

Building a clinical and academic health research network using data science and artificial intelligence

Hope James Ekpa

H.Ekpa@hull.ac.uk

University of Hull

Thomas Phillips

University of Hull

William Stephen Jones

University of Hull

Research Article

Keywords: Research Network, Data Science, Artificial Intelligence

Posted Date: August 20th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10345362/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

Abstract

Background

This study explored the use of data science and artificial intelligence (AI) to develop a clinical and academic health research network focused on alcohol, addiction, and mental health research in Northern England. The aim was to create a scalable and systematic alternative to traditional manual networking approaches.

Methods

We identified potential members of the research network by systematic searching of PubMed/MEDLINE (via Metapub API using Python) for peer-reviewed articles published between 2020 and 2025, focusing on 26 journals selected for their relevance to alcohol, addiction, and mental health related research. From the retrieved articles, we extracted metadata (including author names and affiliations), performed automatic disambiguation using ORCID data, and used a graph-based model to construct and visualise the potential research network.

Results

The approach identified 420 unique authors. Regional analysis showed the Yorkshire and Humber had the highest output (181 publications, 45% of the total), followed by the North West (162) and the North East (55). The University of Sheffield had the highest number of publications (94) followed by the University of York (57) and the University of Liverpool (52). Network analysis revealed moderate interconnectivity with authors collaborating with average of seven co-authors and there is also evidence of inter-university collaborations.

Conclusion

The study demonstrates the feasibility of using AI and data science methods to systematically identify and develop clinical and academic research networks at scale. We consider this to be a substantial, scalable and systematic advancement compared to the manual, *ad hoc* techniques for developing a research network, which are time-consuming, inconsistent, and potentially biased, and so not fit for purpose. Future work should look to address the limitations of this research by capturing more publications, performing author disambiguation ORCID ID (and via alternative methods), embedding this approach in an open-source app (for ease of use for all researchers), and exploring techniques for quantitative validation. Additionally, future iterations should consider ways identify relevant stakeholders who may not be actively publishing but are engaged in the research ecosystem.

Background

Research networks and alliances are organized collaborations between researchers, clinicians, departments, institutions, industry partners, countries, and other relevant entities working on particular research are (Andres, 2014; Li et al., 2018). In healthcare, these networks are particularly valuable because they bring together broad expertise, resources, and patient populations to address complex health problems that cannot be solved by individuals or individual groups alone. They enable the building of research capacity through knowledge exchange, resource sharing, coordinated data collection and patient recruitment (e.g. multisite clinical studies), rapid evidence generation, staff development, joint grant applications, and translation of findings into clinical practice and policy (Brown et al., 2016).

The impact of research networks has been demonstrated in multiple domains. During the COVID-19 pandemic, rapid knowledge exchange between clinicians, epidemiologists, public health agencies, and other stakeholders (largely on X, formally Twitter) allowed for the rapid evaluation of treatments and the development of evidence-based guidelines (Breiman et al., 2025). The National Institute of Health and Care Research (NIHR) Statistics Group brings together methodological experts to advance the use of robust statistical approaches in health research. It provides a platform for sharing best practice, promoting novel methods, and fostering collaboration between statisticians, clinicians, and scientists (NIHR, n.d.). In addiction research, collaborative networks such as the Centre for Addiction and Mental Health Research (CAMHR), the NIHR Mental Health Research Group at the University of Hull, have supported and hosted joint grant applications, cross-institutional studies, and policy engagement to address regional substance use challenges (NIHR Award Ref: NIHR207584). The Northern Addiction Research Alliance (NARA) is one such initiative hosted by CAMHR and funded by the NIHR Yorkshire and Humber Research Delivery Network with the specific aim of developing research collaborations across the North of England.

Despite their value, building and sustaining research networks presents significant challenges (Scott et al., 2020). Traditional approaches often rely on manually mapping collaborators through personal contacts, surveys, or institutional records. These methods are labor-intensive, prone to bias, and quickly become outdated as teams evolve and new expertise emerge (Marsden, 2011). They also struggle to scale, that is, identifying potential collaborators across hundreds or thousands of websites or publications is rarely feasible without automation. Furthermore, developing and maintaining such networks requires substantial time and financial investment, resources that are often scarce or unavailable.

