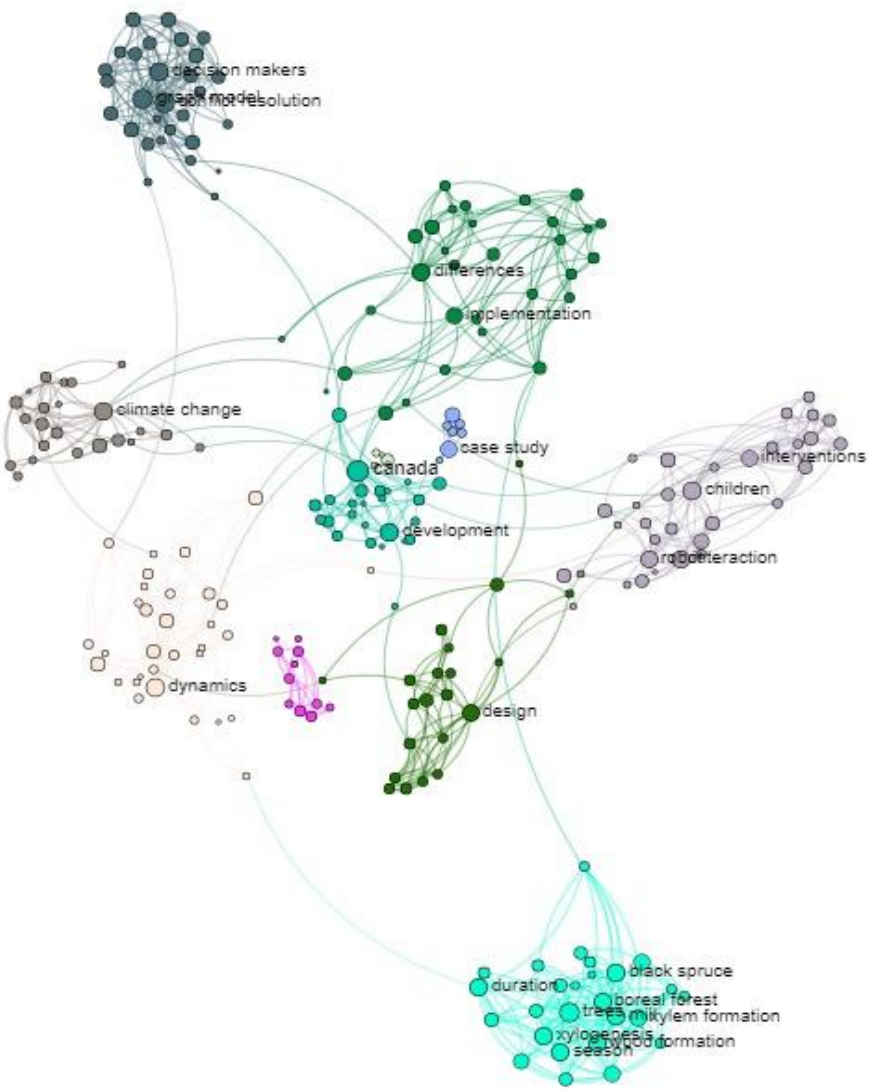
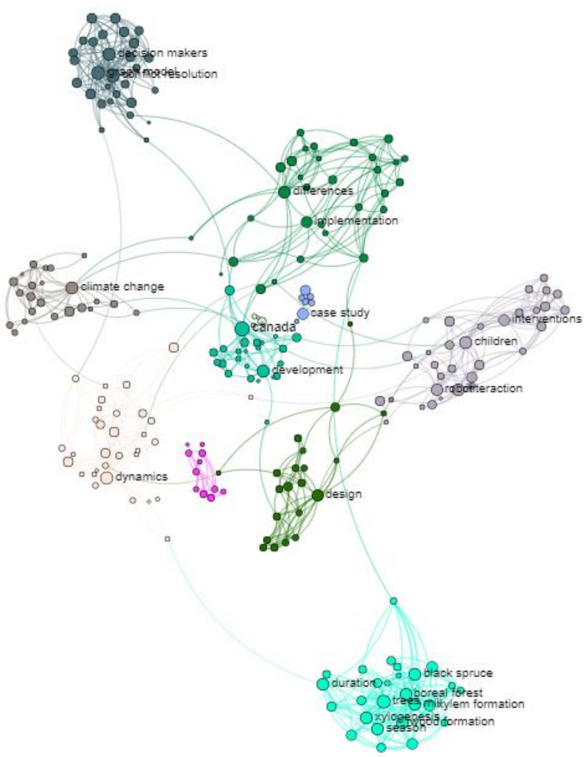
WATERLOO INSTITUTE for COMPLEXITY & INNOVATION



