

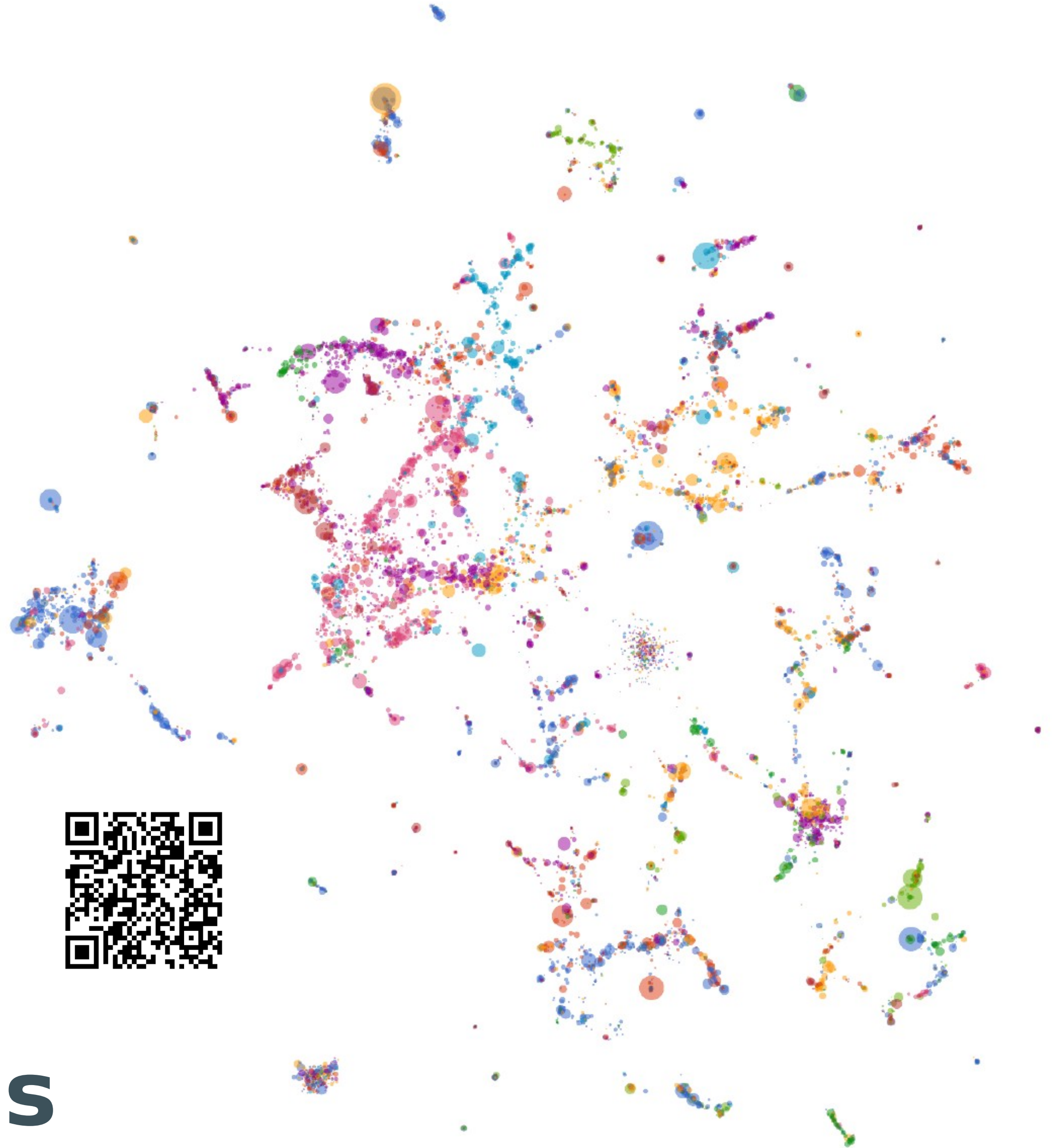
Revealing Connections in QCD with ML

Patrick L.S. CONNOR (CDCS, Universität Hamburg), Antonin ŠULC (HZB)

