Bibliometric Network Visualization with OpenAlex: An Analysis of the Quantum Computing Hardware Ecosystem

Martin Sand, Alain Mermoud and Julian Jang-Jaccard

martin.sand@ar.admin.ch
Cyber-Defence Campus, armasuisse Science and Technology, EPFL Innovation Park, Switzerland

As the field of bibliometrics gradually shifts towards open data, OpenAlex becomes an increasingly popular data source for analysis. However, bibliometric analysis through network visualization is still mostly conducted with proprietary databases, which impacts the reproducibility of the results. In this paper, we provide a feasibility study for analyzing the academic landscape surrounding emerging technologies, using the case of quantum computing hardware, in order to assess the construction of bibliographic networks with OpenAlex data and VOSviewer. The tools were able to produce subject similarity networks through bibliographic coupling, collaboration networks through coauthorship analysis and co-occurrence networks with OpenAlex concepts. However, this study highlights several limitations, notably the lack of interactability with static VOSviewer representations, which can conceal important data points, and no support for co-citation analysis.

Keywords: Bibliometrics; Scientometrics; Open-Source; OpenAlex; VOSviewer; Quantum

1. Introduction

As new potentially disruptive technologies emerge, experts and policymakers need efficient methods for assessing the scientific ecosystem before performing in-depth analyses of the subject matter. Without the proper tools and methodology, scanning the research landscape can be complex and time-consuming, leading, at worst, to an inaccurate or incomplete understanding of the relevant topic. Fortunately, over the years, information about published documents (e.g., abstracts, publication date, source, etc.) has been systematically stored in bibliographic databases such as Web of Science or SCOPUS. Accessible in digital form and over the Internet, these create an opportunity to evaluate the research state through bibliometric analysis. Bibliometrics is the field of studying bibliographic data quantitatively, providing insight such as the number of publications in a specific field over time, the number of citations per institution, the most cited works, etc. Thus, bibliometric analysis can produce a reliable overview of the scientific output relating to the relevant research topic.

Criticism has been levied against commonly used proprietary databases, notably regarding their cost and availability (Wilder, 2021), which impacts scientific reproducibility, their profit incentive and their structural bias against non-Western and non-native English-speaking publications (Tennant, 2020). These considerations have led institutions like the Sorbonne University or Leiden University to unsubscribe from their Web of Science license in favor of open bibliometric tools (Sorbonne, 2023; Brooks, 2023); furthermore, open science has been declared a policy priority by the European Commission to bolster research and innovation (European Commission, 2024), and widely promoted by the United-Kingdom Research and Innovation (UKRI, 2024) and the Swiss Confederation (SERI, 2024).

With the growing shift towards open data, this feasibility paper aims to evaluate a methodology for using OpenAlex to assess the research landscape around specific emerging technologies through network visualizations in a reproducible way, specifically using the VOSviewer software. This methodology will be illustrated with the analysis of quantum computing hardware, which is the most active research area in quantum technology, a field of research that has received much attention over the past years. More specifically, it will seek to visualize the semantic proximity between journals, the state of collaboration among countries and institutions and the research focus on quantum hardware.

One prime example of an open bibliographic database is OpenAlex, a fully open research catalogue developed by the non-profit OurResearch (Priem, 2022). Launched in January 2022, OpenAlex indexes over 250 million works from various sources, notably MAG, Crossref, PubMed and arXiv, and tracks associations between them through countries, institutions, co-authors, etc. While OpenAlex comprises a higher absolute number of indexed works than most proprietary databases, it appears to be lacking when it comes to abstracts (Culbert, 2024), which may influence any search based on the content of the works. Nevertheless, OpenAlex, while being open source, delivers comparable performance for bibliometric analysis, which has led to its increasing use by institutions and academics.

In the following section, the state of the research around the use of OpenAlex and the literature surrounding bibliographic network visualization and VOSviewer (2) are presented. Then, the methodology for accessing the relevant subset of OpenAlex data and visualizing it in VOSviewer (3) is presented. This methodology is then applied to extract results that clarify the research questions (4). The last section (5) discusses the method and its limitations.

2. Related Work

While many papers have performed evaluations of OpenAlex as a source of bibliographic data, notably in comparison to other proprietary databases (Culbert, 2024 ; Jiao, 2023 ; Velez-Estevez, 2023), it has also already been used for actual bibliometric analysis, e.g. for forecasting trends and monitoring technological convergence in encryption technologies (Ismail, 2023 ; Tavazzi, 2024), for measuring proximity between social and academic interest in research topics (Arroyo-Machado, 2023) or for analyzing cited reference patterns in institutions (Schares, 2023).