WATERLOO INSTITUTE for COMPLEXITY & INNOVATION



MAPPING CANADIAN COMPLEX SYSTEMS SCHOLARSHIP

A Bibliometric and Network Analysis

ISBN: 978-0-9918878-7-3

ABSTRACT

This study reports a bibliometric and network analysis of complex systems research from 2009-2018 within the Waterloo Institute for Complexity and Innovation, across the University of Waterloo, and across Canada. Study objectives were: (1) to assess the WICI member community in terms of overall research performance, areas of research, and applications of complexity science; (2) to identify complex systems scholars at the University of Waterloo who are not yet connected to WICI; and (3) to conduct preliminary Canada-wide literature search for scholars engaging in complex systems research. As part of the work, we developed a customized "complex systems" keyword query. Based on analysis from several analytical tools/platforms, the report lays the groundwork for deeper understanding and more regular assessment of the community's research activities. In particular, the study finds that WICI members' research fall into nine thematic clusters, and that there is significant activity on topics related to health and nutrition, and environment and climate change. The project also identifies complex systems research/researchers at the University of Waterloo and at other institutions across Canada, in support of a future Canadian Network of Complex Systems Scholars. The report highlights a lack of keyword standardization across the literature, creating challenges for automated community mapping. Finally, the report offers recommendations to mobilization of these data and findings to strengthen WICI and support complex systems research Canada-wide.

Jinelle Piereder, Lead Author; Dawn Cassandra Parker, Supervisor and collaborating author

October 2020

Table of Contents

List of Tables and Figures

Tables

Figures

Executive Summary

Objectives

Our goals for this project were to better understand (1) the WICI member community in terms of overall research performance, areas of research, and applications of complexity science; (2) to identify complex systems research and scholars at the University of Waterloo that is/are not yet connected to WICI; and (3) to conduct preliminary Canada-wide literature search for scholars engaging in complex systems research. All these objectives serve to aid WICI in positioning itself as a leading Canadian complex systems research institute and lay the foundations for a Canadian Complex Systems Network.

Design

This study reports a bibliometric analysis via a detailed examination of WICI members (including members external to UW) research activity from 2009-2018 and a further systematic search of the academic literature applying complexity science to any and all disciplines and subject areas across the University of Waterloo and Canada. We conducted the literature search in Scopus and then exported the search results to SciVal—Elsevier's analytics platform—for more detailed analysis and reporting. We also used the exported the WICI community results in Gargantext and VOSviewer—two platforms for citation and network analysis, and community and theme detection.

Methods & Data

To understand the research profile of existing researchers in the WICI Member Community, we identified three overlapping non-student researcher *populations*: *Core Members* (UW researchers, 11 authors); *Extended Core Members* (including core and external core members, 17 authors); and *All Members* (60 authors). We identified these members in Scopus and added them to respective Saved Author lists. We also added each of these authors as entities in SciVal for individual analysis and reporting. These lists served as the populations for the rest of the analyses and enabled more streamlined subsequent literature searches (using Scopus Author ID's).