Goal $\mathcal{L}_{\text{QCD}} = \bar{\psi}_i (i\gamma^\mu (D_\mu)_{ij} - m \delta_{ij}) \psi_j - \frac{1}{4} G_{\mu\nu}^a G_a^{\mu\nu}$

- Complex structure behind apparent simplicity
- Need to identify hot topics or unexplored connections
- Handbook may quickly get outdated
- Develop complementary methods to bibliometrics

Citation graph



Graphs

- One colour per topic
- Each point corresponds to an article
- The narrower the more focused a topic is
- Blob size is proportional to the number of citations

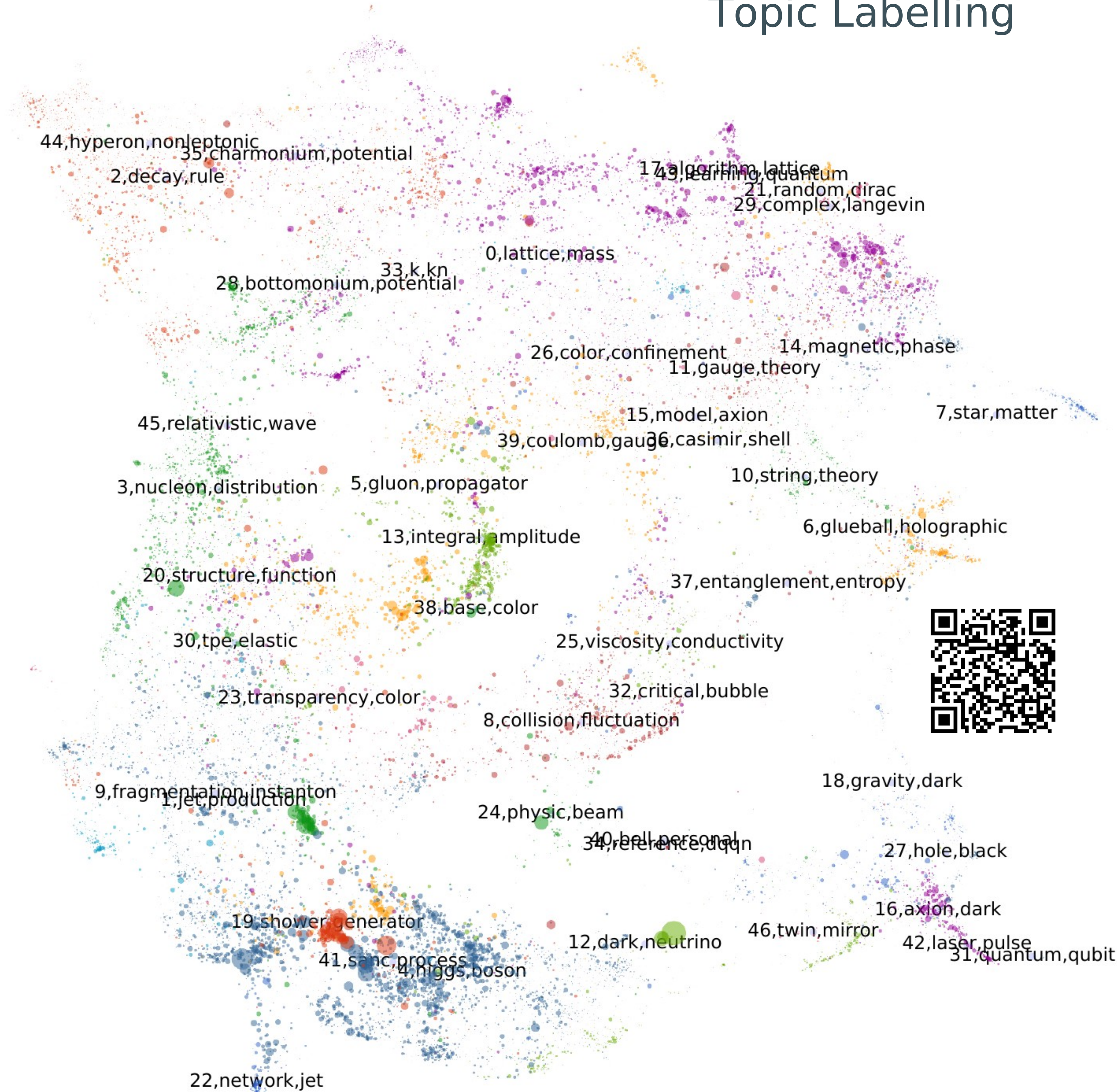
Conclusions

- Topics loosely related to QCD are found at the edges of the map (e.g. 7, 31)
- Existing and unexplored connections are visible in the embedding map
- These connections are distinct from the ones suggested by the citation graph

