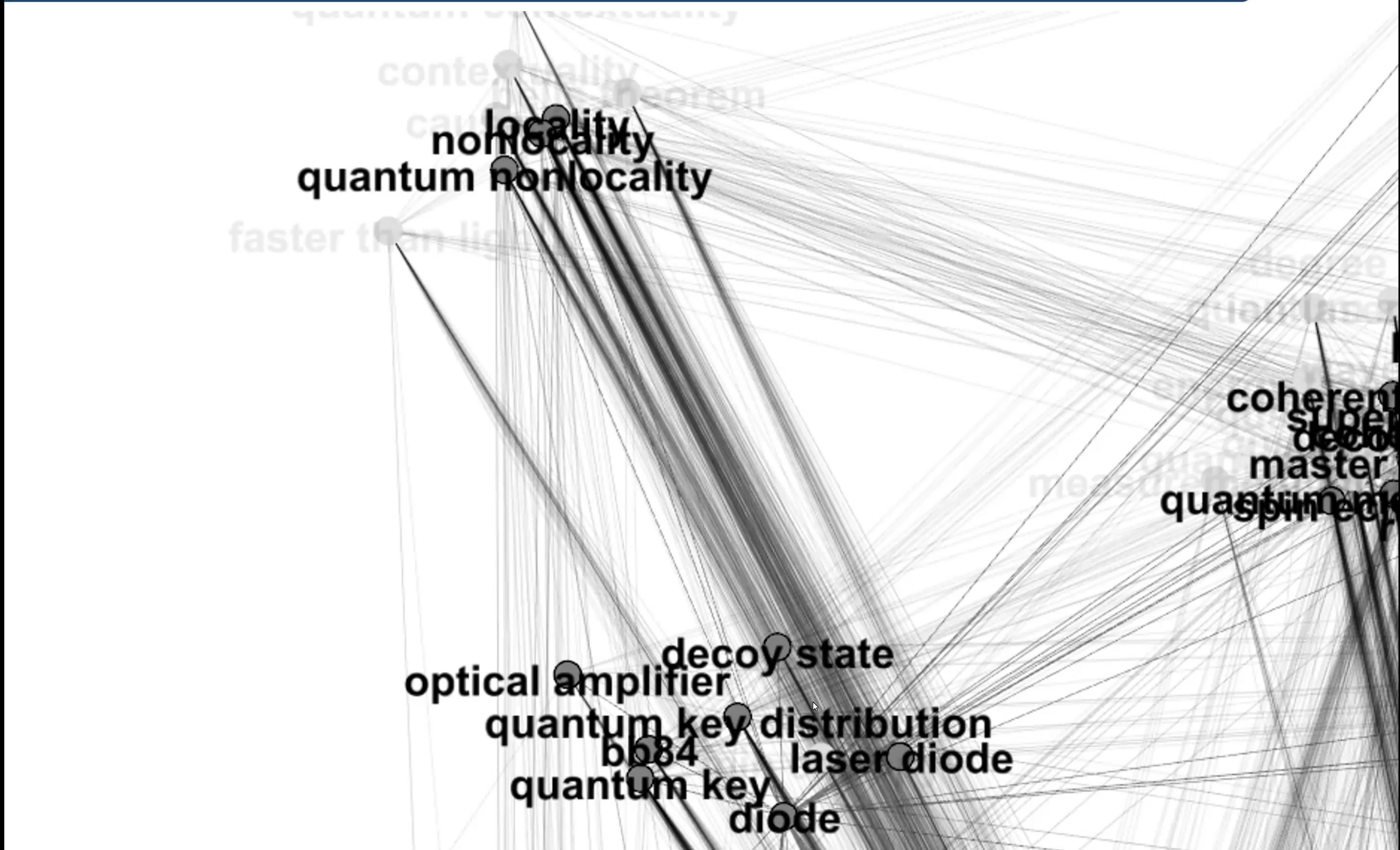
Krenn, Zeilinger "Predicting Research Trends with Semantic and Neural Networks with an application in Q.Physics", *PNAS* **117**(4), 1910 (2020).

