Exploring the Strong Coupling Through Natural Language Processing

Patrick L.S. Connor, **Antonin Sulc**
Hamburg, 22.02.2024

HELMHOLTZ

Strong Coupling

- The strong force describes how the proton constituents (quarks and gluons) hold together.

Strong Coupling

- The strong force describes how the proton constituents (quarks and gluons) hold together.
- The intesity of the force is determined by the **strong coupling**.

Strong Coupling

- The strong force describes how the proton constituents (quarks and gluons) hold together.
- The intesity of the force is determined by the **strong coupling**.
- At high energy scales or short distance scales, the strong coupling is small (~ 0.1) - quarks behave almost freely (asymptotic freedom).

Strong Coupling

- The strong force describes how the proton constituents (quarks and gluons) hold together.
- The intesity of the force is determined by the **strong coupling**.
- At high energy scales or short distance scales, the strong coupling is small (~ 0.1) - quarks behave almost freely (asymptotic freedom).
- As energy scale decreases or distance scale grows, the strong coupling increases.

Strong Coupling

- The strong force describes how the proton constituents (quarks and gluons) hold together.
- The intesity of the force is determined by the **strong coupling**.
- At high energy scales or short distance scales, the strong coupling is small (~ 0.1) - quarks behave almost freely (asymptotic freedom).
- As energy scale decreases or distance scale grows, the strong coupling increases.
- Strong coupling increases at lower energies, in contrast to the electromagnetic and weak forces.

Strong Coupling

- The strong force describes how the proton constituents (quarks and gluons) hold together.
- The intesity of the force is determined by the **strong coupling**.
- At high energy scales or short distance scales, the strong coupling is small (~ 0.1) - quarks behave almost freely (asymptotic freedom).
- As energy scale decreases or distance scale grows, the strong coupling increases.
- Strong coupling increases at lower energies, in contrast to the electromagnetic and weak forces.
- The behavior of the strong coupling constant is an important feature of the strong force in the Standard Model. Yet it is one of the least known parameters (only known at percent-level accuracy).

Approaches

Embedding

Analysis of semantic similarity through vector representation of abstracts.

`...Space-Time Geometry Using Young diagrams...` $\rightarrow \mathbf{e} = [0.1, 0.3, -0.2\ldots]$

`...Status of NNLO 3-jet calculations..` $\rightarrow \mathbf{e} = [0.1, -0.3, 0.2\ldots]$

Approaches

Embedding

Analysis of semantic similarity through vector representation of abstracts.

`...Space-Time Geometry Using Young diagrams...` $\rightarrow \mathbf{e} = [0.1, 0.3, -0.2\ldots]$

`...Status of NNLO 3-jet calculations..` $\rightarrow \mathbf{e} = [0.1, -0.3, 0.2\ldots]$

Topic Modeling

"Clustering" of abstracts and extraction of their most characteristic words.

`...Space-Time Geometry Using Young diagrams...` $\rightarrow Topic_1 = \{young, diagram\}$

`...Status of NNLO 3-jet calculations..` $\rightarrow Topic_2 = \{status, jet\}$

Approaches

Embedding

Analysis of semantic similarity through vector representation of abstracts.

`...Space-Time Geometry Using Young diagrams...` $\rightarrow \mathbf{e} = [0.1, 0.3, -0.2 \ldots]$

`...Status of NNLO 3-jet calculations..` $\rightarrow \mathbf{e} = [0.1, -0.3, 0.2 \ldots]$

Topic Modeling

"Clustering" of abstracts and extraction of their most characteristic words.

`...Space-Time Geometry Using Young diagrams...` $\rightarrow Topic_1 = \{young, diagram\}$

`...Status of NNLO 3-jet calculations..` $\rightarrow Topic_2 = \{status, jet\}$

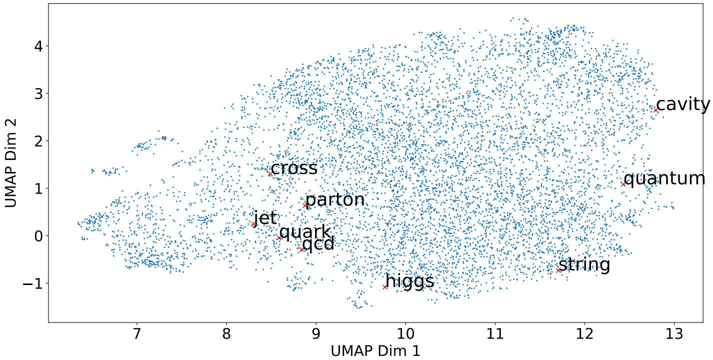
Citation Graph

Building a directed graph on citations and identification of context from reference list.

Word Embedding

Representing individual words as vectors.