In recent years, advances in data science and artificial intelligence (AI) have offered new possibilities for overcoming some of these challenges. Xu et al (2020) developed the PubMed Knowledge Graph, an open, large-scale, and integrated dataset that links biomedical concepts, authors, affiliations, funding information, and publications from over 29 million PubMed abstracts. The researchers used a domain-adapted transformer, deep learning model (BioBERT) which automatically extracts and normalizes biomedical entities (e.g. genes, diseases, drugs, species, mutations), uniquely identifies authors through

high-accuracy name disambiguation, and enriches records with career histories, geographic locations, and funding data. This enables automated, fine-grained mapping of research expertise and collaborations at scale, across disciplines and institutions, and over time. Priem et al (2020) developed OpenAlex, a fully open scholarly metadata index linking works, authors, venues, institutions, and concepts. It draws data from Crossref, PubMed, repositories, and the Microsoft Academic Graph, using automated parsing, normalization, and disambiguation. The dataset covers 209M works (journal articles, theses, books, datasets), 213M authors, and 65k concepts. ORCID aids author disambiguation, supplemented by publication and citation histories to improve accuracy (OpenAlex, 2022).

Alongside large-scale scholarly metadata infrastructures such as Open Alex and PubMed Knowledge graph, several researcher-finder platforms have also existed to connect and find experts. Examples include Scinapse Expert Finder (Scinapse, 2026), Profiles Research Network Software (Harvard) (Profile Research Network, 2026), and ResearchGate (ResearchGate, 2026). These systems allow users to search for experts by topic, institution, or publication record, and in some cases recommend potential collaborators or reviewers based on shared research interests.

However, while these studies and systems have contributed to components of developing a research network, they have not gained significant traction across clinical and academic settings, and key gaps remain. For instance, most existing systems focus on broad, high-level mappings of research activity and lack functionality for building networks within narrow, specialized areas. They do not allow for filtering author affiliations to specific national or institutional scopes, such as UK Higher Education Institutions. Some of the systems are not open-source and so their costs can be prohibitive to individual researchers, particularly early career researchers, who most need support with network development.

In this study, we present a scalable, automated method for identifying and profiling researchers (academics and clinicians) who have published peer-reviewed articles within a defined domain, as a means of mapping potential members of a thematic research network. To demonstrate the method, we apply it to a case study in alcohol, addiction, and mental health related research between 2020 and 2025, using Natural Language Processing (NLP) and bibliographic databases.

Methods

The AI and Data Science Methodological Approach

Figure 1 illustrates the methodological steps involved in the proposed AI and data science approach to developing a research network, including identifying the domain, journal selection, metadata extraction, author and affiliation filtering, ORCID retrieval, and the construction of the research network.

[Insert Figure 1]

Figure 1. Conceptual methodological workflow of the AI and data science approach for building the research network.

The diagram outlines sequential steps starting from domain identification (e.g., Mental Health, Addiction, Alcoholism), followed by selection of relevant journals, retrieval and extraction of articles and metadata (e.g., via PubMed and parsing), retrieval of authors' ORCID IDs using the ORCID API, filtering for institutional affiliations (e.g., UK Higher Education Institutions), and finally building of the research network through nodes and edges.

Identify Domain

The first step in this research is to identify the domain from which publications will be retrieved. Defining a domain is essential as it helps narrow the search space and ensures relevant literature is captured. A well-defined domain guides the scope of the research, reducing noise from unrelated fields and improving the accuracy of subsequent analyses such as the network construction.

Select Relevant Journal(s)

After identifying the domain, the next step is to identify journals that frequently publish research within this area. Doing this ensures that the publications to be retrieved are focused and grounded in the truth of publications in the domain.

Query Database

Once the journals are identified, the next step is to define a list of research databases to retrieve articles from. In this research we focused on utilizing PubMed as a case study. PubMed is a free resource provided by the U.S. National Library of Medicine (NLM) that primarily provides access to the Medline database of citations and abstracts for biomedical literature. It provides comprehensive metadata in XML format for peer-reviewed biomedical literature, including author names, affiliations, publication titles, journal names, and publication dates.

To interact with PubMed, one needs to utilize an API. In this research, the Metapub API was utilized, as it is the most reliable among the available options (Metapub, 2024). To ensure fair use, PubMed limits the number of articles retrieved per query to 10,000. So, to achieve a higher number of articles, the search query must be executed in two six-month intervals per year (January–June and July–December). The results for each half-year period can then be subsequently consolidated into a single dataset.

Retrieving Articles and Extracting Metadata

After the search query is executed, the retrieved PubMed records, in XML format, must be parsed to extract relevant metadata. The XML files contain structured information about each publication, including authors, affiliations, publication year, and unique identifiers.

Python's `lxml.etree` package was used to parse the XML structures in this research. Key attributes extracted include: Author information from `<Author>` elements; full author names are constructed using the `<ForeName>` and `<LastName>` tags; primary affiliation obtained from the first `<Affiliation>` tag under `<AffiliationInfo>`; research title extracted from the corresponding title field; PMID, the unique identifier for each publication, is used to detect and remove duplicates.