Semantic Network of Quantum Physics



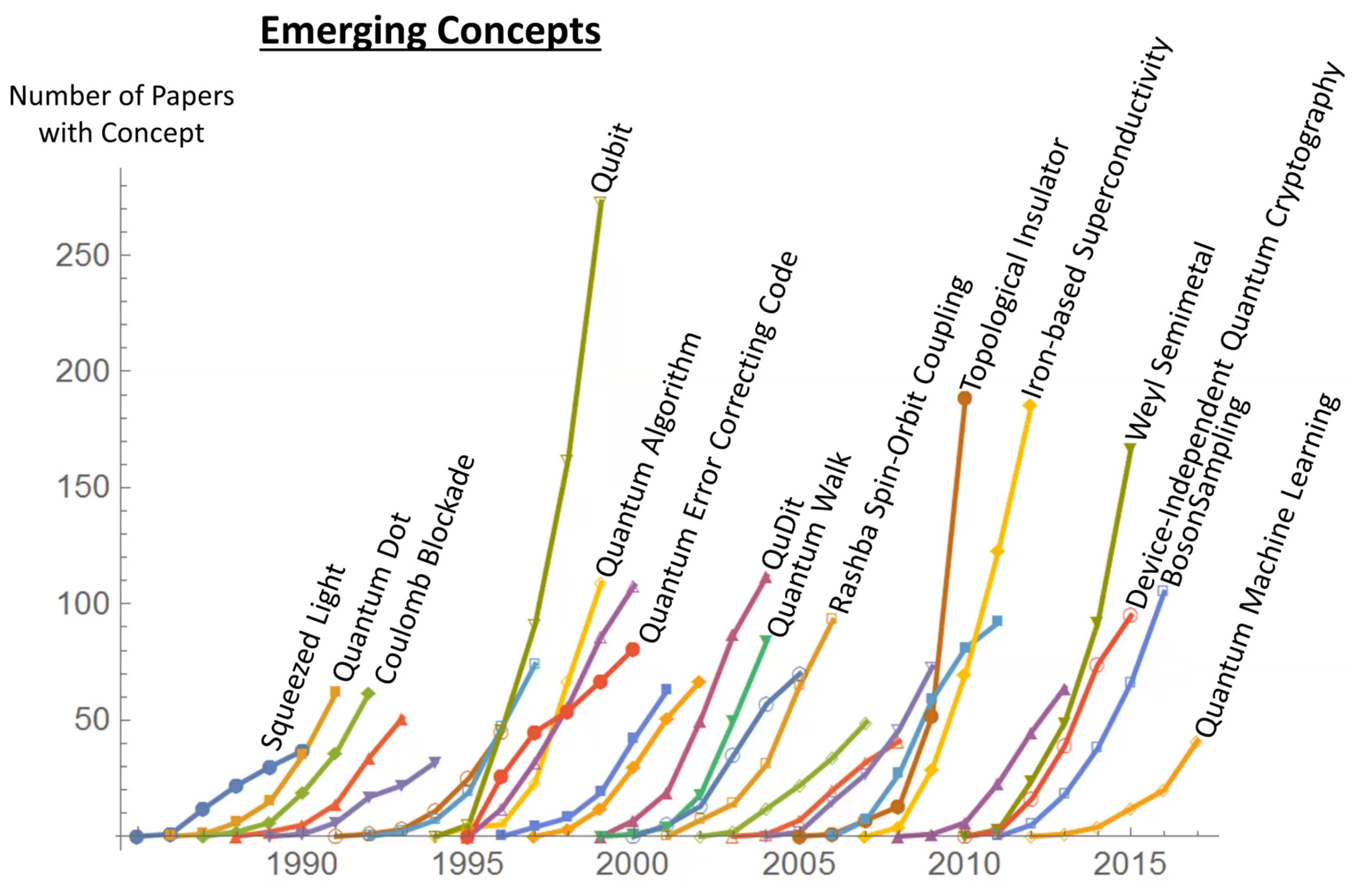
Krenn, Zeilinger "Predicting Research Trends with Semantic and Neural Networks with an application in Q.Physics", *PNAS* **117**(4), 1910 (2020).