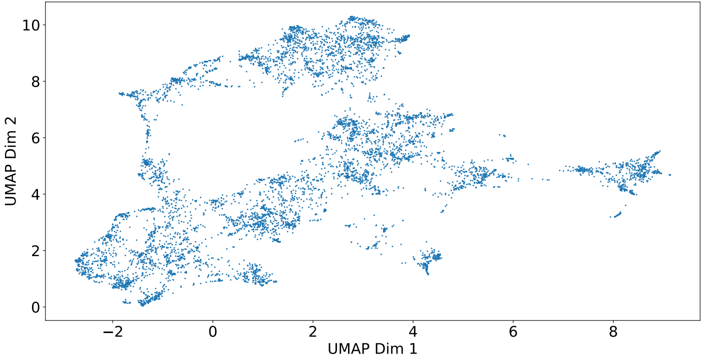
$$\begin{aligned}
\texttt{cavity} &\quad \rightarrow \mathbf{w} = [0.1, -1, 2, \dots] \\
\texttt{qcd} &\quad \rightarrow \mathbf{w} = [0.2, 0.8, 1.8, \dots] \\
\texttt{photon} &\quad \rightarrow \mathbf{w} = [0.18, 1.1, 1.8, \dots]
\end{aligned}$$



Sentence Embedding

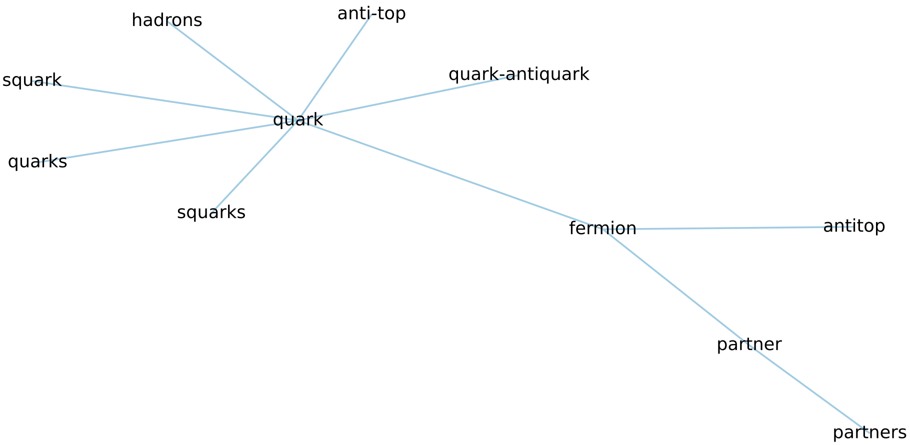
Representing sentences as vectors.

$$\texttt{...using Young ...} \rightarrow \mathbf{e} = [0.1, \ldots]$$
$$\texttt{...NNLO 3-jet ...} \rightarrow \mathbf{e} = [0.8, \ldots]$$



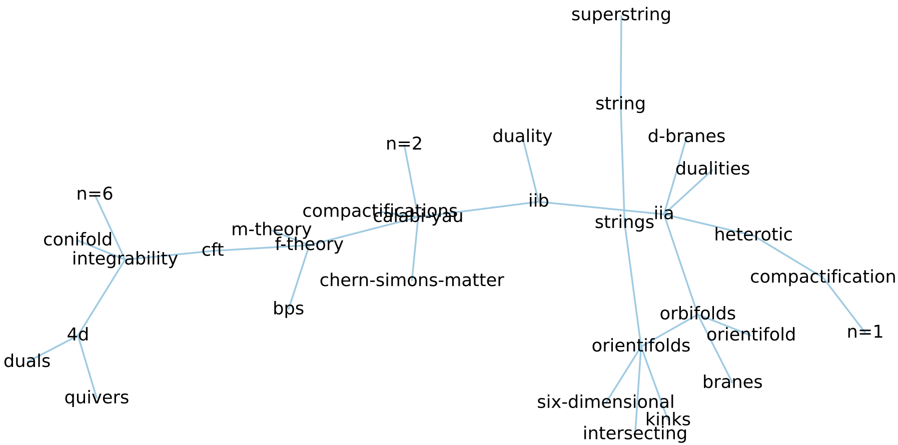
Word Embedding - Quark Similarity Graph

Quark



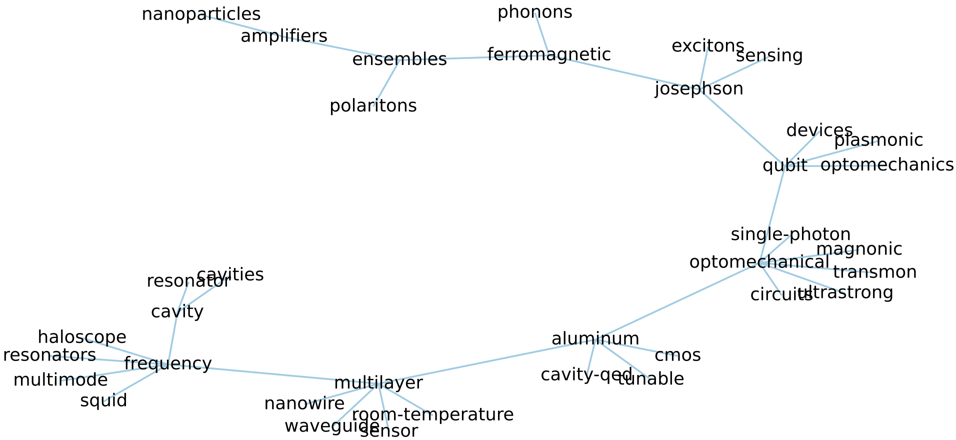
Word Embedding - String Similarity Graph