To build our initial *corpus*, we collected self-reported complex systems-related publications from the past ten years' Annual Reports and Productivity Reports. This first corpus provided a validation dataset for identifying complex systems research. Further corpora were created using the Scopus IDs of the populations defined above in a search query and adding the results to Saved Documents lists. Given the university's desire to map health-related research, we additionally generated a corpus that included All WICI Members publications that related to health and medical subject areas or topics, from 2009-2020. Note that these corpora only include Scopus-listed publication outlets. Note also that we expect these corpora to include non-complex systems scholarship.

To identify the subset of members' publications related to complex systems, we developed a broad keyword list and search query (based on a survey of WICI members and complex systems identifiers of other institutes) to further "filter" each corpus (see section 3). We later used the query again to identify complex systems scholars and research across the University of Waterloo and Canada as a whole. Each corpus (both those filtered by the complex systems query and those not filtered) was exported to SciVal (known on that platform as Publication Sets) for more detailed analysis and reporting.[1]

To gain deeper insight into WICI Members scholarship, one corpus—*All Member' publications 2009-2018*—was also imported into Gargantext and VOSviewer for further bibliometric and network analysis.

Results: Scopus and SciVal

Our first key finding relates to the bibliometric search process itself. Our complex systems search query captured only (1) 25 per cent of the *Extended Core Members' self-reported publications* (which were all self-identified as related to complex systems—this corpus was meant to serve as our validation set), (2) 17 per cent of the *Core Members' publications in Scopus*, (3) 19 per cent of the *Extended Core Members' publications in Scopus*, and (4) 16.5 per cent of *All Members' publications in Scopus.* The results of our complex systems search query suggest one of three things: (1) that either the majority of publications by WICI members are *not* related to complex systems, (2) that our complex systems search query is missing some important terms, or (3) that WICI members do not always explicitly use complexity terminology in their work, even when their underlying ontology is grounded in a complex systems perspective.

To test these possibilities, we compared Extended Core member's self-reported publications to the complex systems publications identified by our query. While our query captured 31 of the 124 self-reported publications we identified in Scopus, in another search, it also captured 150 of the total 889 publications in Scopus by Core and External Core members. Our findings suggest that our query picked up some publications by Core or External Core Members that they did not self-identify as being complex systems related. To validate this finding, future work could involve providing a list of those additional publications to the authors and confirm whether they are, indeed, complex systems research. We could do the same with the full WICI membership to determine whether our query missed publications that the authors consider to be complex systems research.

We took a closer look at the terms in our search query compared to the official Author Keywords attached to the publications in each corpus. Taking the subset of publications by Extended Core members that was *not* captured by our complex systems query, we extracted the official Author Keywords from each publication. We identified as many as 40 keywords (of approximately 300) that are or may be related to complex systems but were not included in our search query. Many of these keywords represent the research of one or two authors of the Extended Core Members population (e.g., Keith Hipel's graph model, or Vanessa Schweizer's scenario work). However, some of them capture areas of complex systems research that are not well-represented in our search query, and should be included in future bibliometric searches (e.g., dynamic analysis, robust decision making, vulnerability based scenario analysis, human-environment coupling, relational ontology, systems thinking).

We conducted further keyword analysis on the set of publications from our UW-wide complex systems search, first by calculating keyword frequency and then by comparing the list of most frequent terms (58

[1] At the time of analysis, SciVal had not yet completed its' 2020 "rollover" (which occurs in June) where all 2019 publications were considered added, and the year "complete". So, while we could identify 2019 publications in our Scopus search, they could not be included in the SciVal publication sets for analysis (though they could be easily updated in future years). Further, the Canada-wide query results, which were too large to export to SciVal, had to be identified endogenously within SciVal via their "Research Area" feature.

keywords used at least 10 times) to the list of terms in our own query. Approximately 35 of our search terms were present as Author Keywords in the dataset, meaning 44 were not (note that some of our search terms may still have been present in the abstracts of publications). A review of the full set of Author Keywords showed that approximately 70 terms/phrases could be considered relevant to complex systems research. Future bibliometric research on complex systems should consult these lists to further refine a complexity search query.

