Navigating Multidisciplinary Research Using Field of Study Networks

Eoghan Cunningham[1,2], Barry Smyth[1,2], and Derek Greene[1,2]

[1] School of Computer Science, University College Dublin, Ireland
[2] Insight Centre for Data Analytics, University College Dublin, Ireland

Abstract. This work proposes Field of Study networks as a novel network representation for use in scientometric analysis. We describe the formation of Field of Study (FoS) networks, which relate research topics according to the authors who publish in them, from corpora of articles where fields of study can be identified. FoS networks are particularly useful for the distant reading of large datasets of research papers, through the lens of exploring multidisciplinary science. To support this, we include case studies which explore multidisciplinary research in corpora of varying size and scope; namely, 891 articles relating to network science research and 166,000 COVID-19 related articles.

Keywords: Network analysis, Scientometrics, Multidisciplinarity

1 Introduction

In line with recognised benefits of multidisciplinary and interdisciplinary collaboration in scientific research [10, 15], a trend has established towards greater levels of interdisciplinary research [11]. A common means of understanding these research processes is through the lens of network analysis. For instance, given a collection of research papers and their associated metadata, we can construct a variety of different network representations, including co-authorship networks [6, 7] and citation networks [8]. Such representations serve to highlight the collaboration patterns between individuals researchers at a micro level. However, in other cases we might be interested in examining collaboration patterns between researchers coming from different disciplines at the macro level. In particular, we might wish to study how these patterns evolve over time in response to changing research funding landscapes or exogenous events, such as the COVID-19 pandemic.

In this work, our aim is to propose a practical "distant reading" approach to help reveal collaborative research patterns in large scientific corpora in order to understand better the nature and implications of these patterns. This concept of distant reading has been considered in other contexts as a means of exploring large volumes of data from a macro level perspective, to identify specific niche areas of interest for closer inspection [13]. In this work, we present a novel graph representation, the *Field of Study* (FoS) network, which facilitates the investigation of multidisciplinary and interdisciplinary research in corpora of scientific research articles at the macro level. A core contribution of the field of study networks is the use of author-topic relations; a FoS network is populated by fields of study (or research topics), which are related to one another according to the authors who publish in them. In Section 3 we describe how these networks can be constructed from the topics/fields of study that have been assigned to research papers. In Section 4 we describe two exploratory cases studies, which analyse the FoS networks arising from datasets of differing scope and size. These case studies suggest that FoS networks can provide a useful tool for the distant reading of large corpora of research articles, as well as conducting quantitative analysis to understand the relationship between scientific disciplines.

2 Related Work

Multidisciplinary research is most commonly defined as research which draws on expertise, data or methodology from two or more disciplines. Most formal definitions distinguish *interdisciplinary* research as an extension of multidisciplinary research, which involves the *integration* of methodologies from the contributing disciplines [4]. There are numerous analyses which explore multi- or interdisciplinary research, and investigate the relationship between scientific disciplines. Many studies define metrics to quantify research interdisciplinarity at the author or paper level [17, 16], often in order to investigate a correlation between interdisciplinarity and research impact [15, 10], productivity or visibility [12]. Typically, works which integrate methods and ideas from a diverse set of disciplines are found to have greater research impact and visibility compared to those that do not [12, 15]. As such, we can identify several examples of analyses which investigate cross-disciplinary collaboration and map areas of multidisciplinary research, often drawing on methods from network science [6, 20, 8, 18, 9, 18].

Co-authorship networks can provide an effective means of representing research collaborations. Here researchers are represented by nodes and collaborations are encoded via the edges between them. Thus, research teams are identified as fully-connected components of the graph. In cases where research backgrounds can be identified among the authors in the network, this can be used to quantify the level of multidisciplinary collaborations. These methods have been used to reveal a strong disciplinary homophily between researchers, despite showing those with diverse neighbourhoods tend to have higher research impact [6].