String



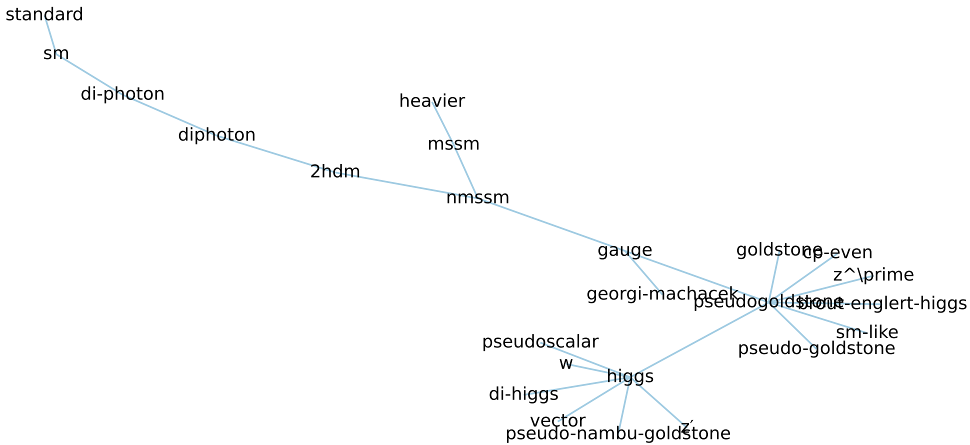
Word Embedding - Cavity Similarity Graph