With the global body of research growing larger and larger every year, network visualization provides a powerful tool in bibliometric studies for analyzing relationships between large amount of produced material, e.g., citation or co-authorship networks (van Eck, 2014). VOSviewer has been a popular option for academic network visualization. VOSviewer is a freely available Java-based software tool released in 2010 by Nees Jan van Eck and Ludo Waltman from Leiden University. The program has been primarily developed to construct and visualize large bibliometric networks based on citations, bibliographic couplings, co-citations, co-occurrences, and co-authorship. These networks can be built upon bibliographic database files (e.g., Web of Science, Scopus, Dimensions, Lens, PubMed), reference manager files (e.g., RIS, EndNote, RefWorks), and API data (e.g., OpenAlex, Crossref, Europe PMC, etc.). It relies on distance-based mapping, which means that the distance between each node reflects their relatedness, e.g. how many sources they have in common, how many co-authors they share, etc. This makes it more suitable for visualizing large networks than the graph-based approach, where the relatedness is only represented by the weight of the links (van Eck, 2014). Furthermore, the variety of sources and representation parameters offer a multi-faceted approach to understanding which actors are contributing to the development of a particular technology (van Eck, 2011; van Eck, 2011).

However, the data used for these visualizations tend to be taken from proprietary sources, such as Web of Science or Scopus (Scheidsteger, 2021; Bitzenbauer, 2021), which can be explained by the fact that the emergence of platforms such as OpenAlex is rather recent. Indeed, the combined use of open data and network visualizations appears to be rare, even though tools like VOSviewer have functions to load OpenAlex data directly into their visualizations. This paper seeks to bridge the gap by studying the combination of OpenAlex and VOSviewer.

3. Methodology

3.1. Accessing the Data – OpenAlex API

OpenAlex provides an API to query lists of works, coupled with a search function to narrow them down to the relevant topic. Since VOSviewer uses this API to access OpenAlex data, it will be used to produce a set of works exploring the quantum hardware topic over the past ten years (2014-2023). Considering that this set will determine the validity of the analysis, the filters and the search terms must be carefully chosen: the scope of the chosen topic was defined through conversations with Swiss quantum experts at different events, notably Swissnex Quantum Summit (held on 14 October 2023, EPFL Rolex Learning Center), Quantum Industry Day in Switzerland (held on 16 October 2023, SwissTech Convensation Center), Swiss Quantum Days 2024 (held on 31 January to 2 February 2024, Villars-sur-Ollon), and Global Quantum Symposium 2024 (held 18-20 March, 2024, Quantum Basel). However, the methodology for defining the relevant search terms will vary depending on the width of the subject of interest.

Table 1. Chosen Filters for OpenAlex API request.

Chosen subject matter	Search Terms	Concepts[1]	API URL
Quantum Computing: Hardware	• Quantum Coherence Times • Quantum Gates • Quantum Circuits • Quantum Control Electronics • Quantum Error Correction • Quantum Processor Cooling	• Computer Science • Engineering	https://api.openalex.org/works?filter=title_and_abstract.search:"quantum coherence time"|"quantum gates"|"quantum circuits"|"quantum control electronics"|"quantum error correction"|"quantum processor cooling",publication_year:2014-2023,concept.id:c41008148|c127413603

The API will look for these search terms in the title and the abstracts but not the full text (*title_and_abstract.search*). They were chosen to be very specific and will be enclosed in quotes to return exact matches[2]. These measures will help ensure that only relevant works will

[1] OpenAlex concepts are semantic tags applied to a record with a reference to their corresponding Wikidata concept. The OpenAlex concepts are being deprecated in favor of "Topics". The methodology for this transfer can be found in OpenAlex Documentation. This analysis will continue using Concepts since the following tools are not necessarily current.

[2] This feature has returned inconsistent results during experimentation: while the higher-ranked results will always contain exact matches, the API will often send back lower-ranked records with only approximate matching; furthermore, it has been noticed that the usage of quotation marks will sometimes unexpectedly return more results

appear in the dataset, but they will also introduce a bias towards false negatives, because works treating the subject of quantum hardware without these exact search terms will be excluded. Further filtering will be achieved by only choosing works with specific level 0 concepts such as “Computer Science” and “Engineering” (Table 1). The query returns, at the time of writing, a total of 33'611 results.

3.2. Preprocessing the Data - VOSviewer

As mentioned, VOSviewer has a function to load OpenAlex API data directly into the software: the chosen API request will, therefore, be inserted as the bibliographic data source.

3.3. Visualizing the Data - VOSviewer

3.3.1. Semantic Proximity Between Journals

For analyzing the semantic proximity between journals, the bibliographic coupling method will be used on the unit of analysis "Sources", which encompasses the assigned sources where the works are hosted, e.g. journals, conferences, preprint repositories, etc. Bibliographic coupling is a form of citation analysis where two works are coupled if they cite the same reference; thus, the strength of the bibliographic coupling is an indicator of how probable it is that the two works are treating the same subject matter (van Eck, 2014). The link weight will be calculated using the full counting method, meaning each bibliographic coupling will contribute equally to the total link strength.