Semantic Network of Quantum Physics



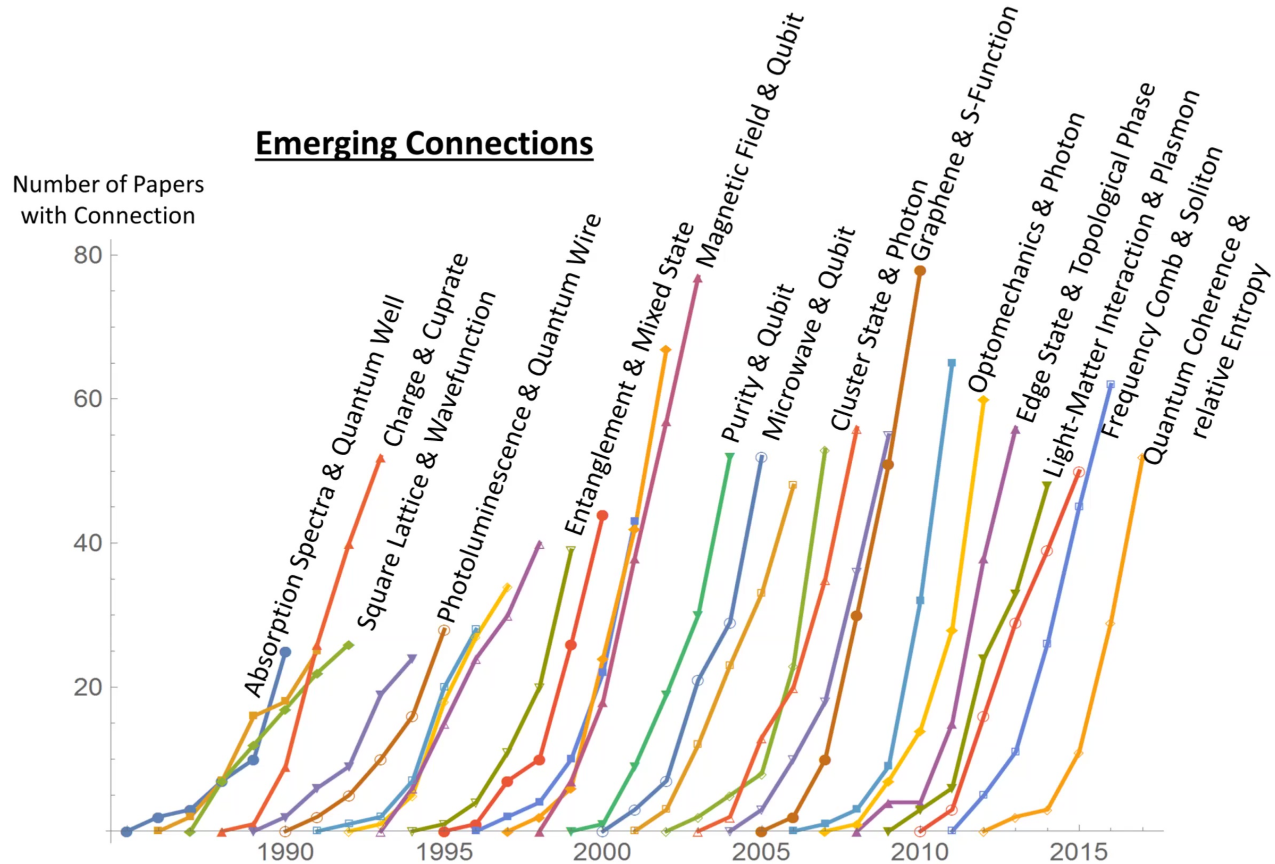
Krenn, Zeilinger "Predicting Research Trends with Semantic and Neural Networks with an application in Q.Physics", *PNAS* **117**(4), 1910 (2020).

Some History of Quantum Physics



Krenn, Zeilinger "Predicting Research Trends with Semantic and Neural Networks with an application in Q.Physics", *PNAS* **117**(4), 1910 (2020).

Some History of Quantum Physics

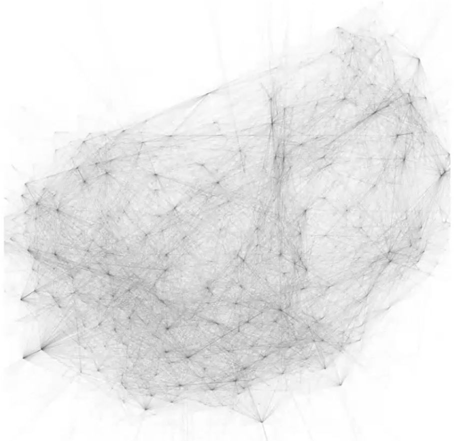


MK, Zeilinger "Predicting Research Trends with Semantic and Neural Networks with an application in Quantum Physics", *PNAS* **117**(4), 1910 (2020).