Another common representation used to investigate interdisciplinary research is the citation network, typically constructed at the article or journal level. Analyses of citation networks can highlight influential or "disruptive" articles in interdisciplinary research [20], as well as "boundary" papers which span multiple disciplines [8]. Indeed community finding approaches have been employed to automatically group articles in citation networks into their respective fields of study [18], so that interdisciplinary interactions can then be explored at the macro level. An alternative strategy is to apply text analysis to article abstracts in order to cluster articles together which relate to similar research topics [9, 18]. This is typically based on term co-occurrence patterns, rather than based on article citation patterns. Of course, connections between topics in each of these representations can differ greatly, as fields of study which are distant in their citation patterns may be closely linked semantically.

Here we propose an alternative network representation, which relates fields of study according to the authors who typically publish in those fields. This Field of Study network may be used in conjunction with more conventional network representations — in much the same way that semantic networks have been shown to complement citation networks [18] — but in Section 4 we show that, on their own, FoS networks can provide an effective means of exploring large collections of research articles, particularly in revealing author multidisciplinarity.

3 Methods

In this section we formalise the definition of a Field of Study (FoS) network and explain how these networks can be generated from existing research resources. In Sections 3.2 and 3.3 we describe two FoS variations: the *static* FoS network and the *temporal* FoS network respectively.

3.1 Field of Study Networks

Formally, a Field of Study (FoS) network is defined as a general graph representation of a collection of research articles (R), written by a set of authors (A), and denoted $F = (N, E)$. The nodes (N) represent identifiable research topics (i.e. the fields of study) and the edges (E) represent authorship relations between pairs of topics. These relations are aggregated across multiple associated research papers. Below we describe how a FoS network can be constructed from a more conventional authorship graph and we argue that FoS networks are particularly well-suited to analysing the nature of collaboration within the scientific literature, especially as they relate scientific fields of study according to the researchers/authors who publish in them.

The formation of a FoS network depends on the availability of fields of study labels for a given set of research papers. These could be derived via manual annotations, the application of automated text mining methods, or some combination of the two. For instance, topic modelling techniques have been shown to be successful in extracting research topics from corpora of research articles and assigning papers to those fields [9].

In fact, many research databases and search engines employ these techniques (or manual classification) to assign research articles or academic journals to fields of study. For example, the Microsoft Academic Graph (MAG)[3] maintains a deep hierarchy of *Fields of Study* which they assign to papers; Web of Science (WOS)[4] group journals in 258 *Subject Categories*; Scopus[5] employs experts to assign *All Science Journal Classification (ASJC)* codes to all journals covered

[3] https://www.microsoft.com/en-us/research/project/microsoft-academic-graph/
[4] https://clarivate.com/webofsciencegroup/solutions/web-of-science/
[5] https://www.scopus.com/home.uri

by their index. For the purpose of the case studies described later in Section 4, we use MAG fields of study to categorise research papers and construct FoS networks. The deep MAG field of study hierarchy is desirable as it supports the construction of FoS networks at varying levels of detail, from the broadest research disciplines (level 0) to the specific topics and sub-topics that exist within a particular discipline (levels 4 and 5).

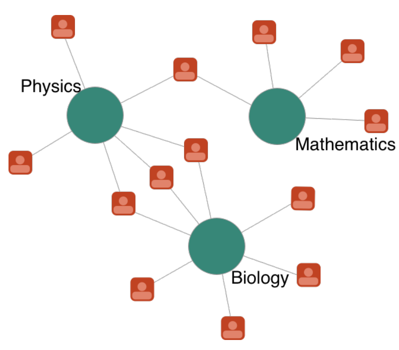
It is worth noting that the Microsoft Academic Graph may not always be an appropriate source for field of study data. For instance, the corpus does not provide full coverage of all research disciplines and the massive FoS hierarchy may contain some spurious connections due to its size and semi-automated construction. However, the methods that we propose are not specific to the MAG hierarchy, and are designed to generalise to any scenario where fields of study can be identified at the appropriate level of detail.