Beyond corpus comparison and keyword analysis, a key outcome of this project is the thorough and highly detailed reports we produced in SciVal for individual authors, our various WICI member corpora, and our broader UW and Canada search results. All these results are included in the Supplement Report accompanying this document. We produced Author Reports for each of the WICI Core Members and External Core Members (see pages 1-290 of the Supplemental Report), which include analyses on research performance, keyphrases, subject areas and topics, most cited publications, and co-authorship and collaboration. We then produced Corpus or Publication Set Reports for each sub-population within the WICI community (see Appendix B for full list; see full reports starting on page 291 of the Supplemental Report). These include analyses on performance summaries related to citation percentiles and journal percentiles, keyphrases, subject areas and topics, author and institutional collaboration, top authors and journals. we produced similar Publication Set Reports for the UW and Canada-wide search results (see the UW report on pages 618-686 of the Supplemental Report, and the Canadian report on pages 687-840).

Both the Author Reports and Corpus/Publication Set Reports are useful for assessment, benchmarking, and getting a clearer picture of scholarly output and research activities of WICI members. One key finding is that over 52 per cent of complex systems related publications by WICI members (2009-2018) are co-authored with internationally-based scholars. Furthermore, 21 per cent of these publications are in the top 10 per cent most cited worldwide, and 48 per cent of them are published in the top 10 per cent of journals (by CiteScore percentile).

The Reports also tell us about the makeup of WICI member and UW-wide research in terms of disciplines, and the differences in makeup across several corpora. For example, in our SciVal analyses of publications by subject area, we find evidence that (1) both environmental science and social science account for a larger proportion of the WICI member complex systems subset than they do in the full set of WICI member publications, as well as in the UW-wide complex systems search results; (2) there is a significant amount of complex systems research being conducted in engineering, mathematics, physics and astronomy, and materials science that is *not* well represented in the WICI membership; (3) WICI members account for the majority of complex systems related work in the arts and humanities, immunology and microbiology, and multidisciplinary research happening across UW campus; and (4) complex systems related research is more often multidisciplinary than non-complex systems work at UW.

Finally, the Author and Corpus Reports offer some direction for the developing Canadian Network of Complex Systems, and for tapping into existing international complexity networks/communities. Our analysis of the UW community yielded the names of approximately 355 (of the top 500) authors engaged in complex systems research (as per our search query), only 16 of whom are already WICI members (see page 53 of this Report for the list of members). Additionally, the lists of “Collaborating Authors” in each Extended Core Member’s Author Report and the full lists of “Top Authors” across UW and Canada (see page 60 of this report) may provide leads on potential new WICI members (at UW or external). Similarly, the lists of "Collaborating Institutions" in the Author Reports and the lists of "Top Institutions" across Canada (see page 61 of this report) may provide leads on potential future nodes for the Canadian Network of Complex Systems, as well as international partners.

The Reports also give us insight into where WICI Members are publishing most often—a computer science journal and an environmental modelling journal ranked highest, but we found that members publish very broadly—and which journals are most receptive to complex systems research. Finally, while they are too numerous to list here, the reports give us a sense of the main topics and topic clusters that WICI members contribute to, which can help us identify research trends, gaps, and opportunities within the community (see the individual Author Reports and each Corpus Report's Topics Wheel visualization in the Supplemental Report).