Cavity

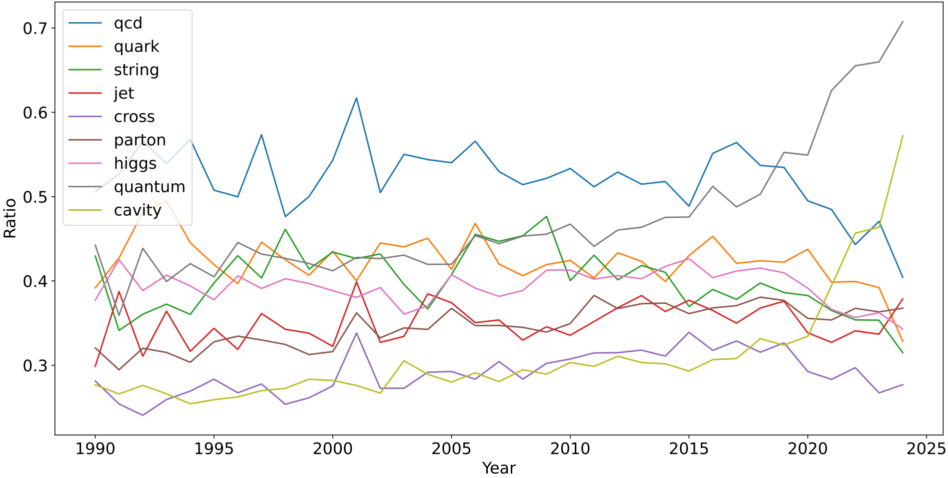


Word Embedding - Higgs Similarity Graph

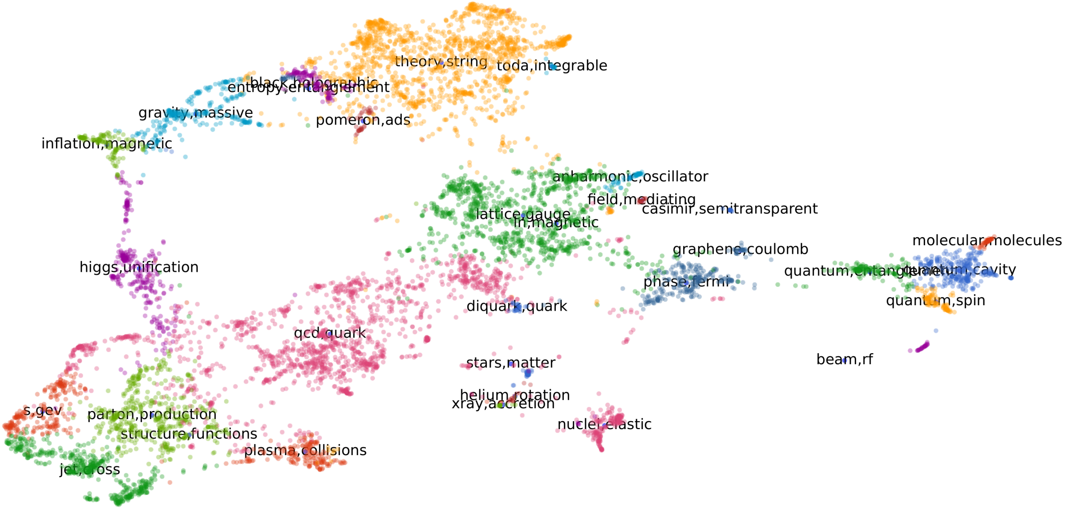
Higgs



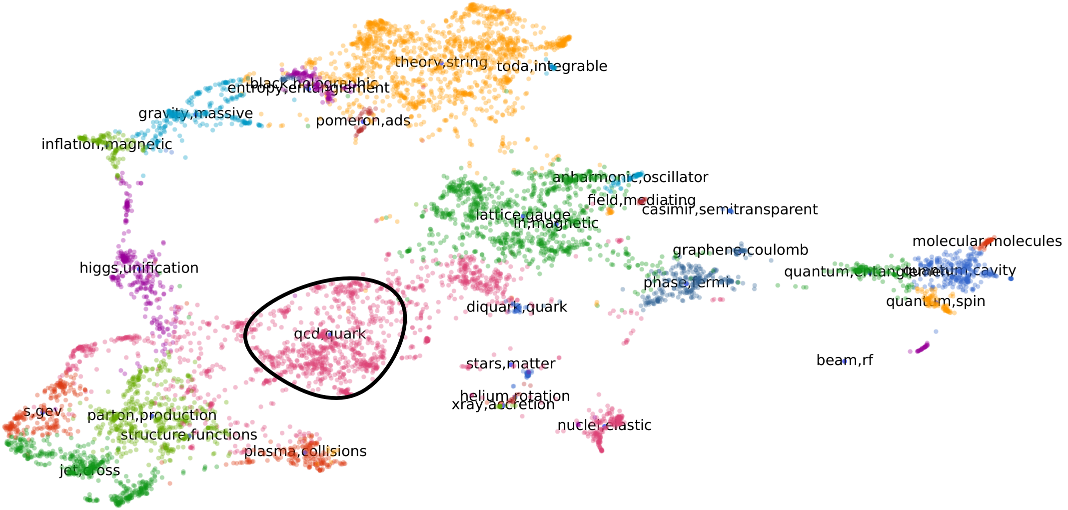
Trends of Selected Words



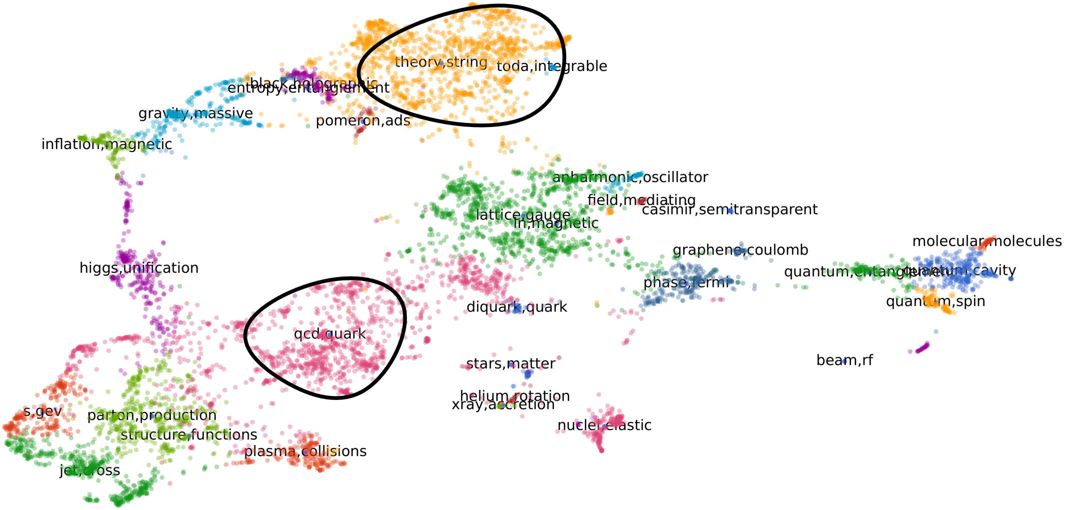
Topic Modelling



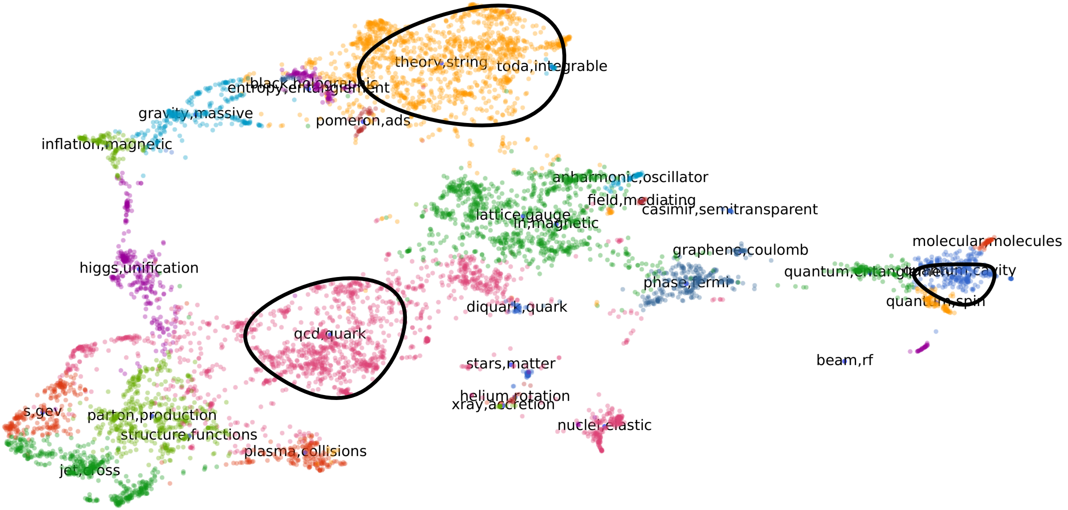
Topic Modelling - QCD, Quark



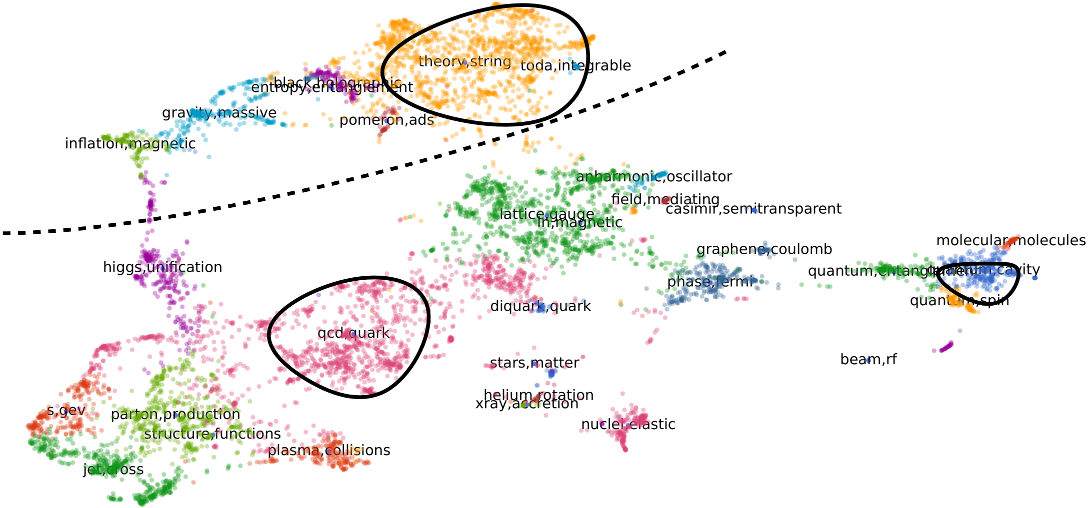
Topic Modelling - String Theory



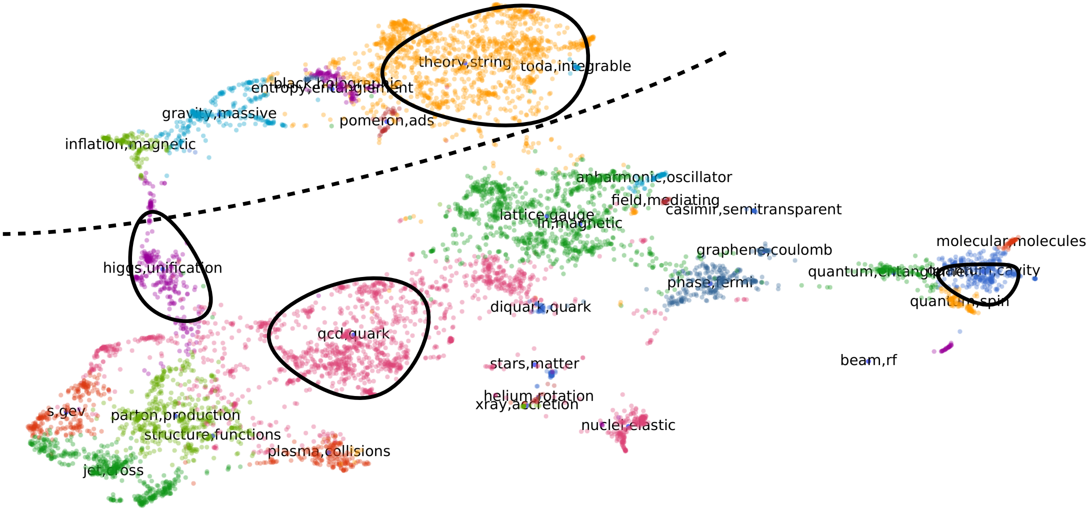
Topic Modelling - Quantum



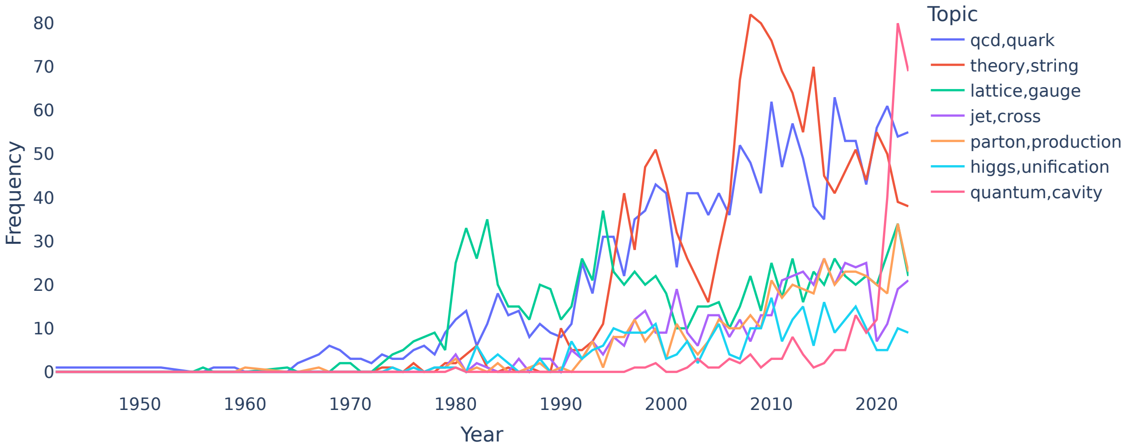
Topic Modelling - Border



Topic Modelling - Higgs, Unification