3.2 Static FoS Networks

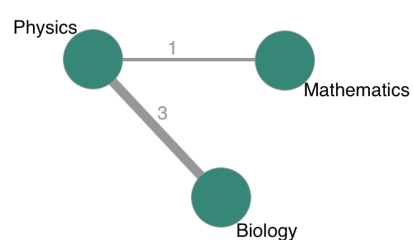
The formation of a static FoS network from a collection of research articles is best described as the two-step process illustrated in Figure 1. In the first step, an unweighted bipartite graph is generated from identifiable fields of study and their contributing authors; see Figure 1a. In the second step, this graph is used to generate a projection (the FoS Network) in which a weighted undirected edge exists between two fields if and only if at least one author has published research in both fields; see Equation 1 for all $a \in A$, where N is the set of fields identifiable in R. The resulting edge weights correspond to the number of such authors who publish in both fields (Equation 2).

$$E = \big\{(n_i, n_j) \ : \ published(a, n_i) \ \wedge \ published(a, n_j)\big\} \tag{1}$$

$$w\big(n_i, n_j\big) = |\big\{a \ : \ published(a, n_i) \ \wedge \ published(a, n_j)\big\}| \tag{2}$$

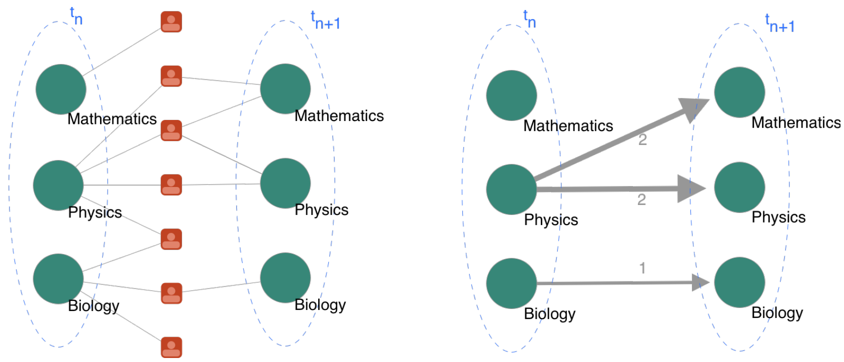


(a)



(b)

Fig. 1: The formation of a *static* Field of Study (FoS) network involving two steps: (a) creation of a bipartite network of authors and fields; (b) projection to an *undirected* network of fields.



(a) (b)

Fig. 2: Illustrative example of a *temporal* Field of Study (FoS) network, involving two steps: (a) creation of a bipartite network of authors and fields; (b) projection to a *directed* network of fields.

3.3 Temporal FoS Networks

It is further possible to encode temporal information in a FoS Network as *directed* edges, which allows us to study changes in multidisciplinarity research patterns over time. Temporal FoS networks can be visualised in a time-unfolded representation, where the data is divided into a sequence of two or more discrete *time steps*, as frequently employed in dynamic network analysis tasks. Nodes are duplicated for each time step so that authors can be connected to any fields in which they publish research during a given time step.

As an example, Figures 2a and 2b illustrate the two stages in the formation of a temporal FoS network, showing an instance of a temporal FoS network with respect to two time-points (t_n and t_{n+1}) on either side of some event (e); thus $t_n < t_e < t_{n+1}$). The temporal FoS network in Figure 2b contains a directed edge between two fields (n_i, n_j) if an author published in field n_i at time t_n (before event e) and in field n_j at time t_{n+1} (after event e), as given in Equation 3. Later, in Section 4.3, we present COVID-19-related research in the context of the research backgrounds of the contributing authors with the start of the pandemic serving as the defining event.

$$E' = \left\{ (n_i, n_j) \ : \ published(a, n_i, t_n) \ \land \ published(a, n_j, t_{n+1}) \right\} \tag{3}$$

4 Case Studies

In what follows we describe two illustrative examples to demonstrate the utility of FoS representations. In the first case study, presented in Section 4.1, we consider the use of static FoS networks to explore aspects of multidisciplinary research in the area of network science. The second case study demonstrates the use of both static and temporal FoS networks in the context of a large-scale dataset of research articles relating to the COVID-19 pandemic and is presented in Sections 4.2 and 4.3.