Furthermore, VOSviewer's clustering algorithm will be applied to the graph, helping identify groups that treat similar topics. The entities' size will be scaled by citation number to maximize the appearance of academically relevant journals.

3.3.2. Collaboration Between Organizations/Countries

Coauthorship networks of both organizations and countries will be constructed to visualize the state of collaboration in the quantum hardware field. A coauthorship network is fairly straightforward: two authors are linked in the network if they have co-authored a work, and the strength depends on how many works they have collaborated. The network will be built upon the fractional counting method, which means that each link will be weighted proportionally to the total number of authors for each work: this method is meant to prioritize exclusive collaboration between entities.

The entities' size will be scaled by citation number to maximize the apparition of academically relevant journals. VOSviewer's overlay visualization for these networks will be used with scores based on the average publication date. Concretely, each entity will be put on a heat gradient, showing which has increased their output in recent years.

3.3.3. Research Focus through Concept Analysis

Co-occurrence analysis will be used, specifically in regard to OpenAlex concepts, to identify the research focus over the past ten years, which may help identify future research trends. Two

---

than without them. This analysis will continue under the premise that quotation marks result in more precision, as described in the OpenAlex documentation, but it is important to take note of these abnormalities.

concepts co-occur when they are present together in the same works. The links will be constructed using the full counting method.

Generic concepts (level 0) will be filtered out to avoid overloading the graph. VOSviewer does not provide an automatic filtering function for the concept level, so this step must be done manually in the preview table. The remaining concepts will then be scaled by the total number of occurrences and presented with cluster visualization

4. Results

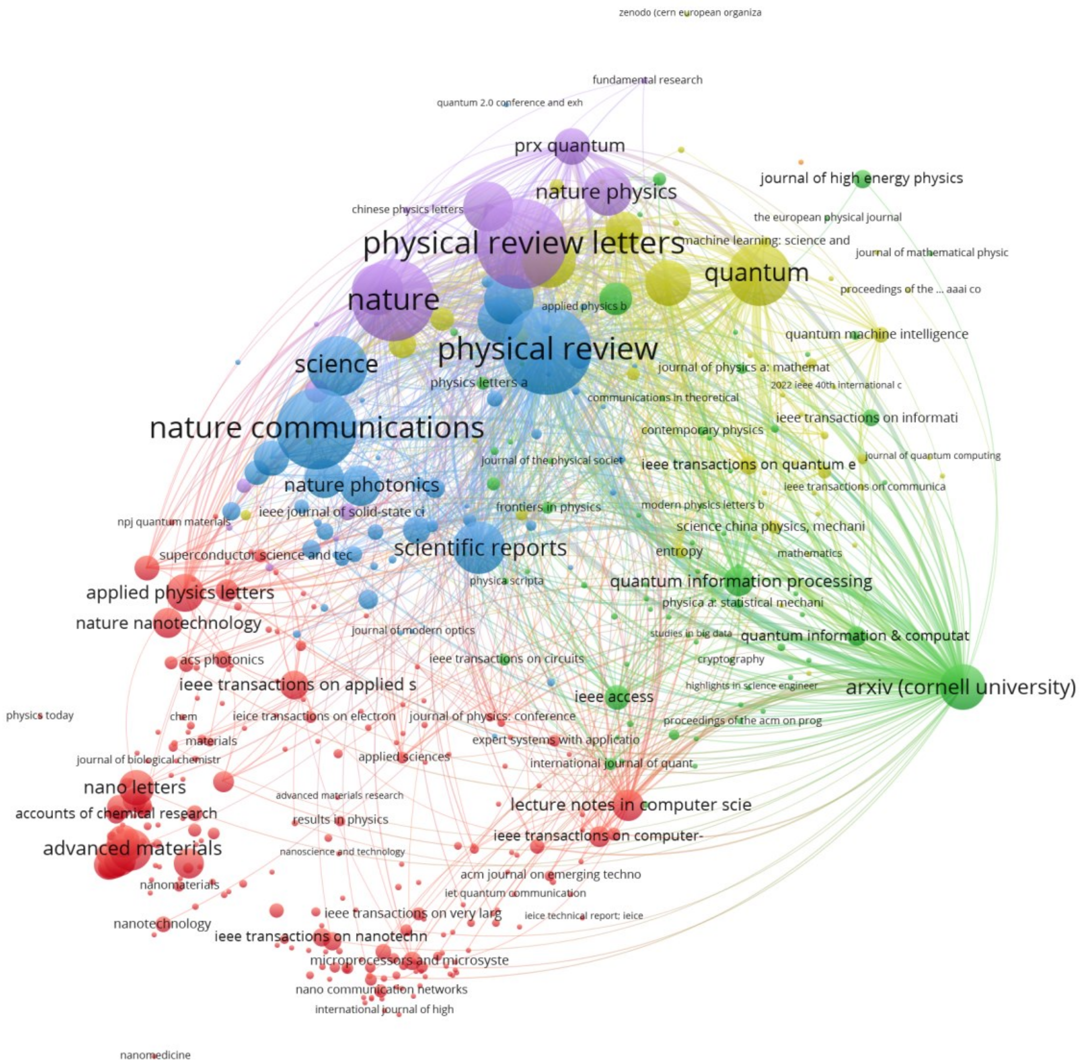
4.1. Similarity between journals and other sources