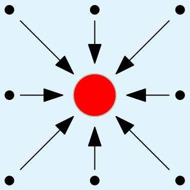
Topics over Time



Analysis of Citation Graph - Degree Centrality

Degree centrality

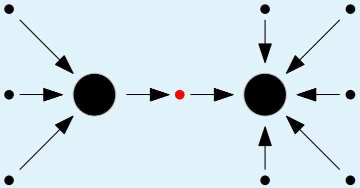
Most cited papers



- **0.043** Gauge theory correlators from noncritical string theory
- **0.015** Confinement of Quarks
- **0.014** The Shear viscosity of strongly coupled N=4 supersymmetric Yang-Mills plasma

Analysis of Citation Graph - Between-ness Centrality

Between-ness centrality

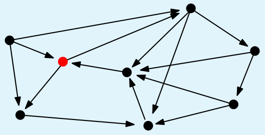


- > **0.22** Confinement of Quarks
- > **0.21** Gauge theory correlators from noncritical string theory
- > **0.08** The QCD Running Coupling

Analysis of Citation Graph - PageRank

PageRank centrality

A node with high PageRank means: routes to other nodes, and has many visits over iterative passes.



- **0.005** `Gauge theory correlators from noncritical string theory`
- **0.003** `Cavity quantum electrodynamics`
- **0.002** `Confinement of Quarks`

Citation Graph Embedding - {Theory, String,...} Topic

Citation Graph Embedding - {QCD, Quark,...} Topic

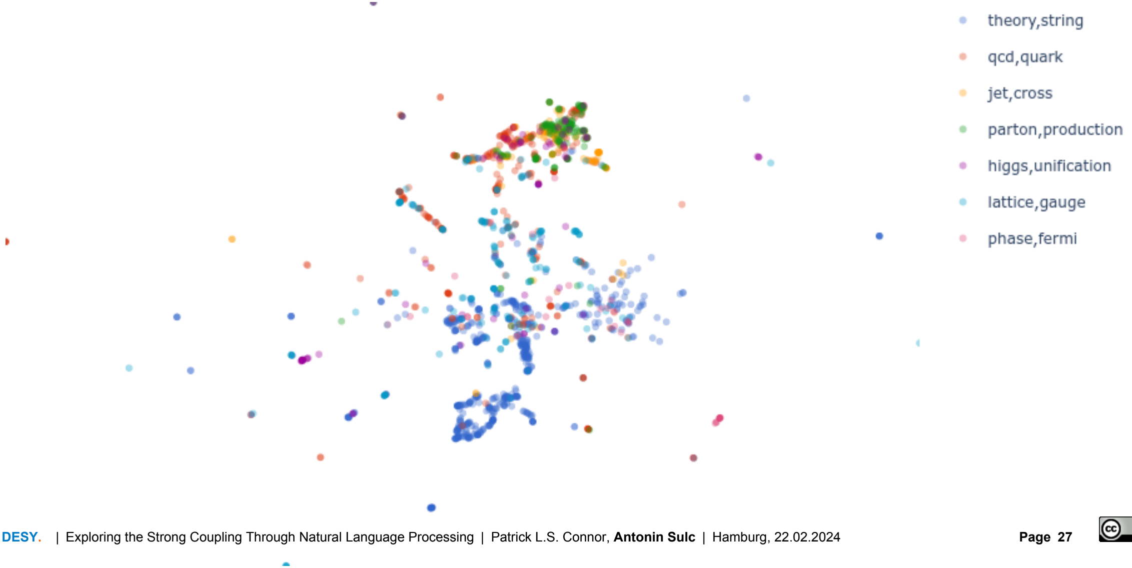
Citation Graph Embedding - {Jet, Cross,...} Topic