4.1 Multidisciplinary Research in Network Science

Figure 3 presents two static FoS networks produced using Microsoft Academic Graph metadata for 891 research articles published in the area of network science. To form this dataset, we collect all available papers published in 5 network science journals between the years 2015 and 2019 inclusive: *Applied Network Science*, *Social Network Analysis and Mining*, *Network Science*, *Complex Systems*, and *Computational Social Networks*. We use MAG *fields of study* metadata to categorise these research papers. The MAG uses hierarchical topic modelling to identify and assign research topics to individual papers, each of which represents a specific field of study [19]. To date, this approach has identified a hierarchy of over 700,000 topics within the Microsoft Academic Knowledge corpus, and the average paper published in the set of 891 network science articles is assigned to 9 such topics.

To produce a more useful categorisation of articles, we first reduce the number of topics, by replacing each field with its parent, to consider topics at two levels in the FoS hierarchy:

1. The 19 FoS labels at level 0, which we refer to as 'disciplines'.
2. The 292 FoS labels at level 1, which we refer to as 'sub-disciplines'

In this way, each article is associated with a set of disciplines (e.g. 'Medicine', 'Physics', 'Engineering') and sub-disciplines (e.g. 'Virology', 'Particle Physics', 'Electronic Engineering'), which are identified by traversing the FoS hierarchy from the fields originally assigned to the paper. Note that some MAG sub-disciplines belong to more than one discipline. For example, Biochemistry is a child of both Chemistry and Biology. Figure 3a illustrates the resulting FoS network when network science articles are categorised at the *discipline* level. Each node (or discipline) in this FoS network can then be decomposed into its *sub-disciplines* as shown in Figure 3b.

From Figure 3, we can begin to understand the respective roles of the many fields of study represented in network science. Highly central in Figure 3b are the fields which represent the technical and methodological foundations of network science research. The sub-disciplines of Mathematics and Computer Science such as 'Algorithm', 'Combinatorics', and 'Statistics' have high degree centrality (ranked 3rd, 8th and 9th respectively), because they are identified across the majority of network science research papers. Some fields beyond the disciplines of Computer Science and Mathematics, such as 'Social Psychology', 'Social Science', and 'Law' have high betweenness centrality in the FoS Network (ranked 1st, 4th and 6th, respectively). This is likely because they help to bridge network science methods to their interdisciplinary applications. In particular, in the upper right corner of Figure 3b we can see a group of fields which reflect the proliferation of recent studies of social media networks from the perspective of sociology and political science.

Community detection methods can be used to categorise the topics in the FoS network too. Figure 4 shows the network from Figure 3b, but with the nodes colour-coded to show cluster memberships identified using the Louvain method [2]. This technique identified 6 clusters in the graph, containing as few as 2, and as many as 16 topics. The clusters shown in Figure 4 differ from the MAG categorisation illustrated in Figure 3b, because they show how these topics relate in the context of network science specifically, rather than in the MAG hierarchy as a whole. Broadly, the clusters could be categorised as: (i) the core theoretical and methodological topics in the network science (15 topics: statistics, theoretical computer science, combinatorics, etc), (ii) research relating to computer networks [3] (5 topics: telecommunications, distributed computing, etc), (iii) social network analysis (16 topics: social science, social psychology, media studies, etc), (iv) networks in machine learning [1] (6 topics: artificial intelligence, computer vision, natural language processing), (v) applications in biology (3 topics: molecular biology, genetics, biochemistry), (vi) applications in medicine (2 topics: pathology and surgery).

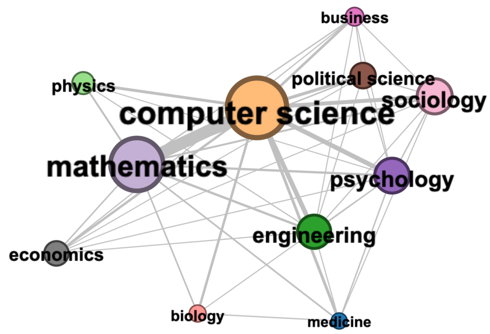
4.2 COVID-19 Research and the Effect on Multidisciplinarity

FoS networks can be used to evaluate the degree of an author's *multidisciplinarity*, that is, the extent to which they publish in different disciplines. For example, [5] describes an in-depth analysis of the effect of COVID-19 research on author multidisciplinarity using static FoS networks and for completeness, we summarise the construction and use of FoS networks in this way for this case study.