Each article's extracted metadata was stored as a dictionary object and appended to a list for further processing.

Filtering for Institutional Affiliation

In bibliometric analysis, it is often necessary to apply filtering to isolate a specific sector of interest, such as government, industry, or academia, to ensure the analysis aligns with the research objective. This general approach allows for a focused examination of a specific institutional dynamic.

In this research, the dataset was filtered to retain only publications authored by researchers affiliated with Higher Education Institutions (HEIs). This filtering ensures that the final dataset represents the specific ground truth required for developing the research collaboration network.

Retrieve Authors' ORCID IDs

After filtering the dataset, the next critical step is to resolve author ambiguity. In bibliometrics, the problem of multiple authors sharing the same name can lead to inaccurate network representation. To address this, ORCID (Open Researcher and Contributor ID) IDs were retrieved. ORCID is a free, persistent identifier utilized to uniquely identify researchers even if researchers have similar names as they engage in research, scholarship, and innovation activities. It is increasingly a standard requirement for submissions to many major journals, ensuring widespread adoption in recent literature (Carter et al., 2017).

To obtain these IDs automatically, each author's last name, first name, and affiliation were combined to query the ORCID database. This retrieval process identified the most likely ORCID record for each author. The essence of doing this is to ensure that no authors are duplicated when constructing the research network. However, it is important to acknowledge the limitations of this approach. Not all ORCID IDs will be retrieved, and automated retrieval based on name and affiliation can occasionally yield false positives or miss records if the affiliation or name is incomplete.

Build a Research Network

Once the ORCID IDs were successfully retrieved, the next step is to build the research network or co-authorship network. This was built by utilizing a graph-based approach consisting of nodes (entities) and edges (relationships) between the nodes. Each unique author, identified by their ORCID ID, was

represented as a node within the network. Nodes were also color-coded according to their institutional affiliation to allow for visual interpretation of university-level collaboration and clustering.

The edges between nodes were created using the shared publication identifiers (PMIDs). An edge represents a co-authorship relationship, formed when two or more authors collaborated on the same article. This structure enabled the development of a detailed research collaboration network that illustrated the connectivity, collaboration frequency, and institutional relationships among the researchers.

Case study

Scope and rationale

To demonstrate this methodological approach, we applied it to the Northern Addiction Research Alliance (NARA) case study, shown in Table 1, which focuses on identifying potential early-career and senior researchers and clinicians, and developing a research network of individuals who have published on alcohol, addiction, and mental health research during the 2020–2025 period across the northern region of England, including Yorkshire and the Humber, the North East, and the North West. We selected 26 journals detailed in S1 Table to search for alcohol, addiction and mental health related articles. Additionally, the HEIs used to identify potential authors in the North of England are also shown in S2 Table.

The NARA was established as a collaborative clinical-academic network hosted by the CAMHR and funded by the National Institute of Health and Care Research (NIHR) Regional Research Delivery Network. It’s a publicly funded initiative led by Professor Thomas Phillips, with the primary mission of increasing leadership and participation in public health and clinical applied addiction research across the Northern region of England. The alliance promotes research applications led by its members, creates opportunities for early- and mid-career researchers, and supports the development of embedded research roles within clinical settings.

The network’s focus is to facilitate multidisciplinary collaborations between early-career and senior clinical academics with a special interest in addiction; however, identifying members is labor-intensive and time-consuming. Therefore, we recognized the potential for AI to accurately and efficiently target potential members within the northern regions of England.

Table 1 Title: Methodological steps taken in this research.

Methodological steps	Details
Identify Domain	The target domain includes alcohol, addiction, and mental health.
Select Relevant Journal(s)	26 peer-reviewed journals were selected based on disciplinary relevance, including <i>Addiction</i> , <i>Alcohol and Alcoholism</i> , and the <i>International Journal of Drug Policy</i> (see <i>Supplemental material for full list</i>).
Query Database	The query was developed to capture articles on alcohol, addiction, and mental health then ran across the 26 journals. (see <i>supplemental material for query</i>).
Retrieving Articles and Extracting Metadata	The articles were retrieved after running the query, and relevant metadata such as the authors information were extracted.
Filtering for Institutional Affiliation	The dataset was filtered to retain only authors affiliated with 21 Higher Education Institutions (HEIs) located in the North East, North West, and Yorkshire & The Humber regions.
Retrieve Authors ORCID IDs	Authors' ORCID IDs were retrieved to perform author disambiguation.
Build a Research Network	A research network was built where authors serve as nodes and articles act as edges. The network was colour-coded based on the authors' institutions. This helps to show patterns of institutional collaboration within the northern region of the UK.

Table 1 Legend: This table shows the methodological steps and how they were applied to the NARA case study to identify authors in the UK Northern region who have published on alcohol, addiction and mental health from 2020–2025.