Citation Graph Embedding - {Parton, Production,...} Topic

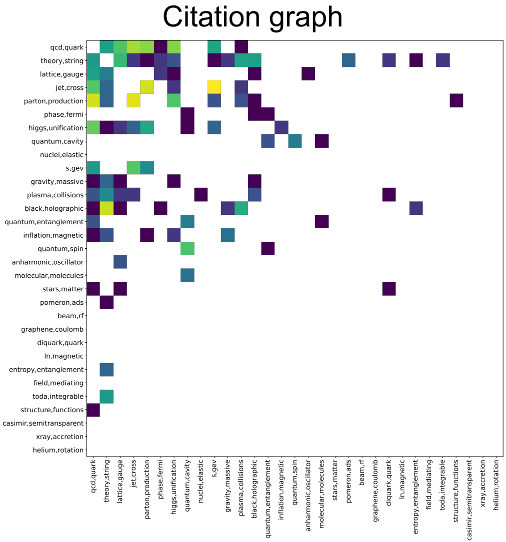
Citation Graph Embedding - {Higgs, Unification,...} Topic

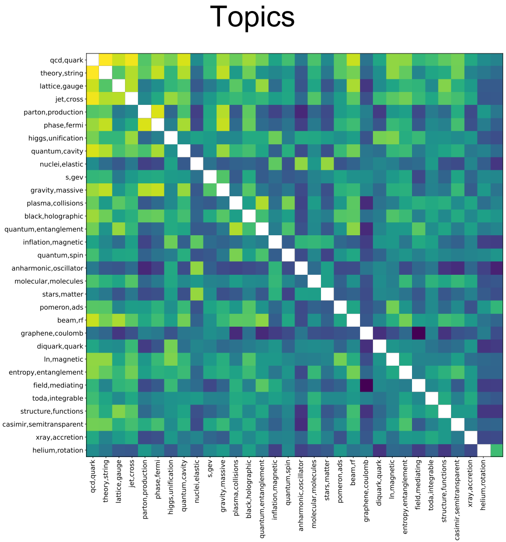
Citation Graph Embedding - {Lattice, Gauge,...} Topic

Citation Graph Embedding - {Phase, Fermni,...} Topic

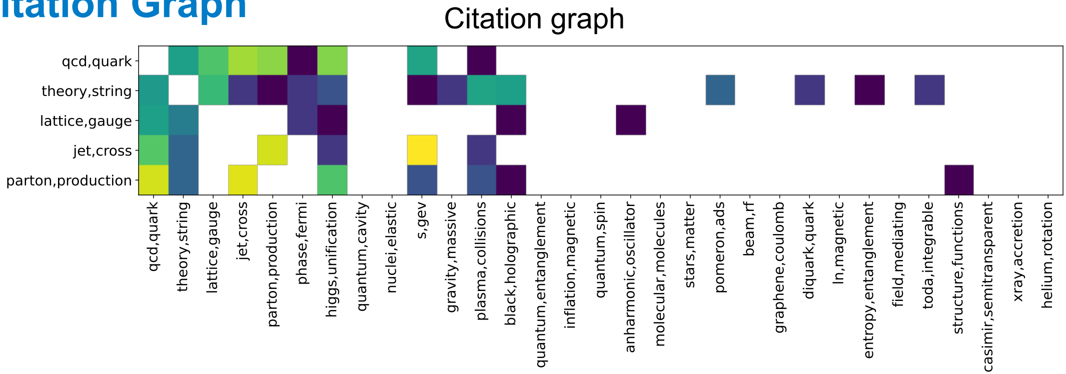


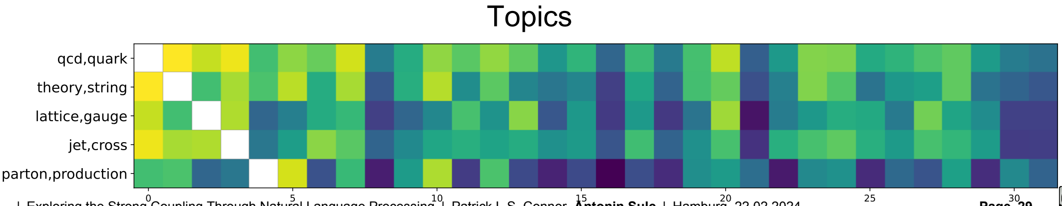
Citation Graph





Citation Graph





Conclusion & Summar

- You heard about embedding of abstracts into vectors.
- You learned about (unsupervised) topic modelling of the current dark matter publication landscape from InspireHEP.
- You saw that you can built knowledge graph just based data from abstracts.
- You learned something about about community and citations.