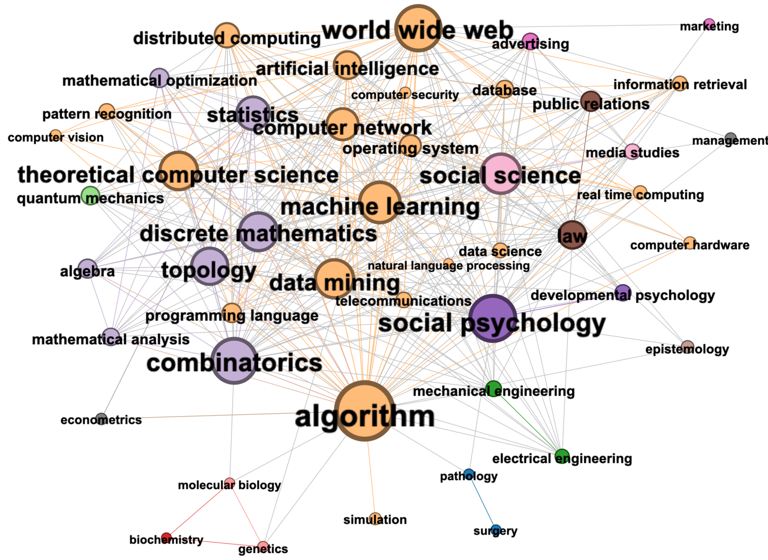
We construct five annual FoS networks from all available research articles by authors who published work related to COVID-19 using the COVID-19 Open Research Dataset (CORD-19)[6]. From CORD-19, we identify all authors who published COVID-19 related research in 2020, and collect MAG metadata for any available research articles they published between 2016 and 2020 inclusive. In total, we collect 5,389,445 articles published 2016-2020, including 166,356 articles which relate to COVID-19.

Next, using the 292 MAG sub-disciplines, we build a FoS network for each year in the dataset. The nodes in these networks represent MAG sub-disciplines, and they can be divided into 19 overlapping communities based on their assignment to MAG disciplines. This facilitates the characterisation of edges in the FoS network: an edge *within* a community represents an author publishing in two sub-disciplines within the same parent discipline, while an edge *between* communities represents an author publishing in two sub-disciplines from different parent disciplines. For example, if an author publishes research in 'Machine Learning' and 'Databases', then the resulting edge is *within* the community/discipline of 'Computer Science'. Conversely, if an author publishes in 'Machine Learning' and 'Radiography', the resulting edge is *between* the 'Medicine' and 'Computer

[6] https://www.semanticscholar.org/cord19



(a) Disciplines or level 0 fields of study.



(b) Sub-disciplines or level 1 fields of study.

Fig. 3: FoS Networks for research published in 5 network science journals during 2015–2019. Node size encodes the number of papers attributed to a field of study. In (b) nodes are coloured to represent the parent discipline of the field of study. Edges are coloured to show the parent discipline if the edge is within a discipline/community. Edges between communities are not coloured.

Science' communities. In this way, an edge between disciplines may represent either a single instance of interdisciplinary research or two separate stances of research, in two different disciplines, by the same author. To explore changes in author multidisciplinarity, we compare the proportion of the total number of edges in the network that are external (i.e. between communities). Figure 5 plots the odds ratio effect sizes when the proportion of external edges in an annual FoS network is compared with that of the previous year. We report these