Results: Network Analysis

Using Gargantext, and the *All WICI Members' All Publications 2009-2018* publication set*,* we identified eight thematic clusters based on a co-word analysis. These clusters consisted of words related to: (1) the graph model of conflict resolution, (2) forestry research, (3) disease dynamics and vaccination, (4) ecosystem services and management, (5) child health and health interventions, (6) diet, food, and assessment design, (7) Canada and development, (8) and climate change and adaptation. It is clear from looking at the sets of terms that some of these thematic clusters represent the work of just a few scholars (e.g., cluster 1 obviously captures Keith Hipel and colleague's work on the graph model of conflict resolution; cluster 2 represents Madhur Anand, Chris Bauch, and Dawn Parker's research related to trees). Others, however, cover a broader set of work despite still clustering around key terms (e.g., cluster number nine on climate change is clearly quite multi-disciplinary, including terms related to environmental issues, social concerns, data and methods, and different geographies). In these cases, there may be several authors represented in a single cluster who are not actually aware of each other's work or who do not collaborate yet.

In addition to the contents of the clusters, the relationships between the clusters offer more insights. For example, we found that the more author-homogenous clusters also tend to have fewer outward connections to other clusters (e.g., clusters 1 and 2). This observation points to potential untapped collaboration opportunities and suggests a need to better control for publication counts when working across disciplines. Furthermore, some terms serve as clear "bridges" from one cluster to another (e.g., clusters 3 and 6 are linked via "social innovation," and clusters 5 and 6 are linked via "interventions" and "physical activity"), demonstrating what network science has—for decades—called "the strength of weak ties." These weak ties, or bridges, could point to potentially novel areas of collaboration, and even offer empirical support to multidisciplinary funding applications.

Using the same dataset, we conducted a co-occurrence analysis of terms (similar to the co-word analysis above) using another tool—VOSviewer—and identified six thematic clusters. Three of the clusters were similar to the Gargantext results (1) the graph model and conflict resolution, (2) health and (3) environment. The other three clusters were related to (1) international trade, (2) engineering problems and solutions, and (3) behaviour (of systems, people, study participants, governments, and robots). We find a significant, but not complete, alignment of these clusters with the SciVal-defined subject areas. Notably, environmental science, social science, and medicine/health are more strongly represented in the network analysis in both software platforms than in the SciVal subject area analysis. This finding underlines the importance of augmenting standard bibliometrics with additional ways of assessing impact, prevalence, and epistemic boundaries of a research field.

VOSviewer also enables more *types* of analysis because the unit of analysis (i.e., the network nodes) can be publications, journals, authors, or keywords (Gargantext uses only automatically extracted or manually added terms). Each of these analyses provide different ways to examine all the relationships that exist in a set of publications. In addition to a co-occurrence of terms analysis (similar to the Gargantext analysis), we also conducted VOSviewer analyses related to co-authorship, journal co-citation and bibliographic coupling, and keyword co-occurrence. These network analyses reveal (1) thematic clusters (the keyword co-occurrence analyses were more effective in identifying tighter, more coherent clusters), (2) the relatedness of WICI members' research and possible future author collaborations (see Figures 19a-d for specific observations), and potential publication targets for complex systems research, among other insights. These analyses also validate many of the results from the SciVal Author Reports and Publication Set Reports.

Future versions of this report should conduct network analysis on multiple publication sets in an attempt to narrow and/or nuance some of these findings.

Conclusions

This study represents the first comprehensive bibliometric and network analysis of the WICI, UW, and Canadian complex systems scholarship community. Through its design and integration of analytical tools/platforms, we lay the groundwork for more regular assessment of the community's research activities, which we anticipate will enable and strengthen complex systems research at the University of Waterloo and across Canada.

Our SciVal analysis and reporting, specifically, offers many insights and leads for growing WICI and establishing a Canadian Network of Complex Systems by identifying potential future partners, collaborations, and co-authors. Our search revealed at least two dozen scholars who are based at UW and engage in complex systems research but are not yet connected to WICI, and hundreds more across Canada. We also highlight the relative impact and degree of international collaboration of complex systems research at the University of Waterloo, and by WICI members especially. Fifty-seven per cent of the publications by Extended Core members that were captured by our complex systems query had international co-authorship. As for the whole WICI membership, 52.4 per cent of publications had international collaboration. Furthermore, 50.7 per cent of UW-affiliated publications that we captured in our complex systems search query were co-authored with institutions in other countries (see pages 463, 553, and 619 of the Supplemental Report). This high level of international collaboration points to the value of complex systems research as an avenue for building institutional partnerships and increasing impact.