Future Work

- Use papers (not only abstracts)

Future Work

- Use papers (not only abstracts)
- Fine-tune embedding (did not work for now, we used generic `all-MiniLM-L6-v2`)

Future Work

- Use papers (not only abstracts)
- Fine-tune embedding (did not work for now, we used generic `all-MiniLM-L6-v2`)
- Fine-tune a LLM (good example is `phi-1.5`)

Future Work

- Use papers (not only abstracts)
- Fine-tune embedding (did not work for now, we used generic `all-MiniLM-L6-v2`)
- Fine-tune a LLM (good example is `phi-1.5`)
- Incorporate topological information

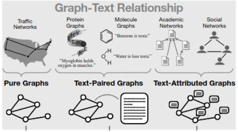
Future Work

- Use papers (not only abstracts)
- Fine-tune embedding (did not work for now, we used generic `all-MiniLM-L6-v2`)
- Fine-tune a LLM (good example is `phi-1.5`)
- Incorporate topological information

Large Language Models on Graphs: A Comprehensive Survey

Bowen Jin*, Gang Liu*, Chi Han*, Meng Jiang, Heng Ji, Jiawei Han

Abstract—Large language models (LLMs), such as GPT4 and LLaMA, are creating significant advancements in natural language processing, due to their strong text encoding/decoding ability and newly found emergent capability (*e.g.*, reasoning). While LLMs are mainly designed to process pure texts, there are many real-world scenarios where text data is associated with rich structure information in the form of graphs (*e.g.*, academic networks, and e-commerce networks) or scenarios where graph data are paired with rich textual information (*e.g.*, molecules with descriptions). Besides, although LLMs have shown their pure text-based reasoning ability, it is underexplored whether such ability can be generalized to graphs (*i.e.*, graph-based reasoning). In this paper, we provide a systematic review of scenarios and techniques related to large language models on graphs. We first summarize potential scenarios of



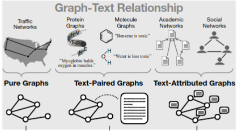
Future Work

- Use papers (not only abstracts)
- Fine-tune embedding (did not work for now, we used generic `all-MiniLM-L6-v2`)
- Fine-tune a LLM (good example is `phi-1.5`)
- Incorporate topological information

Large Language Models on Graphs: A Comprehensive Survey

Bowen Jin*, Gang Liu*, Chi Han*, Meng Jiang, Heng Ji, Jiawei Han

Abstract—Large language models (LLMs), such as GPT4 and LLaMA, are creating significant advancements in natural language processing, due to their strong text encoding/decoding ability and newly found emergent capability (*e.g.*, reasoning). While LLMs are mainly designed to process pure texts, there are many real-world scenarios where text data is associated with rich structure information in the form of graphs (*e.g.*, academic networks, and e-commerce networks) or scenarios where graph data are paired with rich textual information (*e.g.*, molecules with descriptions). Besides, although LLMs have shown their pure text-based reasoning ability, it is underexplored whether such ability can be generalized to graphs (*i.e.*, graph-based reasoning). In this paper, we provide a systematic review of scenarios and techniques related to large language models on graphs. We first summarize potential scenarios of



GraphRAG: Unlocking LLM discovery on narrative private data

Published February 13, 2024

By Jonathan Larson, Senior Principal Data Architect; Steven Truitt, Principal Program Manager

Share this page

Future Work - Cont'd

> Further explore relationship between
  - `dark matter&dark energy`
  - `dark matter&quantum universe`

Future Work - Cont'd

- Further explore relationship between
  - `dark matter&dark energy`
  - `dark matter&quantum universe`
- Recommend - papers, topics

Future Work - Cont'd

- Further explore relationship between
  - `dark matter&dark energy`
  - `dark matter&quantum universe`
- Recommend - papers, topics
- Link, Predict, Associate?

Article

Mathematical discoveries from program search with large language models

https://doi.org/10.1038/s41586-023-06924-6

Received: 12 August 2023

Accepted: 30 November 2023

Bernardino Romera-Paredes[1,4]✉, Mohammadamin Barekatain[1,4], Alexander Novikov[1,4], Matej Balog[1,4], M. Pawan Kumar[1,4], Emilien Dupont[1,4], Francisco J. R. Ruiz[1,4], Jordan S. Ellenberg[2], Pengming Wang[1], Omar Fawzi[3], Pushmeet Kohli[1]✉ & Alhussein Fawzi[1,4]✉

Thank you!

Contact

Deutsches Elektronen-
Synchrotron DESY

www.desy.de

Patrick L.S. Connor, **Antonin Sulc**
0000-0001-7767-778X
MCS
antonin.sulc@desy.de