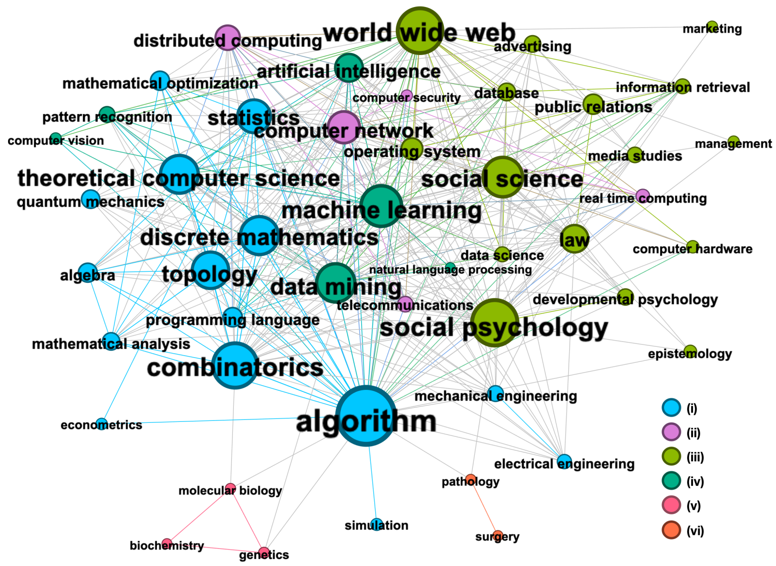


Fig. 4: FoS Network for research published in 5 network science journals during 2015–2019. Nodes are coloured to show clusters identified by Louvain.

scores per community/discipline. We also include a second FoS network for 2020 which excludes any research related to COVID-19, and report an additional odds ratio for the comparison of the 2020-non-COVID network with the 2019 network. Thus, FoS networks have been used to reveal a trend towards greater multidisciplinarity year-on-year. This trend appears to have been accelerated by COVID-19 research, and the increase is shown to be greater in some disciplines.

4.3 Close reading case studies in COVID-19 research

Figure 5 shows an increase in author multidisciplinarity in many fields of study as a result of COVID-19 research and in this section we illustrate how we can further explore this phenomenon by using Temporal FoS networks to compare the pre-COVID (2016-2019) and COVID (COVID-19 related research in 2020) time periods. As an illustration, Figure 6 presents COVID-19 related research in the field of Computer Science, with pre-COVID nodes on the left (representing the authors' research backgrounds) and COVID nodes on the right (representing the FoS characterisation of the COVID related research). To highlight the strongest trends that exist, the FoS network shows only the top-50 edges by weight. We note that authors from diverse research backgrounds contribute articles related primarily to 'Surgery', 'Pathology', and 'Machine Learning'.

To conduct further close reading, we can narrow the list of articles by considering only those papers that contribute a particular edge to the FoS network. For example, we can search for COVID-related papers which result in the edge

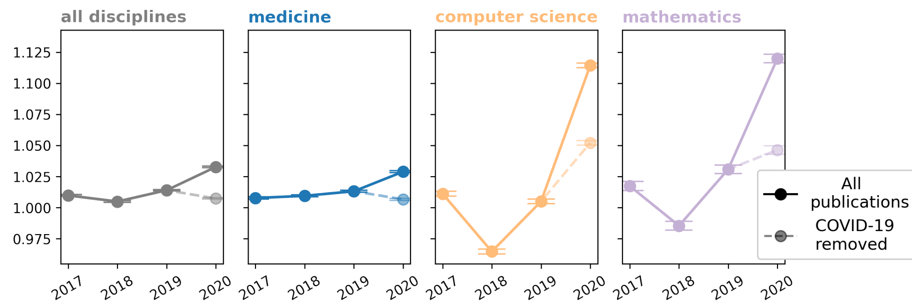


Fig. 5: Multidisciplinarity of authors who published COVID-19-related research, by discipline. For each year, we report the odds ratio effect size when the proportion of edges that are between communities is compared with that of the previous year. 'All disciplines' reports these scores for the entire network. Also reported are scores for individual communities in the graph, which represent disciplines. Bars are plotted to show a 95% confidence interval.

between Pathology and Algorithm; these are COVID-related articles containing the topic Algorithm, in which the authors have previously published research in the field of Pathology. To better understand the papers in this subset, we can explore the lower-level MAG topics that are most commonly identified amongst them, or the keywords which occur most frequently in their titles and abstracts. For additional discussion of close reading of this corpus, see [5].