Future of Quantum Physics Research

Which unconnected pairs of concepts will be investigated together in 5 years?

Neural Network – but how?



Way too large and complex as a direct input

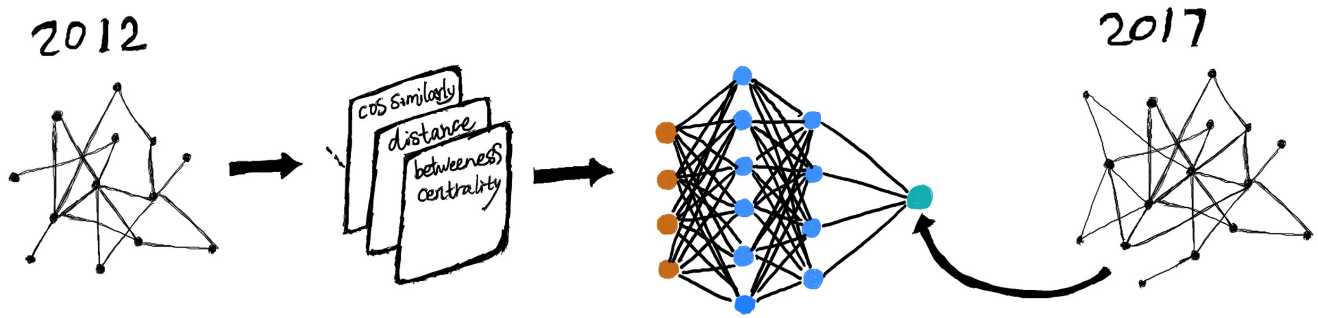
Solution: Network Properties of pairs of concepts

For each concept pair:

- #(papers of concept)
- #(connections of concept)
- #(shared Neighbors)
- Network Distances
- #(Paths of Length 2, 3, 4 over last 3 years)

Krenn, Zeilinger "Predicting Research Trends with Semantic and Neural Networks with an application in Q.Physics", *PNAS* **117**(4), 1910 (2020).

Future of Quantum Physics Research



Link Prediction
via Neural Networks

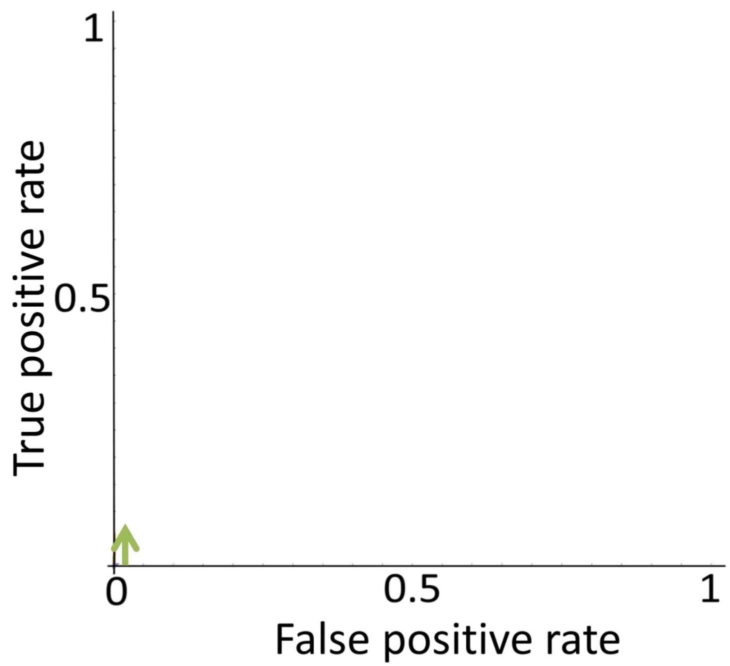
Krenn, Zeilinger "Predicting Research Trends with Semantic and Neural Networks with an application in Q.Physics", *PNAS* **117**(4), 1910 (2020).