Bibliographic coupling of the sources (Figure 1) reveals multiple clusters with the highly cited sources in the field appearing in close relation. It is notable, however, that the most prolific source in this dataset is arXiv (green cluster), an open-access archive, with 7226 documents, dwarfing the highly cited journals by orders of magnitude. This pattern reoccurs between different topics, since arXiv hosts preprints, which can often be found in duplicates once they're published in other journals, and has an extremely high submission rate. The fact that arXiv is positioned further away from the rest of the highly cited journals suggests that it hosts publications on quantum computing hardware that treat the subject matter differently than the highly-cited journals. This result however is very dependent on the specific dataset, as subsequent visualizations have shown arXiv closer to the highly cited journals.

Table 2. Top sources ranked by citation count for the given API request

Journal/source	Citations	Documents
Nature Communications	20’268	237
Physical Review Letters (Print)	13’620	276
Nature	13’402	37
Physical Review Letters	13’296	245
Physical Review A	12’161	481
Quantum (Vienna)	9’874	212
Nature (London)	9’033	51
NPJ Quantum Information	8’839	231
New Journal of Physics	7’648	212
Science (NYC)	6’608	30

Figure 1: Network visualization in VOSviewer based on bibliographic coupling between sources (2014-2023) related to quantum computing (hardware). The entities are scaled by citation number; the clustering is done automatically by VOSviewer.



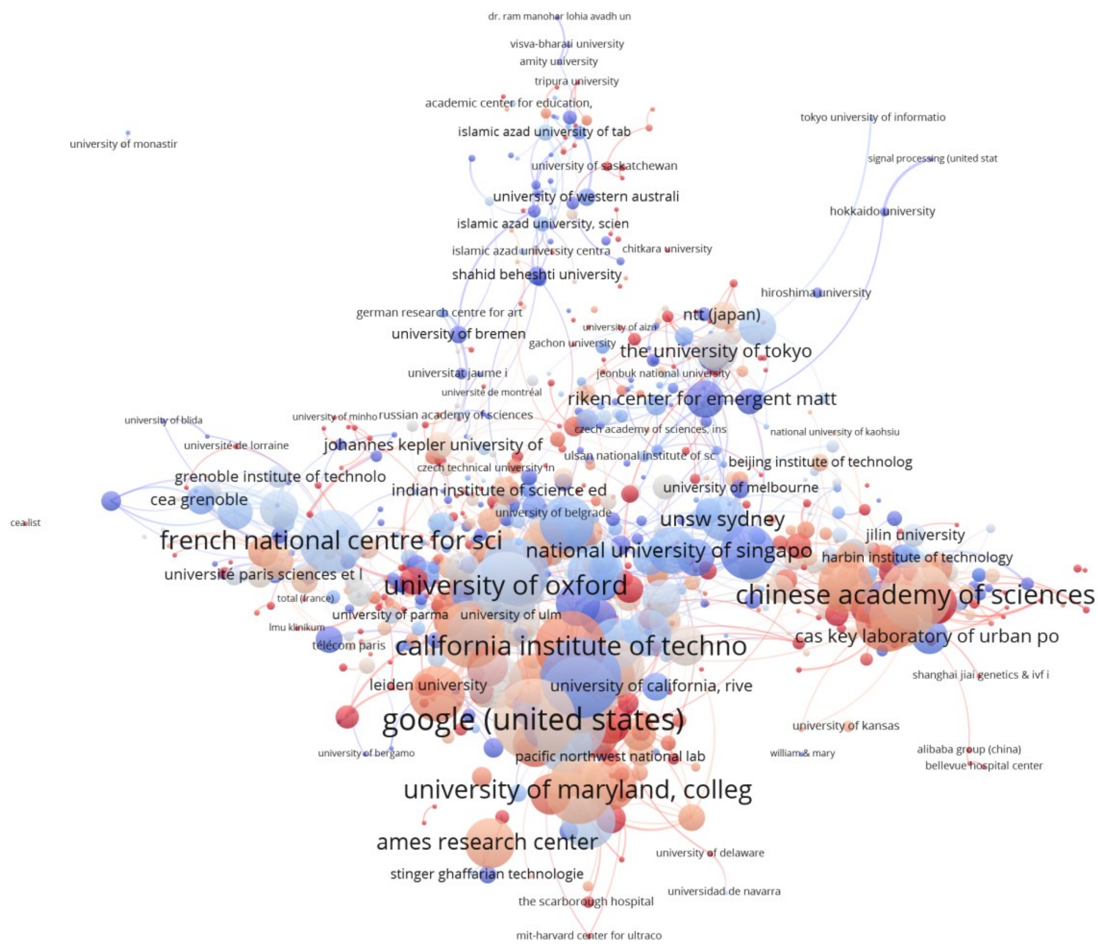
4.2. Collaboration between organizations and countries