With Gargantext and VOSviewer, we identified the emergence of several thematic clusters, which could contribute to setting up various thematic working groups, journal special issues, and events that will engage WICI members more effectively. Within WICI, we see especially strong trends towards complex systems and health research, as well as environment and climate change related research. This report's analysis validates WICI's 2019 proposal to "brand" around complex human-environment interactions. We also see the thematic focal areas identified in 2019 appearing, to some extent, in our bibliometric mapping (e.g., complex coupled human-natural systems, complex health systems, and multi-scale adaptive management and optimization of complex spatial and network problems) (see page 5-6 of the 2019 WICI Annual Report).

Our most important finding, perhaps, is that it is uniquely difficult to effectively and accurately identify "complex systems" research and scholars within large databases. This is partly due to the multidisciplinary nature of a lot of complexity research, but also to a relative lack of self-identification and an abundance of "chaff"—i.e., non-technical use of complex systems terms. Challenge like these act as a hindrance to cross-disciplinary communication, collaboration, and overall accumulation of complex systems science. However, we suggest that a more formal interdisciplinary Canadian Network of Complex Systems could help address this problem by identifying standard complex systems keywords and encouraging members to include them in publications. Funding agencies could then include "complex systems science" as a disciplinary category, enabling scholars to self-identify. More regular analyses, systematic literature reviews, and interdisciplinary reports centred on complex systems research—such as the current report—would also help to consolidate the relatively scattered efforts of Canadian scholars working in this exciting and growing field.

1.0 Introduction

Mapping exercises are increasingly important to inform organizational strategy and assess impact; this is as true for research networks as it is for large institutions. Social network mapping is key for organizational positioning and identifying key players and possible collaborators. Additional bibliometrics and informetrics are important for answering other kinds of questions about emerging topic clusters and epistemic communities, how clusters and communities relate to each other, important research trends, gaps, impact, state-of-the-art assessments, and alternative perspectives.

With these questions in mind, WICI commissioned the lead author to map the network of complex systems scholars at the University of Waterloo, and across Canada. The goal of this mapping exercise was to:

1. evaluate the research activities of WICI members,
2. increase our awareness of complex systems scholars and research,
3. help us better understand how scholars situate themselves as complex systems researchers, and
4. enable more collaboration across institutions, with the eventual goal of establishing a Canadian Complex Systems Network.

The project consisted of several simultaneous research activities, each of which contributed to the others in an iterative way. We (1) conducted a survey of existing WICI members, (2) identified important keywords and tags related to the disciplines, methods, and application areas of complex systems research, (3) developed a Scopus search query based on our keyword database, and (4) conducted bibliometric and network analysis of search results using several cutting-edge tools (Elsevier's SciVal, VOSViewer, and Gargantext).

The WICI community is made up of scholars in many faculties and departments across the University of Waterloo, as well as several other institutions. According to WICI's website,

> **Core Members** are regular, research, or adjunct university faculty who lead a long-horizon research program under the institute's auspices.
>
> **External Core Members** are regular, research or adjunct faculty, outside the University of Waterloo, who lead a long-horizon complex systems research program and actively engage with WICI networks and activities.
>
> **Affiliate Researchers** are regular, research, or adjunct university faculty or non-university researchers, including post-doctoral fellows, who actively participate in institute activities, including its research projects or committees.
>
> **Practitioner Members** include people in government, the voluntary sector, and private sector interested in the institute's research and findings and who actively participate in WICI meetings, workshops, and conferences open to a general audience.[2]

[2] "People: Categories of Membership," Waterloo Institute for Complexity and Innovation (WICI), University of Waterloo, available at: https://uwaterloo.ca/complexity-innovation/people.