The selection of alcohol addiction as the focal case study within the context of the CAMHR initiative and NARA was a deliberate and contextually grounded choice. Alcohol misuse remains a persistent public health concern in the Northern region of the United Kingdom, where social and health inequalities contribute to higher prevalence rates and more severe clinical outcomes compared to other regions. The Northern region continues to experience the highest rates of alcohol dependence, deaths from drug misuse, and smoking-attributable mortality. In light of the Independent Review of Drugs, strengthening the impact of addiction services through the adoption of applied research has become an urgent public health priority. (Black et al., 2021; Cranstoun, 2022; Alcohol Health Alliance, 2024)

Limiting the scope of this case study to northern institutions in England therefore provided an opportunity to address these region-specific challenges, while simultaneously capitalizing on the emerging infrastructure and funding opportunities aimed at enhancing research capacity in addiction and mental health within the North.

Results

Distribution of Authors Identified

The analysis identified 420 unique UK Northern-based authors across 21 institutions who published on alcohol, addiction, and mental health research within the past five years. By region, the North West had the highest number of unique authors (195), followed by Yorkshire and the Humber (168) and the North East (57).

Distribution across publication years

The analysis in Figure 2 of articles published across all queries showed that the top 3 years with the most articles published include 2021, 2020, and 2022. Overall, the trend indicates a gradual decline in the number of articles published on alcohol, addiction, and mental health research over recent years.

[Insert Figure 2]

Figure 2. Number of articles published across years.

2021 recorded the highest number of publications with approximately 4300 articles, followed by 2020 with 3900 articles, 2022 approximately 3800 articles. The overall pattern shows a downward trend in research output related to alcohol, addiction, and mental health topics.

Numbers of Articles Published by Institution (UK Northern Region)

After consolidating all their articles for the specific query ran, the results were then filtered for UK Northern institutions (York and Humber, Northwest, and Northeast regions). The results of the experimentation then showed that, in Figure 3, the top 3 universities with the highest number of articles published in the past 5 years are the University of Sheffield (94 articles), the University of York (57 articles), and the University of Liverpool (52 articles). Additionally, when categorizing the universities based on regions, the top 3 universities in York and Humber include the University of Sheffield (94 articles), the University of York (57 articles), and the University of Leeds (9 articles). In terms of Northeast, the top 3 universities with the highest frequencies of articles include Newcastle University (30 articles), Northumbria University (10 articles), and Teesside University (8 articles). Lastly, for Northwest, the top 3 universities include the University of Liverpool (52 articles), the University of Manchester (33 articles), and Liverpool John Moores University (31 articles).

[Insert Figure 3]

Figure 3. Numbers of Articles published by UK Northern HEIs.

This figure shows the distribution of article published by each UK northern HEIs. The overall top 3 universities are the University of Sheffield (94 articles), the University of York (57 articles), and the University of Liverpool (52 articles). Regional findings show that for York and Humber the top 3

universities are University of Sheffield (94), the University of York (57), and the University of Leeds (9). For the Northeast the top 3 universities are Newcastle University (30), Northumbria University (10), and Teesside University (8). Lastly for Northwest region the top 3 universities are the University of Liverpool (52) the University of Manchester (33) and Liverpool John Moores University (31).

Network Construction [Yorkshire and Humber Region Focus]

The co-authorship network for universities in the Yorkshire and Humber region reveals the collaborative structure among researchers across multiple institutions. Each node represents an author, color-coded by their university, while edges denote co-authored publications. As shown in Figure 4, there is a research network collaboration between the institutions in the region. Zooming in on institutions, it is seen that the University of Sheffield, indicated by the purple node, appears as the most central and densely connected cluster in the region, which is a result of the institution having the highest number of publications (94 articles) in the entire North of England. However, from the network, it is seen that collaboration is mostly internal, but there are also forms of collaboration between the university researchers and researchers from other universities. Nevertheless, it can be attributed that the University of Sheffield plays a pivotal role in regional research collaborations in alcohol, addiction, and mental health research.

In terms of the University of York, it can be seen that there is more diverse collaboration between each university researcher and researchers from other universities than internal collaboration between researchers. Zooming into the University of Leeds, there are also collaborations between the university researchers and researchers from other universities in the region. Likewise, for the rest of the universities, i.e., Leeds Trinity University and the University of Hull, there are also collaborations between each university researchers and researchers from other universities.