Analyzing co-authorship networks of academic organizations across the quantum hardware field (Figure 2) reveals multiple collaboration clusters. These clusters tend to align with the geographic or linguistic proximity of the institutions, e.g., Harvard, MIT, and Google, Delft and Qutech, ETH and EPFL, etc. Additionally, while many institutions are centered around the American academic organizations, two isolated clusters appear further away and are centered around French and Chinese organizations.

Table 3. Top institutions ranked by citation count for the given API request

Organization	Citations	Documents
Google (US)	15’227	142
University of California (Santa-Barbara)	13’851	175
Massachusetts Institute of Technology	12’695	337
Chinese Academy of Sciences	11’623	420
University of Oxford	11’091	260
California Institute of Technology	10’949	200
Delft University of Technology	10’701	298
ETH Zürich	10’643	232
Harvard University	10’535	143
University of Maryland (College Park)	10’030	244

Figure 2: Network visualization in VOSviewer based on co-authorship between organizations (2014-2023) related to quantum computing (hardware). The entities are scaled by citation number; the heat gradient represents the average publication year and ranges from 2018 (cool) to 2021 (warm).

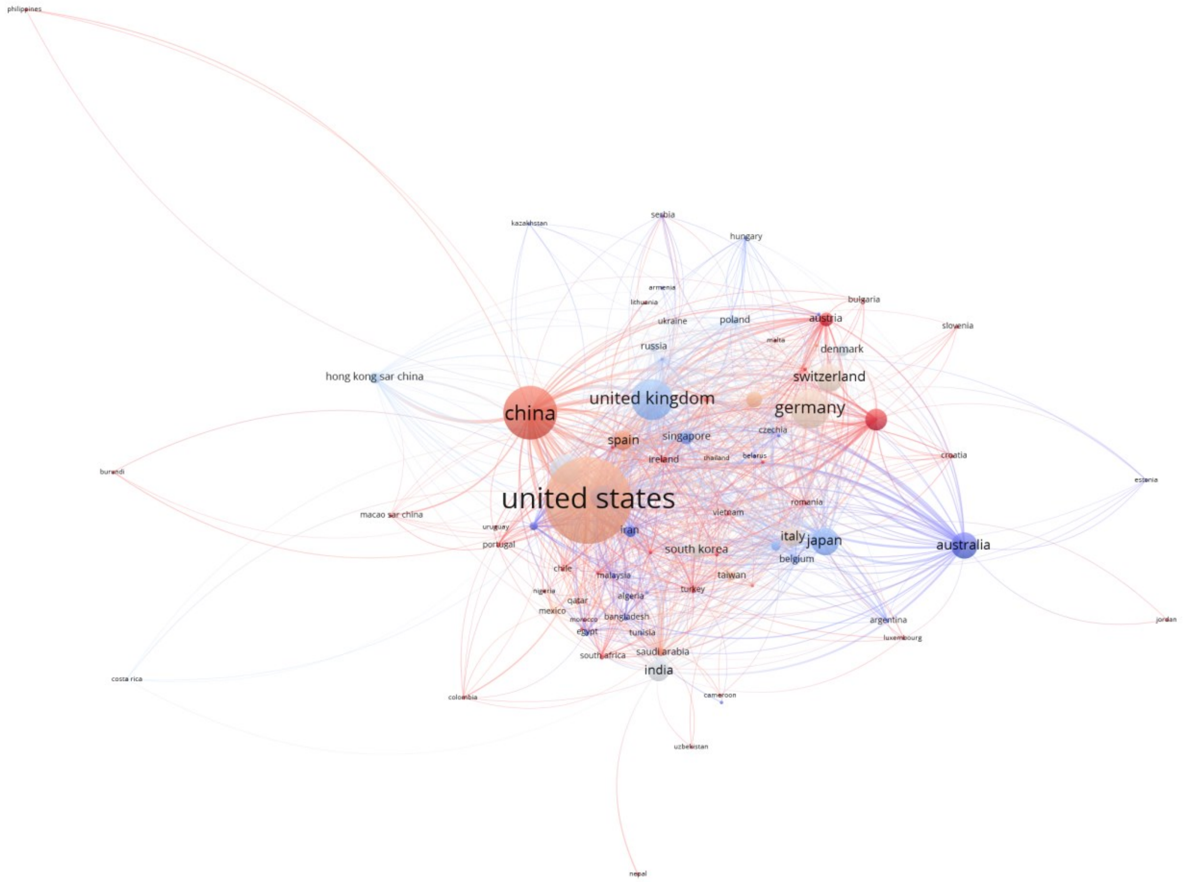


Co-authorship networks on countries (Figure 3) provide a more general view on collaboration. While some clustering patterns map onto those in the organization networks, e.g., Switzerland, Germany, and the Netherlands tend to be closely affiliated, there are some notable differences, e.g., China and France being very close to the United States. These differences are due to the overall distribution of the collaborations of all organizations within a country: while prominent organizations from a certain country may form a cluster with prominent organizations from a different country, e.g., the University of Oxford and ETH Zürich, taking into account every other organization might move these countries further away from each other.