We focused our bibliometric analysis on the following four groups or "populations" of scholars and their publications, imagined as four concentric circles. At the centre, our first and main focus for bibliometric analysis was the WICI Core and External Core members (including those based at the University of Waterloo and those based at other institutions[3]). Following that group, we looked at the research areas and activities of the full WICI membership (excluding students but including researchers and practitioners). Beyond WICI, we looked at the wider University of Waterloo community to identify current or potential co-authors and WICI members, based on their engagement with complex systems research. Finally, we expanded our search and analysis to all Canada-based complex systems scholars, which we identified using our comprehensive search query.

These "populations" of authors formed the framework from which to build corpora or publication sets. We conducted several kinds of analysis on those corpora (using the tools described in the following section).

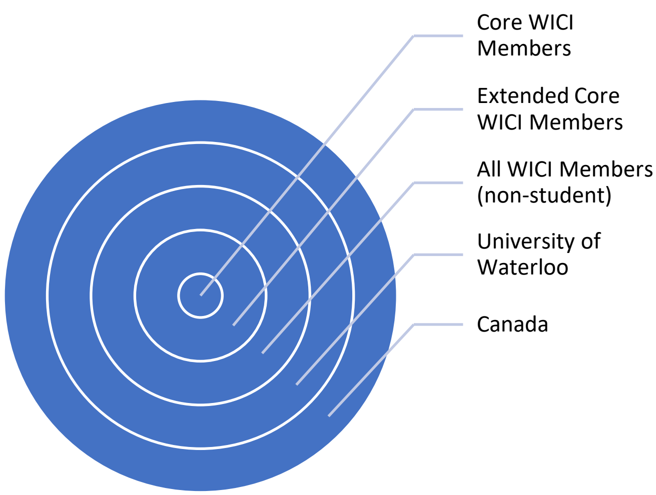


Figure 1: Visualization of research populations

In the following section, we describe the overall methodology of the project and outline our process for developing a complex systems search query. We then describe and compare each of the three analytical tools we use in the rest of the report. Section 3.0 focuses on the data, analyses, and findings related to the WICI Member Community (including the sub-populations outlined above and their relevant corpora). Section 4.0 and 5.0 look at the University of Waterloo and Canada, respectively. Finally, in section 6.0, we offer a synthesis of our findings, some recommendations for using these findings, and a conclusion with suggestions for future research.

2.0 Methods & Data

Since the main purpose of this project was a broad mapping exercise, we designed our methodology to consider multiple populations and corpora and take advantage of the unique capabilities of several analytical tools.

The first step of this mapping project involved a Qualtrics survey of current WICI members. In it, we asked them to highlight the top three disciplines with which they identify, three areas of study where they apply a complex systems lens and/or approach, and three methods they use. These keywords and phrases also fed into our development of a database search query. In the survey, we also asked for citations of what participants would consider their most important complex systems research

[3] The majority of members are based at Canadian institutions. Two members in this study are in the United States, and one is in New Zealand, all of which are Affiliate Members.

publications. A total of thirty-five people completed the survey, and the results of the Qualtrics survey can be found in WICI's 2019 Annual Report.[4]

Next, we developed a complex systems search query to (1) try to see how much WICI member research activity we could "catch" with our keyword string, and (2) to try to identify other potential complex systems research and scholars across the University of Waterloo and across Canada. This process was less straightforward than one would think. While many other review articles looking at complex systems simply limit their search terms to "complexity theory," "complexity science," "complex adaptive system," or "complexity thinking," we knew that this limited language misses a significant amount of relevant work. First, not every scholar who employs complex systems thinking, concepts, tools, or metaphors explicitly uses these keywords. Particularly in subfields where complexity has had significant purchase for a decade or more, researchers can assume some base level knowledge of their readers and have no need to specify that their work is grounded in a complex systems perspective; they can jump right to more technical terms such as "cellular automata" or "agent-based modeling," where a complexity lens is implied. Second, because complexity has been adopted across so many disciplines and fields, some complexity concepts are more prevalent in some areas than others. Finally, it is an ongoing debate as to which concepts in complexity thinking connote the "core" or most important concepts and which ones sit at the "periphery." To be sure, and beyond perhaps four or five terms, importance largely depends on what researchers have found most useful in their work.[5]