Future of Quantum Physics Research

Quantify Quality of Prediction - ROC:

(Receiver Operating Characteristic)

6000x6000=36 Mio unconnected concepts
Sorted from likely to unlikely



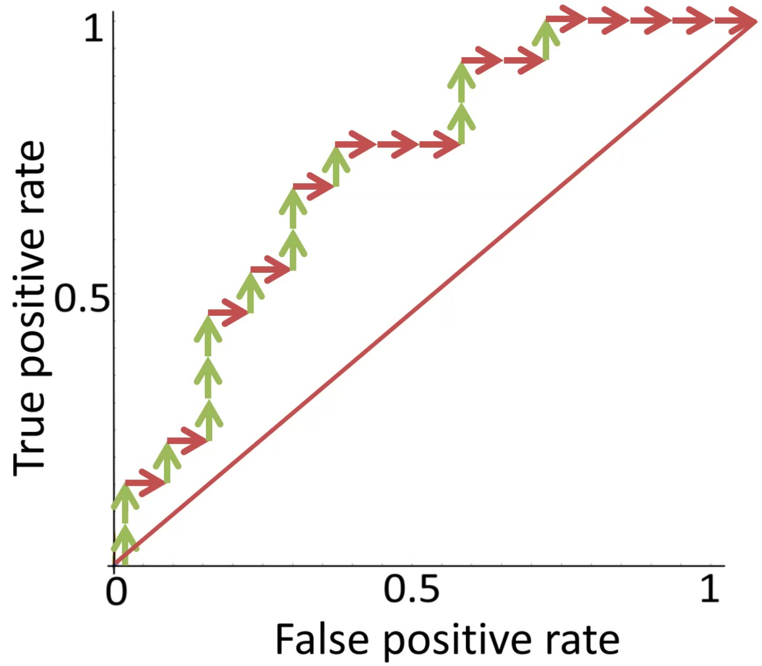
MK, Zeilinger "Predicting Research Trends with Semantic and Neural Networks with an application in Quantum Physics", arXiv:1906.06843

Future of Quantum Physics Research

Quantify Quality of Prediction - ROC:

(Receiver Operating Characteristic)

6000x6000=36 Mio unconnected concepts
Sorted from likely to unlikely



MK, Zeilinger "Predicting Research Trends with Semantic and Neural Networks with an application in Quantum Physics", arXiv:1906.06843

Future of Quantum Physics Research

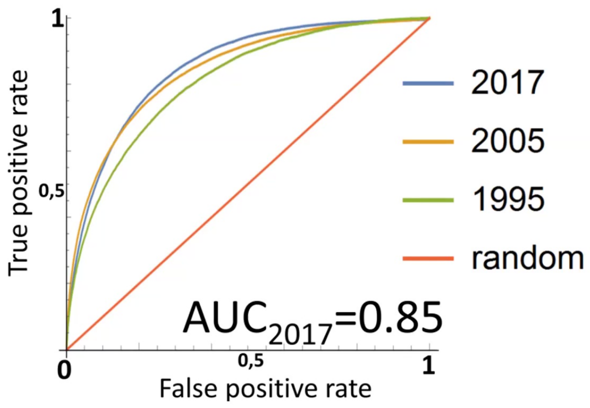
Quantify Quality of Prediction:

(Receiver Operating Characteristic)

Area Under Curve:

% that random true element is ranked higher than random false one.