Table 4. Top countries ranked by citation count for the given API request

Country	**Citations**	**Documents**
United-States	159’704	6’031
China	70’158	3’977
United-Kingdom	46’853	1’586
Germany	42’965	1’819
Canada	27’308	1’000
Japan	24’534	1’562
Switzerland	24’010	654
Australia	19’233	697
Netherlands	17’122	575
France	16’889	940

Figure 3: Network visualization in VOSviewer based on co-authorship between countries (2014-2023) related to quantum computing (hardware). The entities are scaled by citation number; the heat gradient represents the average publication year and ranges from 2019 (cool) to 2020 (warm).



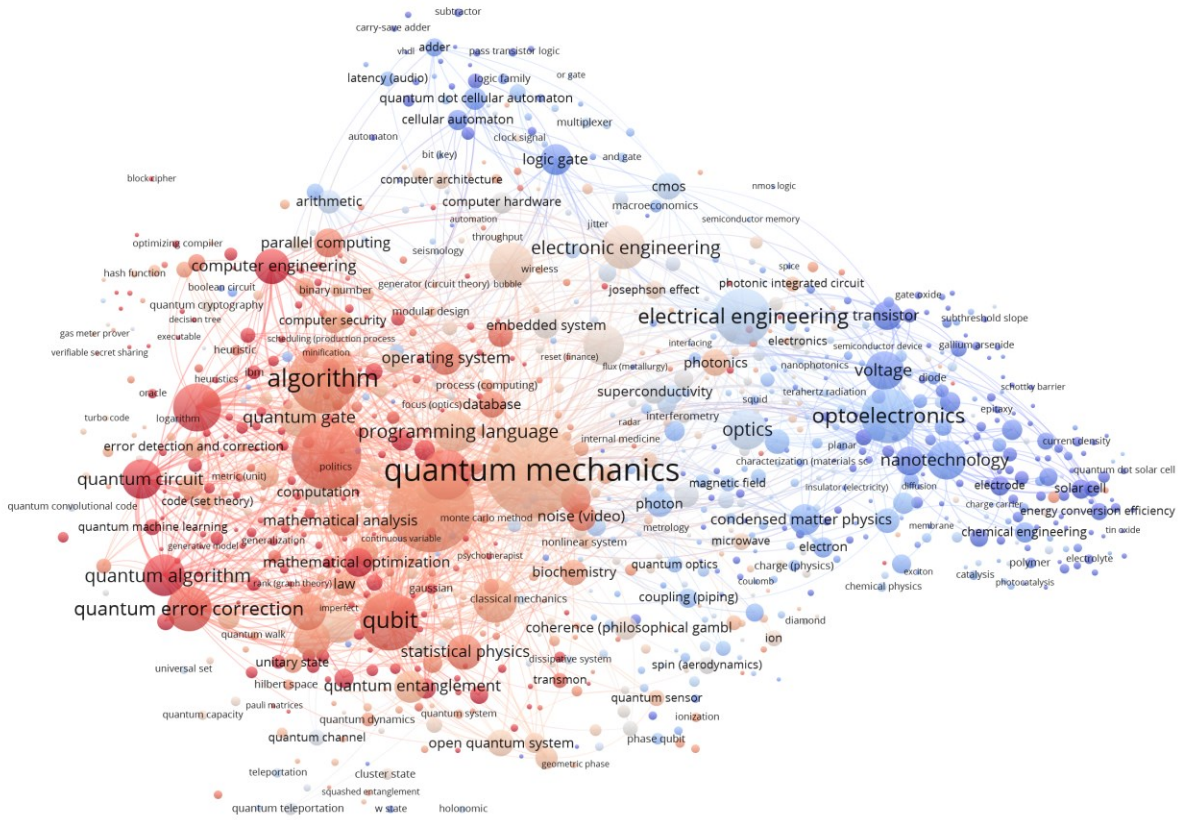
4.3. Research Focus Through OpenAlex Concepts

A co-occurrence network based on a curated list of OpenAlex concepts reveals the most popular topics treated over the past ten years and how they relate to each other (Figure 4). Furthermore, with an overlay visualization based on a heat gradient, it can be observed that topics related to IT (e.g., "qubit," "quantum computer," "algorithm," "operating system," etc.) appear more relevant in recent years than topics related to electronics (e.g., "optoelectronics," "electrical engineering," "transistor," "logic gate," etc.).