Quantitatively there are 168 nodes (authors) and 582 edges (collaborations) in Yorkshire and Humber region, the network density of the region was calculated to be 4%, indicating that 4% of all possible collaborations exist. Additionally, the average degree was 7, suggesting that each researcher has collaborated with an average of seven other researchers. Overall, in the Yorkshire and Humber region, as shown in the network construction (Figure 4), there is a strong research network between the universities in the region, with engagement in collaboration internally as well as extending collaboration to other universities, thereby indicating a strong regional research network.

[Insert Figure 4]

Figure 4. Research Network Construction for Yorkshire and Humber region.

This figure illustrates the research collaboration network between universities in the Yorkshire and Humber region. Quantitatively the network has 168 nodes authors and 582 edges collaborations. The

network density is 4% indicating that 4% of all possible collaborations exist. The average degree is 7 suggesting that each researcher has collaborated with an average of seven other researchers. Overall, the network construction shows a strong regional research network with high engagement in collaboration both internally and with other universities.

The network construction of other regions, including the North East, North West, and the entirety of Northern England, is shown in Supplemental Figure 1, 2, 3. respectively. Additionally, the number of publications for each university across each region is shown in Supplemental Figure 4, 5, 6, 7 respectively.

Discussion

In this study, we present an automated pipeline for identifying UK Northern-Region based researchers in alcohol, addiction, and mental health by analyzing metadata from peer-reviewed articles and constructing a research network.

The results show that the Yorkshire and Humber region has the highest number of publications (181 articles) in this domain over the past five years. The reason for Yorkshire and Humber having the highest publication could be attributed to the presence of the NIHR Applied Research Collaboration (ARC) Yorkshire and Humber, which prioritizes mental health and addiction research. Among institutions, the University of Sheffield, located in the Yorkshire and Humber region, has the highest number of publications across the entire northern UK. The institution has 94 publications, accounting for 52% of publications in the York and Humber region and 24% of publications in Northern England.

Furthermore, on the collaboration between universities, it was seen that there is a strong research network across the university in the Northern England, when focusing on Yorkshire and Humber region, it was seen that University of Sheffield represents the central of research hub in the region, but most of its research are collaborated internally but there are also cases of collaboration with other universities. Additionally, collaborations among other universities in the region indicate a strong research network overall. This may be attributed to the presence of established research groups such as the Sheffield Addictions Research Group (SARG) in the university, which specialize in conducting public health and health economic research on alcohol, tobacco, gambling and addiction, and research centers, such as NIHR-funded units, which aim to strengthen addiction research capacity across Northern England.

While this research has achieved its aim of automatically identifying authors who have published on alcohol, addiction, and mental health, within the past 5 years, and of building a research network to identify relationships, collaborations, and insights between universities, this study has some limitations.

The approach is limited to indexed publications on PubMed and may exclude early-career researchers who haven't yet published or clinicians who don't publish in the selected journals. Additionally, not all researchers' ORCID IDs were retrieved for author disambiguation, prompting us to use authors' full names as unique identifiers for those with missing IDs. Lastly, the researchers identified is limited to the

number of papers retrieved from the PubMed API utilized, which may not capture the full list of publications on alcohol, addiction, and mental health research. Future work should expand on these limitations and apply more advanced network analytics techniques to capture broader collaboration patterns.

Conclusion

This study demonstrated an automated, data-driven approach that leverages data science and artificial intelligence, particularly through NLP, to identify potential collaborators and build a clinical and academic research network with broad cross-specialty applications. By analyzing PubMed metadata, we identified researchers in Northern England specializing in alcohol, addiction, and mental health, and visualized institutional collaborations across the region by constructing a research network. The results of the experimentation showed that the Yorkshire and Humber region is the Northern England with the highest number of research publications in this domain over the past five years. The region has published a total of 181 research publications, while the North West has 162 publications, and the North East with 55 publications.

This study presents a scalable and reproducible framework for automatically creating a research network, providing value for funding agencies and policy strategists seeking to enhance collaboration in mental health and addiction research.

Abbreviations

- Northern Addiction Research Alliance (NARA)
- Centre for Addiction and Mental Health Research (CAMHR)
- National Institute of Health and Care Research (NIHR)
- Applied Research Collaboration (ARC)
- Higher Education Institutions (HEIs)

Declarations

Ethical Statement

The data used in this study were sourced through freely accessible bibliographic databases, such as PubMed/MEDLINE and ORCID. Since this study required the examination of peer-reviewed publication meta-data that was already available and did not presuppose direct contact with human subjects, personal medical records, and confidential patient information, no formal ethical approval or Ethics Committee was necessary. All the data was processed according to the terms of use of the corresponding APIs and databases, which were used.

Consent for publication

Not applicable

Data Availability Statement

The datasets generated and/or analyzed during the current study are not publicly available due to personal identifiable information of authors but are available from the corresponding author on reasonable request.