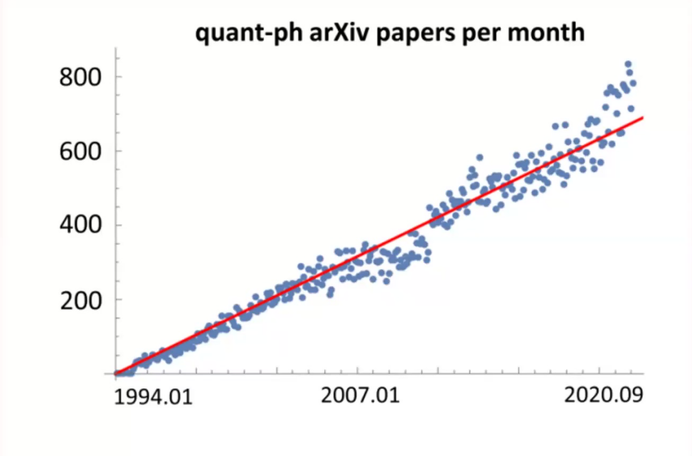
Here: AUC=85%

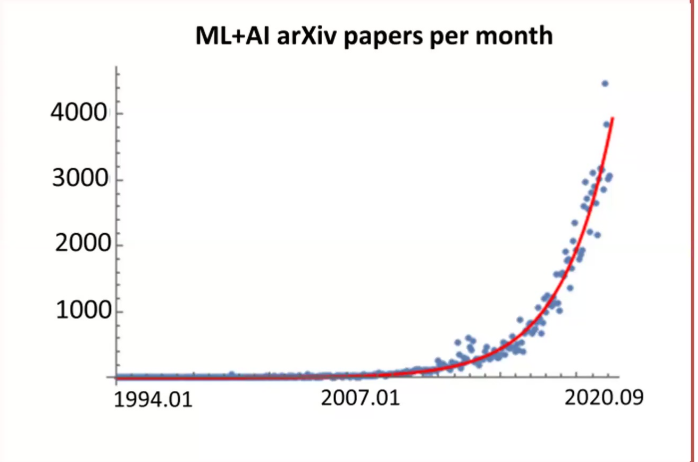


Krenn, Zeilinger "Predicting Research Trends with Semantic and Neural Networks with an application in Q.Physics", *PNAS* **117**(4), 1910 (2020).

Towards personalized Suggestions

Problem: The growing number of publications in all disciplines of science cannot be comprehended by a single human researcher.





Krenn, Zeilinger "Predicting Research Trends with Semantic and Neural Networks with an application in Q.Physics", *PNAS* **117**(4), 1910 (2020).

Towards personalized Suggestions

$$p_{\text{scientist}}(c_i) = \frac{N(c_i)}{\sum_j N(c_j)}$$

$$p_{\text{total}}(c_i) = \frac{M(c_i)}{\sum_j M(c_j)}$$

$$r_{\text{scientist}}(c_i) = \frac{p_{\text{scientist}}(c_i)}{p_{\text{total}}(c_i)}$$

Krenn, Zeilinger "Predicting Research Trends with Semantic and Neural Networks with an application in Q.Physics", *PNAS* **117**(4), 1910 (2020).

Towards personalized Suggestions

Concepts used in papers by Roger Melko

- Used ~104 papers
- ~918 different concepts used, out of 27.067 total concepts (comparison: i used ~500)

Concepts c: $r_{Roger}(c)$

monte carlo: 2167
spin ice: 1142
transverse field ising: 1028
liquid phase: 727
quantum monte carlo: 685
boltzmann machine: 596
ground state phase: 457
spin configuration: 421
lattice boson: 380
spin liquid phase: 380
valence bond solid: 380
machine learning approach: 380
quantum monte carlo simulation: 360
supersolid state: 342
generative model: 336
spin liquid: 316
quantum spin liquid: 302
topological sector: 285
reconstructing quantum state: 253
deep network: 253

$$p_{\text{scientist}}(c_i) = \frac{N(c_i)}{\sum_j N(c_j)}$$

$$p_{\text{total}}(c_i) = \frac{M(c_i)}{\sum_j M(c_j)}$$

$$r_{\text{scientist}}(c_i) = \frac{p_{\text{scientist}}(c_i)}{p_{\text{total}}(c_i)}$$

Krenn, Zeilinger "Predicting Research Trends with Semantic and Neural Networks with an application in Q.Physics", *PNAS* **117**(4), 1910 (2020).

Towards personalized Suggestions

Concepts used in Roger‘s papers–

with other physical concepts – What is predicted?

Roger‘s predictions:

Predicted:

1: quantum phase transition, quantum key distribution
2: quantum phase transition, hidden variable
3: finite temperature , quantum network
4: critical point , fault tolerant

Unorthodox (small degree centrality and cos-Similarity):