Table 5. Top concepts ranked by occurrence count for the given API request

Concept	Occurrences
Quantum Mechanics	24’697
Quantum	20’607
Quantum Computer	12’887
Algorithm	12’206
Qubit	8’980
Electrical Engineering	8’335
Optoelectronics	7’584
Artificial Intelligence	6’374
Theoretical Computer Science	5’874
Quantum Error Correction	4’947

Figure 4: Network visualization in VOSviewer based on co-occurrence between OpenAlex concepts (2014-2023) related to quantum computing (hardware). The entities are scaled by total number of occurrences; the heat gradient represents the average publication year and ranges from 2018 (cool) to 2021 (warm).



5. Discussion and Limitations

In this paper, three core insights in the field of quantum hardware were extracted from the OpenAlex data. The first is the similarity between journals in which subject matter they tend to publish: bibliographic coupling can help identify which journals intersect and where specific expertise lies within the field. The second relates to which institutions and countries have the strongest ties in the academic field on the chosen topic and which have intensified their publication rate in recent years. The third shows recurring topics in the chosen subset, and which of them have gained more traction in the newest publications, forecasting future trends in the field. It is important to note that these insights were chosen to give a broad overview of the quantum computing landscape and illustrate how combining OpenAlex and VOSviewer can help answer these questions, not as a fixed method for tackling the academic ecosystem surrounding an emerging technology.

While screen-shooting the graphs directly from the software allows for a quick glance at the ecosystem, the added value for exploratory research into a specific subject matter lies in the interactivity with the network, i.e., the ability to zoom into more concentrated clusters or to isolate connections for one specific cluster. It could be argued that VOSviewer provides incomplete data visualization, since it is not able to represent a full account of the useful information in a static image, e.g. when important entities overlap. However, using the software with open data should provide easily reproducible networks that can be accessed by anyone for evaluation or deeper analysis.

Unfortunately, since the expected added value of using open sources and tools is more reproducibility, the OpenAlex integration into VOSviewer suffers from non-reproducible results: since the API query works on live data, it will return different results at different times. A solution would be to download a snapshot of the OpenAlex data, parse it to generate a CSV that can be loaded into VOSviewer. Otherwise, tools like *openalexR* or a simple Python script allows to generate data frames from an API request that can be exported to CSV directly. An example is demonstrated on the OurResearch platform in the form of a webinar tutorial on how to use the OpenAlex API (OurResearch, 2024).

Furthermore, these tools are subject to several smaller limitations that need to be considered before starting the analysis and during the interpretation of the results. An early consideration is that combining VOSviewer with OpenAlex data imposes a 50k limit on the number of works that can be loaded into any visualization, which was not a problem for the chosen use case, but could be easily reached when analyzing broad subject matters over long periods of time.

Another consideration is the use of citation number and document count to evaluate the relevance of each entity to the subject matter. The VOSviewer integration of OpenAlex offers no possibility of normalizing by the total number of publications, which introduces a bias towards already prolific organizations or countries.

VOSviewer's co-occurrence network, when using OpenAlex data, is built with Wikidata concepts. Since each work in the API request has several concepts attached to it, they will all contribute equally to the co-occurrence graph. However, OpenAlex provides a relevance score for each concept, determining how compatible the concept is with the publication. In this case, an alternative to the co-occurrence network could be, for instance, a concept cloud as presented in the *openalexR* package docs[3] which scales the concepts based on the sum of their scores.

When trying to analyze the average publication years of each entity with the heat overlay, it is important to note some limitations to the interpretation: warmer colors in an entity should be interpreted as a recent intensification of their activity, and conclusions about which entity has

[3] https://docs.ropensci.org/openalexR/

been contributing more in recent years can only be established between entities with a roughly similar publication output.

Co-citation analysis is a form of citation analysis that helps infer subject similarity between papers. Unlike bibliographic coupling, which links works that share references, co-citation links publications based on whether they are cited in common in other publications. This method is dynamic since the networks can change over time and, therefore, is more suited for forecasting. However, OpenAlex data doesn't support it.

6. Conclusion

As open-source bibliometrics continues to gain attention, understanding the capabilities and limitations of the databases and tools openly available today becomes a necessary step for academics and decision-makers before using them to explore long-standing or emerging technologies. OpenAlex, despite its current lack in abstract metadata and incompatibility with co-citation analysis, is bound to become a popular data source for scientometric and bibliometric studies, and by extension, increasingly used for network analysis. Combining the open data with tools like VOSviewer provides the necessary means to visualize these networks in a reproducible way, allowing for better reliability and innovation from the scientific output.

However, the OpenAlex integration into VOSviewer does not allow for snapshots of the queried data directly. This can be remedied by developing a means for saving the results obtained by the API query in the form of a bibliographic data file, as described in the VOSviewer manual. To this end, tools like *openalexR*, which provides a way of manipulating the data in an OpenAlex API query, need to be explored in the future.