Funding

Not applicable

Author Contributions

HJE Co-conceptualised the solution, performed the modelling, and drafted the first version of the manuscript. WSJ co-conceptualised the solution and provided substantial revision and editing of the manuscript. TP co-conceptualised the solution, provided clinical context, and provided substantial revision and editing of the manuscript.

Authors' Information

Hope James Ekpa is a Research Assistant at the Centre for Addiction and Mental Health Research (CAMHR) at the University of Hull.

Thomas Phillips is the Director of the Centre for Addiction and Mental Health Research (CAMHR), University of Hull.

William Stephen Jones is a Senior Lecturer in the Centre of Excellence for Data Science, Artificial Intelligence, and Modelling (DAIM) at the University of Hull.

References

1. Alliance, A. AHA call on government to tackle 'shameful failure' in public health policy as deaths from alcohol reach all-time high [Internet]. 2022. Available from <https://ahauk.org/news/ons2022/>.
2. Andrés A. Measuring academic research: How to undertake a bibliometric study. Chandos publishing; 2014.
3. Black C. Independent review of drugs by Professor Dame Carol Black [Internet]. 2020
4. Breiman RF, Osoro E, Reithinger R, Wang D, Diamond M, Van Voorhis WC, Wasserheit JN, Rabinowitz PM, Mboup S, Hemingway-Foday JJ, De Oliveira T. Importance of outbreak response research in bridging knowledge gaps on emerging infectious diseases. *BMJ Global Health*. 2025; doi: 10.1136/bmjgh-2024-018297.
5. Brown BB, Patel C, McInnes E, Mays N, Young J, Haines M. The effectiveness of clinical networks in improving quality of care and patient outcomes: a systematic review of quantitative and qualitative

- studies. BMC health services research. 2016; doi:10.1186/s12913-016-1615-z
6. Carter CB, Blanford CF. All authors must now supply ORCID identifiers. Journal of Materials Science. 2017 Jun; doi:10.1007/s10853-017-0919-7
 7. Cranstoun. Cranstoun's Response to "Alcohol-specific deaths in the UK: registered in 2022" ONS Release [Internet]. 2024. Available from: <https://cranstoun.org/news/cranstouns-response-to-alcohol-specific-deaths-in-the-uk-registered-in-2022-ons-release/>.
 8. Li G, Wu Q, Jin Y, Vanniyasingam T, Thabane L. Key factors of clinical research network capacity building. Journal of Venomous Animals and Toxins including Tropical Diseases. 2018; doi:10.1186/s40409-018-0152-0
 9. Marsden PV. Survey methods for network data. The SAGE handbook of social network analysis. 2011.
 10. Metapub. Pythonic access to PubMed [Internet]. 2022. Available from: <https://metapub.org/>.
 11. NIHR Statistical Group. Bringing statisticians and researchers together. <https://statistics-group.nihr.ac.uk/>.
 12. OpenAlex. OpenAlex: A fully-open index of scholarly works, authors, venues, institutions, and concepts [Internet]. Available from: <https://openalex.org/>.
 13. Priem J, Piwowar H, Orr R. OpenAlex: A fully-open index of scholarly works, authors, venues, institutions, and concepts. arXiv preprint arXiv:2205.01833. 2022.
 14. Profiles. Professional Networking and Expertise Mining for Research Collaboration [Internet]. United States. Available from: <https://profiles.catalyst.harvard.edu/>.
 15. ResearchGate. Professional network for scientists and researchers [Internet]. 2008. Available from: <https://www.researchgate.net/>.
 16. Scinapse. Find the Most Suitable Researcher in a Specific Field [Internet]. Seoul. Available from: <https://www.scinapse.io/app>
 17. Scott A, Woolcott G, Keast R, Chamberlain D. Sustainability of collaborative networks in higher education research projects: why complexity? Why now?. Public Management Review. 2018 Jul 3;20(7):1068-87.
 18. Xu J, Kim S, Song M, Jeong M, Kim D, Kang J, Rousseau JF, Li X, Xu W, Torvik VI, Bu Y. Building a PubMed knowledge graph. Scientific Data, 7 (1). 2020; doi:10.1038/s41597-020-0543-2