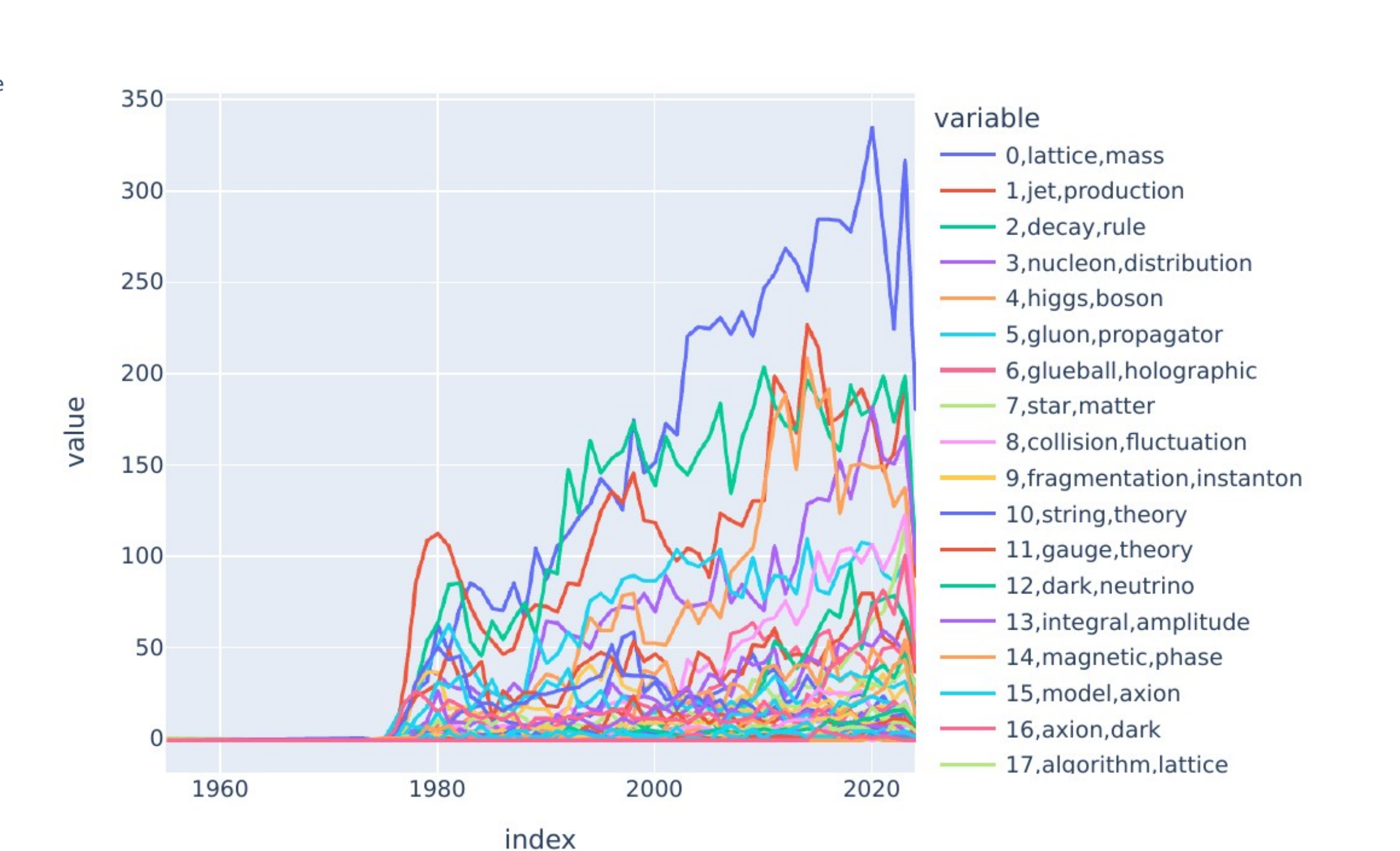
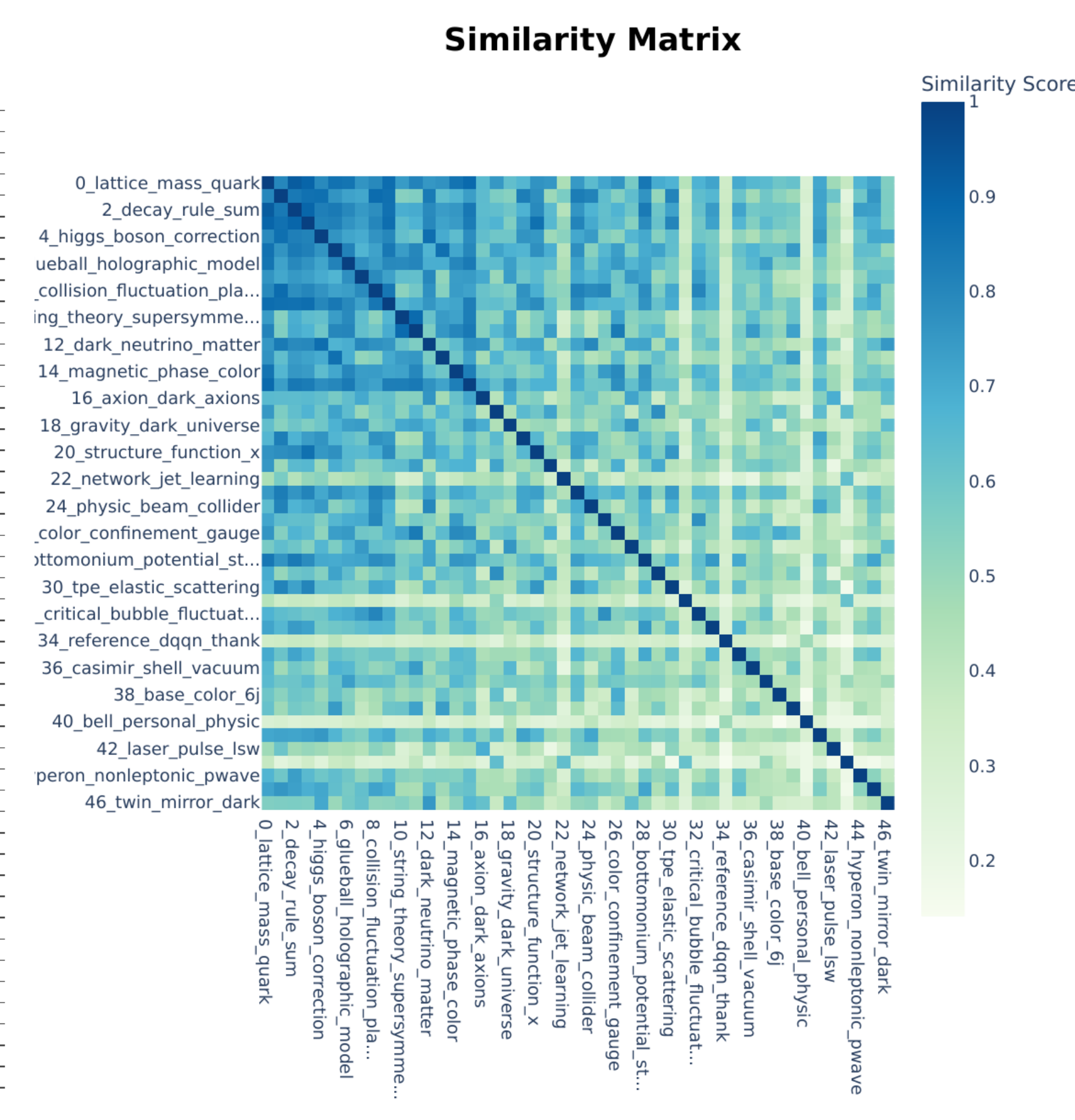
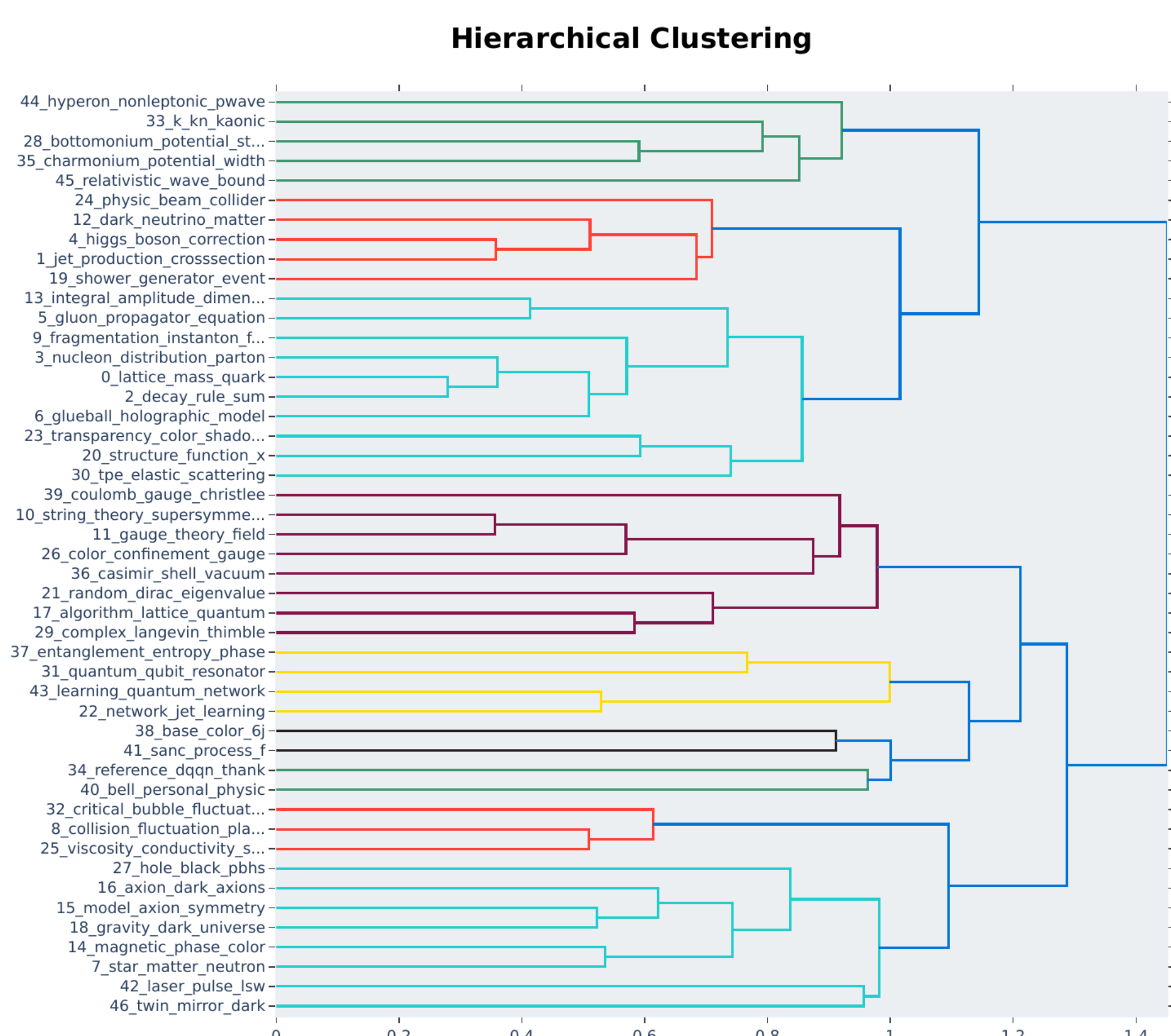
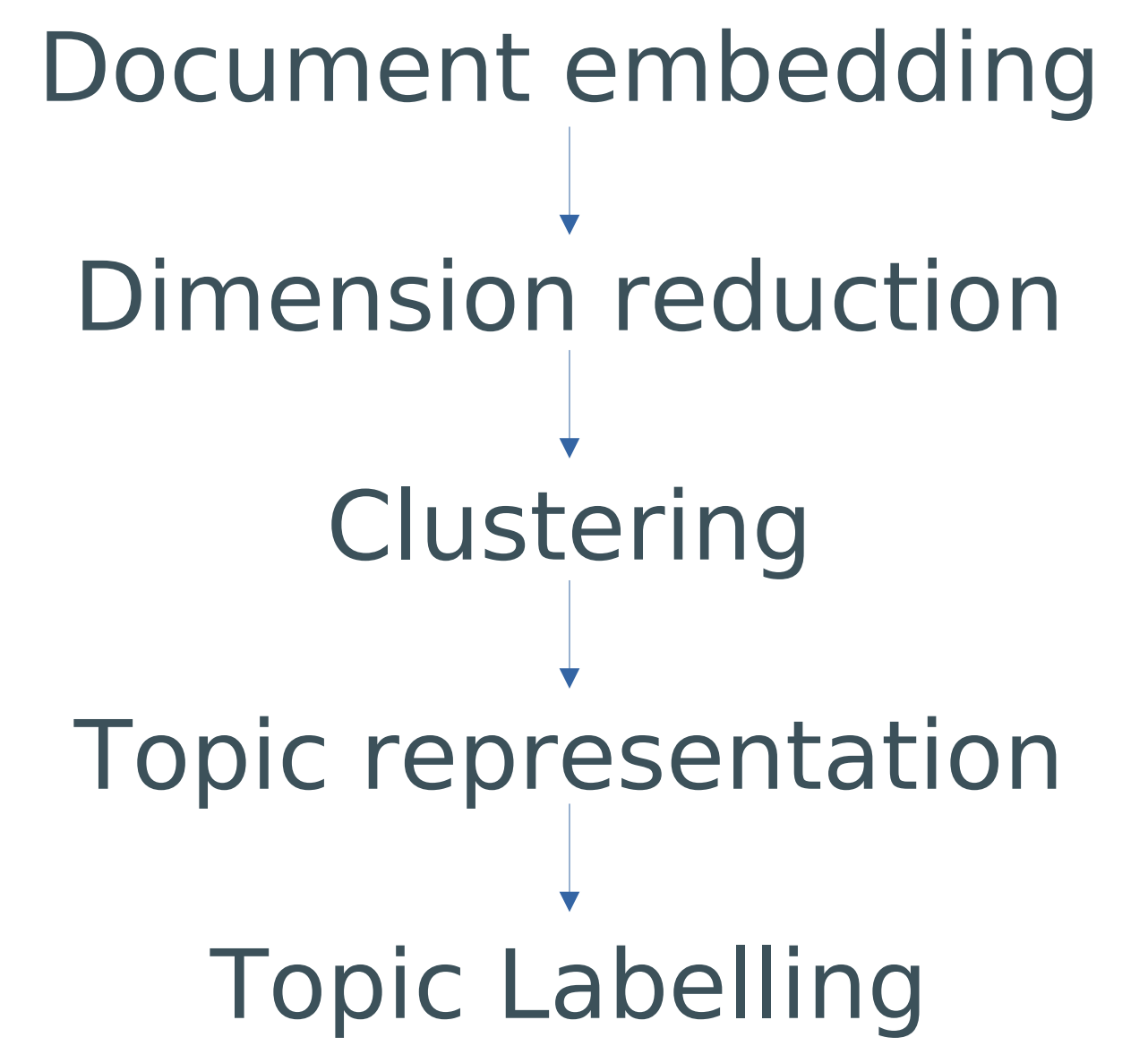
Data

- InspireHEP data base
- Only peer-reviewed articles
- Relying on abstracts only

Embedding



Method [1,2]



Prospects

- Identify key publications
- Identify anomalies
- Provide maps as a function time

[1] Grootendorst, Maarten. "BERTopic: Neural topic modeling with a class-based TF-IDF procedure." arXiv preprint arXiv:2203.05794 (2022)
 [2] Reimers, Nils, and Iryna Gurevych. "Sentence-bert: Sentence embeddings using siamese bert-networks." arXiv preprint arXiv:1908.10084 (2019)