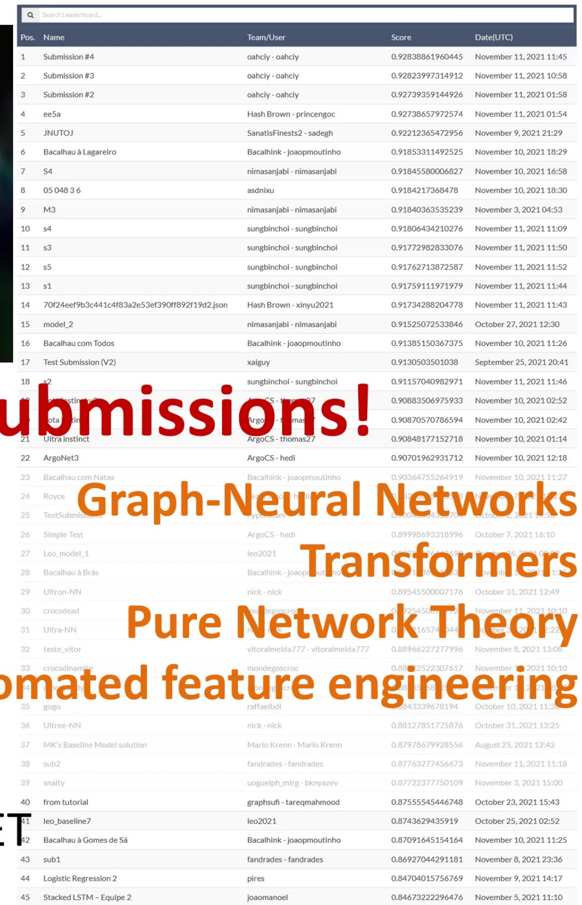
1: bose hubbard model , single photon detector
2: zero temperature, reverse reconciliation
3: triangular lattice, markovian open quantum system
4: scale invariant, optical quantum information

Unpredicted:

1: deconfined quantum critical point - casimir density
2: quantum monte carlo, quantum dot cellular automaton
3: numerical linked cluster expansion , individual quantum trajectory

Krenn, Zeilinger "Predicting Research Trends with Semantic and Neural Networks with an application in Q.Physics", *PNAS* **117**(4), 1910 (2020).

Extension: Science4Cast competition



Pos.	Name	Team/User	Score	Date(UTC)
1	Submission #4	oahciy - oahciy	0.92838861960445	November 11, 2021 11:45
2	Submission #3	oahciy - oahciy	0.92823997314912	November 11, 2021 10:58
3	Submission #2	oahciy - oahciy	0.92739359144926	November 11, 2021 01:58
4	ee5a	Hash Brown - princengoc	0.92738657972574	November 11, 2021 01:54
5	JNUTOJ	SanatisFinests2 - sadegh	0.92212365472956	November 9, 2021 21:29
6	Bacalhau à Lagareiro	Bacalhink - joaopmoutinho	0.91853311492525	November 10, 2021 18:29
7	S4	nimasanjabi - nimasanjabi	0.91845580006827	November 10, 2021 16:58
8	05 048 3 6	asdnixu	0.9184217368478	November 10, 2021 18:30
9	M3	nimasanjabi - nimasanjabi	0.91840363535239	November 3, 2021 04:53
10	s4	sungbinchoi - sungbinchoi	0.91806434210276	November 11, 2021 11:09
11	s3	sungbinchoi - sungbinchoi	0.91772982833076	November 11, 2021 11:50
12	s5	sungbinchoi - sungbinchoi	0.91762713872587	November 11, 2021 11:52
13	s1	sungbinchoi - sungbinchoi	0.91759111971979	November 11, 2021 11:44
14	70f24eef9b3c441c4f83a2e53ef390ff892f19d2.json	Hash Brown - xinyu2021	0.91734288204778	November 11, 2021 11:43
15	model_2	nimasanjabi - nimasanjabi	0.91525072533846	October 27, 2021 12:30
16	Bacalhau com Todos	Bacalhink - joaopmoutinho	0.91385150367375	November 10, 2021 11:26
17	Test Submission (V2)	xaiguy	0.9130503501038	September 25, 2021 20:41
18	s2	sungbinchoi - sungbinchoi	0.91157040982971	November 11, 2021 11:46
[illegible]	[illegible]	[illegible]	0.90883506975933	November 10, 2021 02:52
[illegible]	[illegible]	[illegible]	0.90870570786594	November 10, 2021 02:42
21	Ultra instinct	ArgoCS - thomas27	0.90848177152718	November 10, 2021 01:14
22	ArgoNet3	ArgoCS - hedi	0.90701962931712	November 10, 2021 12:18
23	Bacalhau com Natas	Bacalhink - joaopmoutinho	0.90364755264919	November 10, 2021 11:27
24	Royce	[illegible]	[illegible]	[illegible]
25	[illegible]	[illegible]	[illegible]	[illegible]
26	Simple Test	ArgoCS - hedi	0.89998693318996	October 7, 2021 16:10
27	Leo_model_1	leo2021	[illegible]	[illegible]
28	Bacalhau à Brás	[illegible]	[illegible]	[illegible]
29	Ultron-NN	nick - nick	0.89545500007176	October 31, 2021 12:49
30	crocsdead	[illegible]	[illegible]	November 11, 2021 10:10
31	Ultra-NN	[illegible]	[illegible]	[illegible]
32	teste_vitor	vitoralmeida777 - vitoralmeida777	0.88966227277996	November 8, 2021 13:08
33	crocsdinamite	mondegoscroc	[illegible]	[illegible]
34	[illegible]	[illegible]	[illegible]	[illegible]
35	gogo	raffaelbdl	[illegible]	[illegible]
36	Ultree-NN	nick - nick	0.88127851775876	October 31, 2021 13:25
37	MK's Baseline Model solution	Mario Krenn - Mario Krenn	0.87978679928556	August 25, 2021 12:42
38	sub2	fandrades - fandrades	0.87763277456673	November 11, 2021 11:18
39	snaity	uoguelph_mlrg - bknyazev	0.87722377750109	November 3, 2021 15:00
40	from tutorial	graphsufi - tareqmahmood	0.87555545446748	October 23, 2021 15:43
41	leo_baseline7	leo2021	0.8743629435919	October 25, 2021 02:52
42	Bacalhau à Gomes de Sá	Bacalhink - joaopmoutinho	0.87091645154164	November 10, 2021 11:25
43	sub1	fandrades - fandrades	0.86927044291181	November 8, 2021 23:36
44	Logistic Regression 2	pires	0.84704015756769	November 9, 2021 14:17
45	Stacked LSTM – Equipe 2	joaomanoel	0.84673222296476	November 5, 2021 11:10