Figures

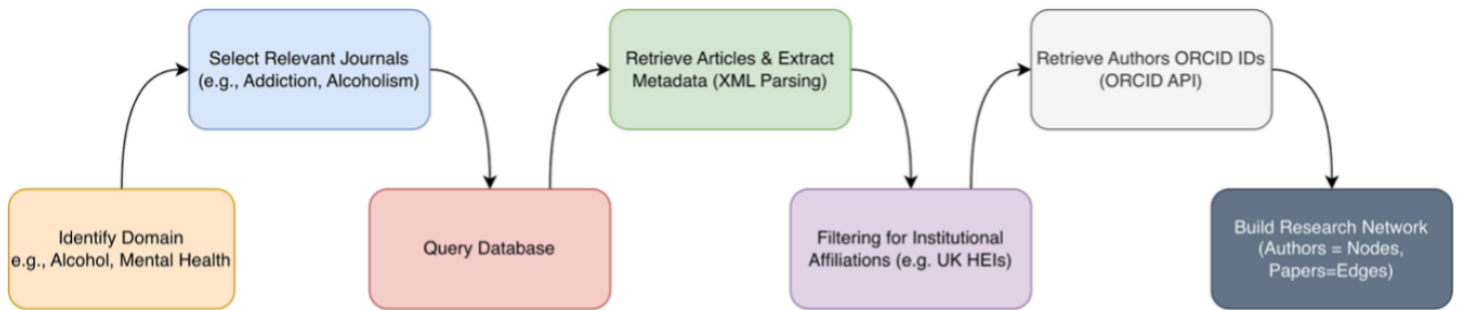


Figure 1

Conceptual methodological workflow of the AI and data science approach for building the research network.

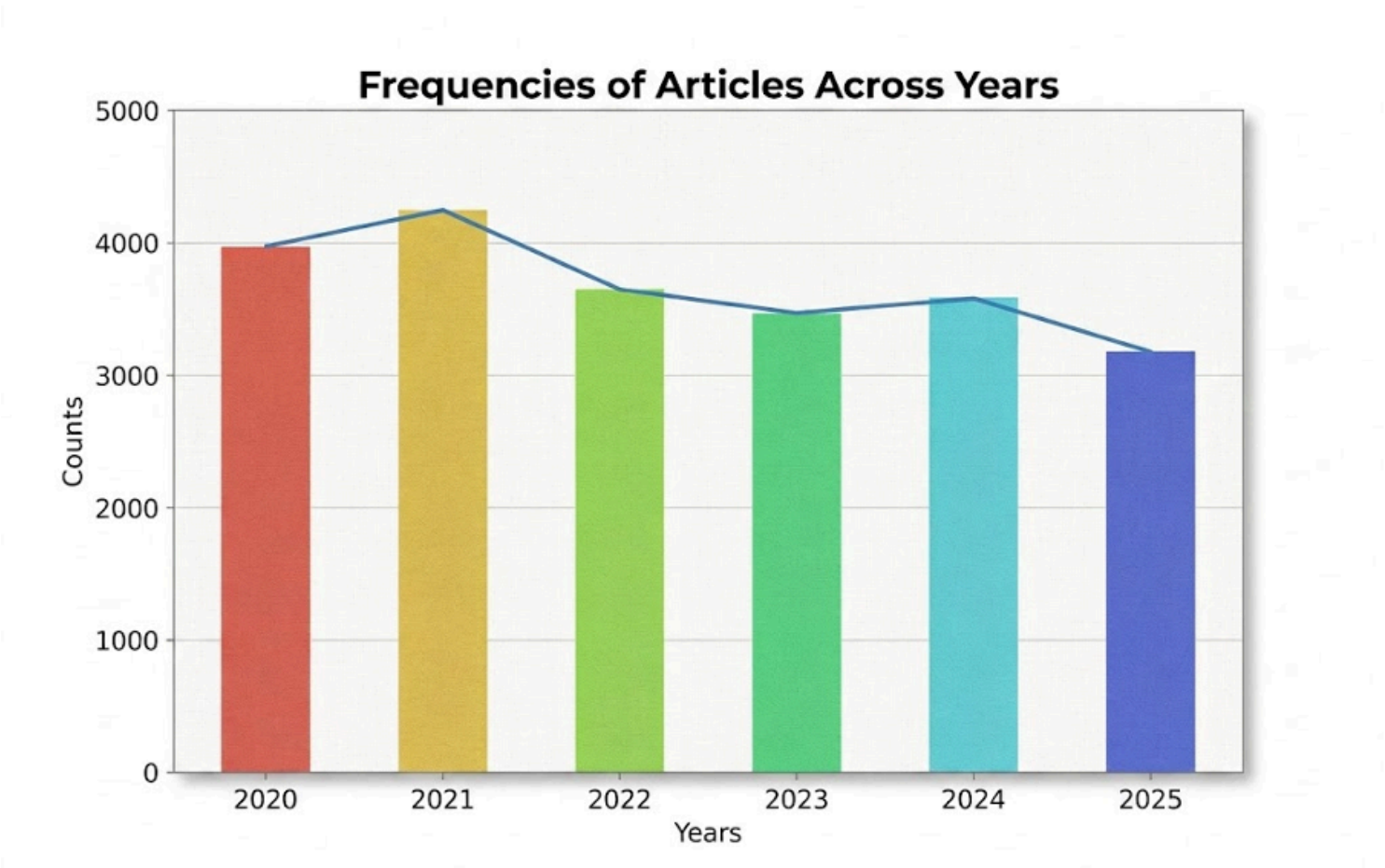


Figure 2

Number of articles published across years.