One approach to close reading is to search for articles which cite a large proportion of the papers in a given subset. For instance, in the case of the papers linking 'Pathology' to 'Algorithm', we find a review paper describing the push for machine learning solutions to COVID-19 detection: "Artificial Intelligence in the Battle Against Coronavirus" [14]. In this way, it is possible to understand in detail, the patterns of multidisciplinarity that were identified at the distant reading level as FoS networks can help to identify novel review papers that bring together ideas from several different fields, papers which may have been hidden in more traditional citation network representations.

5 Conclusions and Future Work

In this work we propose Field of Study (FoS) networks as a novel representation for exploring the relationship between research topics at the macro-level. We describe the formation of two different types of FoS network, and provide case studies which illustrate how these networks can be used in the distant-reading of large corpora of research articles. In the case of network science research, we use FoS networks to explore the roles of different fields of study in multidisciplinary network science, and identify broad topics and applications in network science research. Similarly, in the case of COVID-19 research we investigate the relationship between fields of study within and between scientific disciplines to show

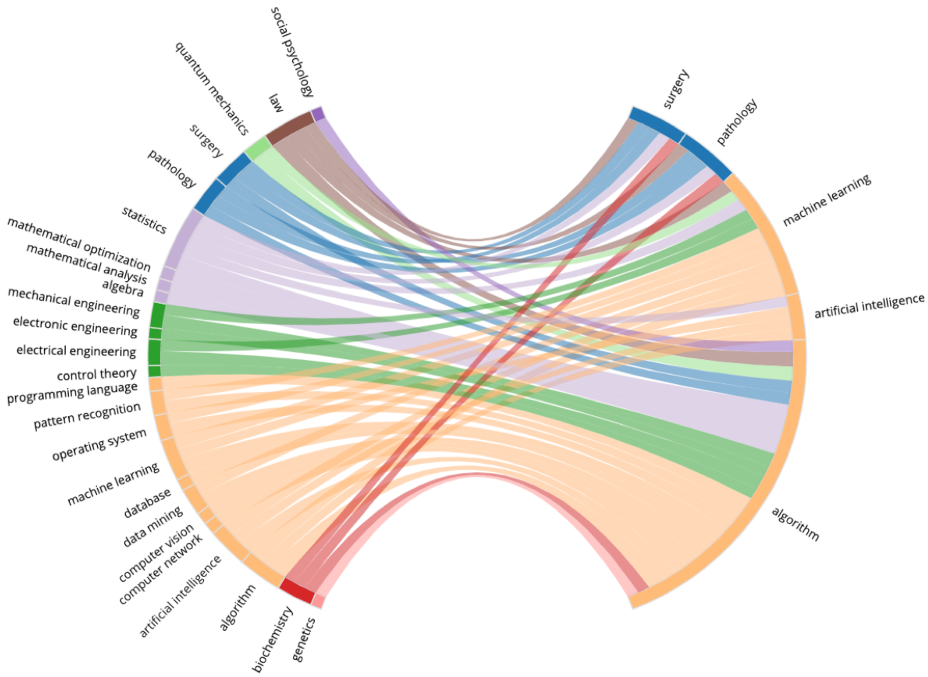


Fig. 6: Temporal FoS Network presenting COVID-19-related research in Computer Science, produced from 9,004 COVID-related research papers which were attributed to the MAG field 'Computer Science'.

an increase in multidisciplinarity in the context of COVID-19 research. Finally, we summarise the use of temporal FoS networks and methods of close-reading conducted on the COVID-19 research dataset in order to understand artefacts of multidisciplinarity identified in FoS networks.

There are a number of avenues for potential further research in this area. For example, in a corpus where full paper texts or abstracts are available, it may be informative to explore semantic relationships between the fields of study represented in the network. Similarly, citation information could be used to explore the flow or diffusion of information between communities. A multi-dimensional approach, which combines these methods (similar to that proposed by [18]), may prove a useful tool for scientometric analysis. Moreover, the FoS network we present may be used to explore multidisciplinarity, but not interdisciplinarity (as per the distinction offered in Section 2). Extending FoS networks to incorporate citation information may allow for the quantification of interdisciplinarity as many studies have used citation information to assess how articles "integrate" methods from different disciplines [16, 17].