72 Submissions!

Graph-Neural Networks

Transformers

Pure Network Theory

Automated feature engineering

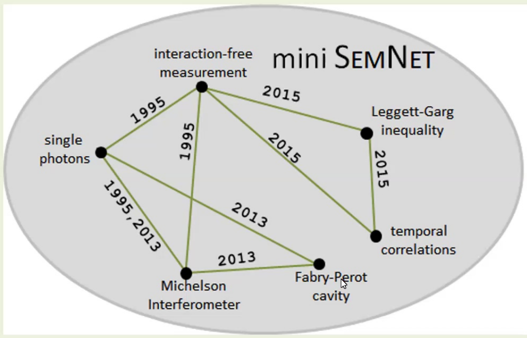
Concluding Special Session:

17.Dezember (Next Friday) – 11.00 ET

Krenn, Zeilinger "Predicting Research Trends with Semantic and Neural Networks with an application in Q.Physics", *PNAS* **117**(4), 1910 (2020).

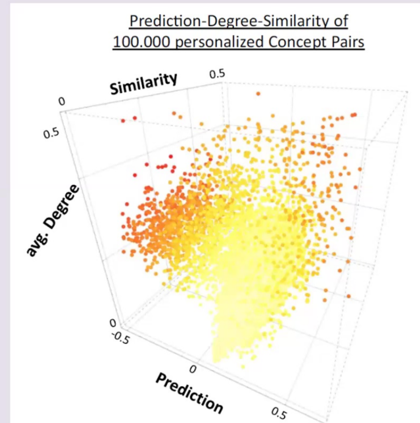
Conclusion & Future

Preditions new research directions via Semantic & Neural Networks



Future:

- **Personalized suggestions** (see before)
- **Surprising suggestions** (network theory or anomaly detection)
- **Fruitful suggestions** (application of proxy of success, e.g. citations)



- **Automated Concepts** via WordEmbedding?

Questions:

How does a human scientist define new research projects?

How can an "automated idea finder" be evaluated?

Krenn, Zeilinger "Predicting Research Trends with Semantic and Neural Networks with an application in Q.Physics", *PNAS* **117**(4), 1910 (2020).