S/N	HEI	Article Published
1	University of Sheffield	94
2	University of York	57
3	University of Liverpool	52
4	University of Manchester	33
5	Liverpool John Moores University	31
6	Newcastle University	30
7	Manchester Metropolitan University	14
8	Edge Hill University	14
8	Lancaster University	10
10	Northumbria University	10
11	University of Leeds	9
12	University of Hull	8
13	Leeds Trinity University	8
14	Teesside University	8
16	Durham University	7
16	University of Chester	4
17	University of Salford	3
18	University of Huddersfield	2
19	Sheffield Hallam University	2
20	Liverpool Hope University	1
21	Leeds Beckett University	1

Figure 3

Numbers of Articles published by UK Northern HEIs.

Co-authorship Network by University [Yorkshire & Humber]

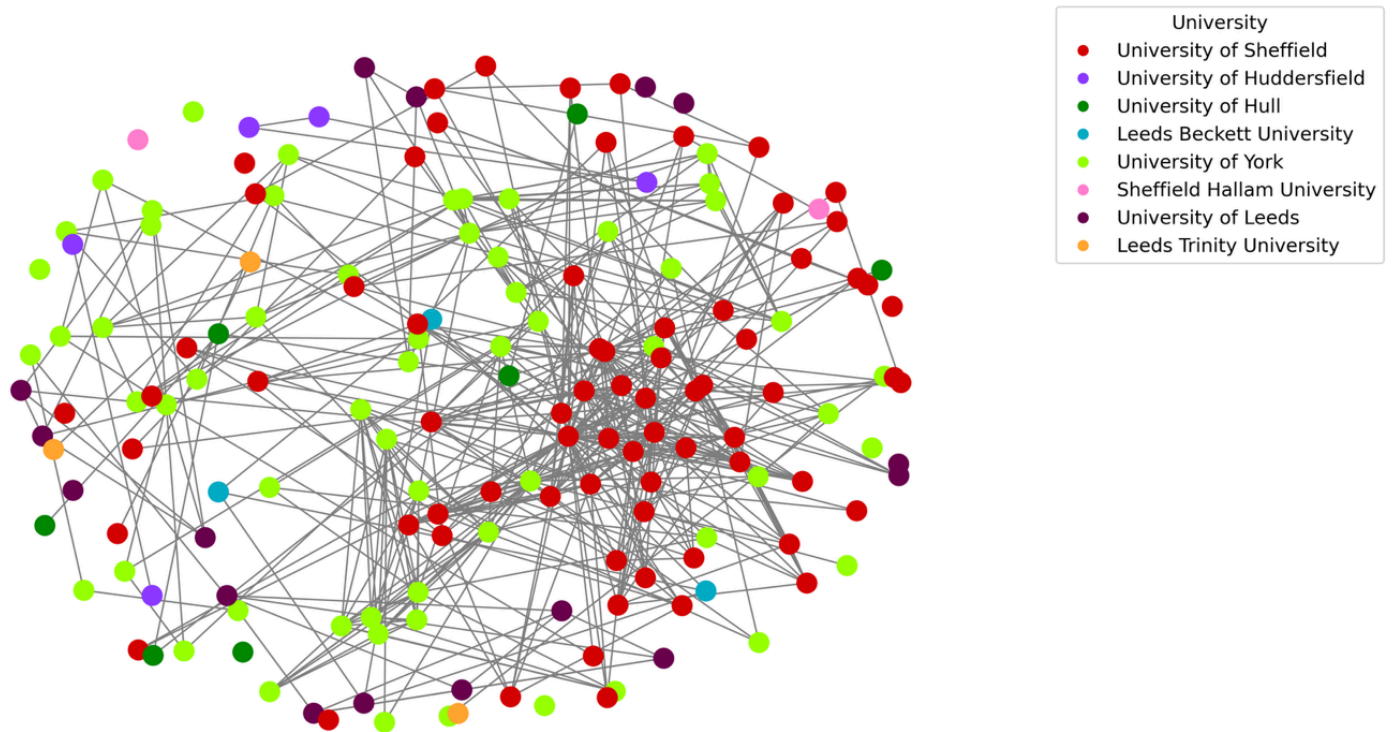


Figure 4

Research Network Construction for Yorkshire and Humber region.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SupplementalMaterial.docx](#)