Acknowledgements

We thank Dr. Gaël Bernard from the EPFL Academic Data Unit for his valuable insights that helped improve the quality of this paper.

This paper has been inspected with Grammarly, an AI typing assistant that has helped detect minor writing mistakes and check for signs of plagiarism.

References

Arroyo-Machado, W., & Costas, R. (2023, April). Do popular research topics attract the most social attention? A first proposal based on OpenAlex and Wikipedia. In *27th International Conference on Science, Technology and Innovation Indicators (STI 2023)*. International Conference on Science, Technology and Innovation Indicators.

Bitzenbauer, P. (2021). Quantum physics education research over the last two decades: A bibliometric analysis. *Education Sciences*, *11*(11), 699.

Culbert, J., Hobert, A., Jahn, N., Haupka, N., Schmidt, M., Donner, P., & Mayr, P. (2024). Reference Coverage Analysis of OpenAlex compared to Web of Science and Scopus. *arXiv preprint arXiv:2401.16359*.

European Commission (2024, 15 March). *Open science*. https://research-and-innovation.ec.europa.eu/strategy/strategy-2020-2024/our-digital-future/open-science_en

Ismail, S., Mermoud, A., Marechal, L., Orso, S., & David, D. P. (2023, April). Capturing Trends Using OpenAlex and Wikipedia Page Views as Science Indicators: The Case of Data Protection and Encryption Technologies. In *27th International Conference on Science, Technology and Innovation Indicators (STI 2023)*. International Conference on Science, Technology and Innovation Indicators.

James Brooks (2023). *Leiden rankings to add open-source version in 2024*. ResearchProfessional News. https://www.researchprofessionalnews.com/rr-news-europe-universities-2023-9-leiden-rankings-to-add-open-source-version-in-2024/

Jiao, C., Li, K., & Fang, Z. (2023). How are exclusively data journals indexed in major scholarly databases? An examination of four databases. *Scientific Data*, *10*(1), 737.

OurResearch. (2024, April 25). *Webinar: The openalex API part 2: Diving in*. YouTube. https://www.youtube.com/watch?v=DLKUgbw7FV4

Priem, J., Piwowar, H., & Orr, R. (2022). *OpenAlex: A fully-open index of scholarly works, authors, venues, institutions, and concepts*. ArXiv. https://arxiv.org/abs/2205.01833

Schares, E., & Mierz, S. (2023, April). Using OpenAlex to Analyse Cited Reference Patterns. In *27th International Conference on Science, Technology and Innovation Indicators (STI 2023)*. International Conference on Science, Technology and Innovation Indicators.

Scheidsteger, T., Haunschild, R., Bornmann, L., & Ettl, C. (2021). Bibliometric analysis in the field of quantum technology. *Quantum Reports*, *3*(3), 549-575.

SERI (2024, 15 March). *Open science*. https://www.sbfi.admin.ch/sbfi/en/home/ihe/higher-education/hochschulpolitische-themen/open-science.html

Sorbonne Université (2023). *Sorbonne university unsubscribes from the web of science*. https://www.sorbonne-universite.fr/en/news/sorbonne-university-unsubscribes-web-science
Tennant, Jonathan. (2020). Web of Science and Scopus are not global databases of knowledge. European Science Editing. 46. 10.3897/ese.2020.e51987.

Tavazzi, A., David, D. P., Jang-Jaccard, J., & Mermoud, A. (2024). Measuring Technological Convergence in Encryption Technologies with Proximity Indices: A Text Mining and Bibliometric Analysis using OpenAlex. *arXiv preprint arXiv:2403.01601*.

UK Research and Innovation (2024, 15 March). *Open research.* https://www.ukri.org/what-we-do/supporting-healthy-research-and-innovation-culture/open-research/

Van Eck, N. J., & Waltman, L. (2014). Visualizing bibliometric networks. In *Measuring scholarly impact: Methods and practice* (pp. 285-320). Cham: Springer International Publishing.

Van Eck, N. J., & Waltman, L. (2011). Text mining and visualization using VOSviewer. *arXiv preprint arXiv:1109.2058*.

Van Eck, N. J., & Waltman, L. (2011). VOSviewer manual. *Manual for VOSviewer version*, *1*(0).

Velez-Estevez, A., Pérez, I. J., García-Sánchez, P., Moral-Munoz, J. A., & Cobo, M. J. (2023). New trends in bibliometric APIs: A comparative analysis. *Information Processing & Management*, *60*(4), 103385.

Wilder, E.I. and Walters, W.H., 2021. Using Conventional Bibliographic Databases for Social Science Research: Web of Science and Scopus are not the Only Options. *Scholarly Assessment Reports*, 3(1), p.4.DOI: https://doi.org/10.29024/sar.36