Acknowledgments. This research was supported by Science Foundation Ireland (SFI) under Grant Number SFI/12/RC/2289_P2.

References

1. M. Arora and V. Kansal. Character level embedding with deep convolutional neural network for text normalization of unstructured data for Twitter sentiment analysis. *Social Network Analysis and Mining*, 9:1–14, 2019.
2. V. D. Blondel, J.-L. Guillaume, R. Lambiotte, and E. Lefebvre. Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, 2008(10):P10008, 2008.
3. A. Celik, J. Tetzner, K. Sinha, and J. Matta. 5g device-to-device communication security and multipath routing solutions. *Applied Network Science*, 4, 11 2019.
4. B. C. Choi and A. W. Pak. Multidisciplinarity, interdisciplinarity and transdisciplinarity in health research, services, education and policy: 1. Definitions, objectives, and evidence of effectiveness. *Clin Invest Med*, 29(6):351–364, Dec 2006.
5. E. Cunningham, B. Smyth, and D. Greene. Collaboration in the Time of COVID: A Scientometric Analysis of Multidisciplinary SARS-CoV-2 Research. *arXiv preprint 2108.13370*, 2021.
6. S. Feng and A. Kirkley. Mixing patterns in interdisciplinary collaboration networks: Assessing interdisciplinarity through multiple lenses. *arXiv preprint 2002.00531*, 2020.
7. W. Glänzel and A. Schubert. Analysing scientific networks through co-authorship. In *Handbook of quantitative science and technology research*, pages 257–276. Springer, 2004.
8. K. Karunan, H. H. Lathabai, and T. Prabhakaran. Discovering interdisciplinary interactions between two research fields using citation networks. *Scientometrics*, 113(1):335–367, 2017.
9. S. Lafia, W. Kuhn, K. Caylor, and L. Hemphill. Mapping research topics at multiple levels of detail. *Patterns*, 2(3):100210, 2021.
10. V. Larivière, S. Haustein, and K. Börner. Long-distance interdisciplinarity leads to higher scientific impact. *PloS one*, 10(3):e0122565–e0122565, 2015.
11. E. Leahey. From sole investigator to team scientist: Trends in the practice and study of research collaboration. *Annual Review of Sociology*, 42(1):81–100, 2016.
12. E. Leahey, C. M. Beckman, and T. L. Stanko. Prominent but less productive: The impact of interdisciplinarity on scientists' research. *Administrative Science Quarterly*, 62(1):105–139, 2017.
13. F. Moretti. *Distant reading*. Verso Books, 2013.
14. T. T. Nguyen, Q. V. H. Nguyen, D. T. Nguyen, E. B. Hsu, S. Yang, and P. Eklund. Artificial Intelligence in the Battle against Coronavirus (COVID-19): A Survey and Future Research Directions. *arXiv preprint 2008.07343*, 2021.
15. K. Okamura. Interdisciplinarity revisited: evidence for research impact and dynamism. *Palgrave Communications*, 5(1):141, Nov 2019.
16. A. Porter, A. Cohen, J. David Roessner, and M. Perreault. Measuring researcher interdisciplinarity. *Scientometrics*, 72(1):117–147, 2007.
17. I. Rafols and M. Meyer. Diversity and network coherence as indicators of interdisciplinarity: case studies in bionanoscience. *Scientometrics*, 82(2):263–287, 2010.
18. J. Raimbault. Exploration of an interdisciplinary scientific landscape. *Scientometrics*, 119(2):617–641, 2019.
19. Z. Shen, H. Ma, and K. Wang. A web-scale system for scientific knowledge exploration. *arXiv preprint 1805.12216*, 2018.
20. L. Wu, D. Wang, and J. A. Evans. Large teams develop and small teams disrupt science and technology. *Nature*, 566(7744):378–382, Feb 2019.