So, to build the search query for this project, we decided to cast a wide net. Keeping in mind the "chaff" that the search would likely gather in addition to the actual complex systems research, it was nevertheless important for WICI's mission and vision to gain as broad a view as possible at this mapping stage. Since the Santa Fe Institute in New Mexico is one of, if not the, leading global complexity organization, we began to assemble a keyword database based on their complexity glossary.[6] We also gained access to a search query by the Complex Systems Institute of Paris that was used in their 2008-2009 Complexity Roadmap.[7] We further supplemented our database with keywords from the index of Melanie Mitchell's *Complexity: A Guided Tour*[8] and from the results of the WICI Member Survey. In total, we identified 225 keywords (plus variants). Drawing on our knowledge of the field, we eliminated those terms that could have a non-complex systems use or were so general that they would pick up significant chaff. In the end, we refined the list to 78 keywords plus their variants (the complete query is defined in Appendix A).

Following the development of our search query, we turned to defining our populations and corpora. To define our *populations*, we found WICI members using their Scopus identification numbers and built

---

[4] Dawn Parker, Brenda Panasiak, Vanessa Schweizer, Peter Deadman, Chris Bauch, Chrystopher Nehaniv, Sharon Kirkpatrick, and Igor Grossmann, 2019, "Waterloo Institute of Complexity and Innovation 2019 Annual Report," WICI, available at: https://uwaterloo.ca/complexity-innovation/sites/ca.complexity-innovation/files/uploads/files/2019_wici_annual_report_final_0.pdf.

[5] A potentially valuable line of future research could be a survey of WICI members asking them to identify what they believe to be the top "core" concepts of complexity, as well as the ones they most use in their research.

[6] "Glossary," Complexity Explorer, Santa Fe Institute, available at: https://www.complexityexplorer.org/explore/glossary.

[7] David Chavalarias, Paul Bourgine, Edith Perrier, Fréderic Amblard, François Arlabosse, et al., 2009, "French Roadmap for complex Systems 2008-2009," HAL, available at: https://hal.archives-ouvertes.fr/hal-00392486.

[8] Melanie Mitchell, 2009, *Complexity: A Guided Tour*, Oxford University Press: New York, NY.

several Scopus Saved Author lists: *Core Members* (11 authors), *Extended Core Members* (17 authors), and *All Members* (60 authors). These Saved Author lists served as the populations for the rest of the analyses and enabled more streamlined subsequent literature searches (using Scopus Author ID's). We also added each of these authors in SciVal for individual analysis and reporting.

To build our initial *corpus*, we collected the self-reported complex systems publications by *Core Members* (internal an external to UW) from the past 10 years' Annual Reports and Productivity Reports (233 documents—but 109 were not identified in Scopus). We created additional corpora using the Scopus IDs of the populations defined above and saving the search results as Saved Documents lists: all *Core Members'* publications 2009-2019 (588 documents), all *Extended Core Members'* publications 2009-2019 (780 documents), and *All Members'* publications 2009-2019 (2,162 documents). We then put our complex systems search query to use by "filtering" each of the corpora through the query (these results are listed in section 3).

Beyond the WICI member community, we wanted to find complex systems research and scholars at the University of Waterloo that is/are not yet connected to WICI. We also wanted to develop a preliminary understanding of complex systems research happening across Canada. To do this, we ran our complex systems query in Scopus, first limited by institutional affiliation to the University of Waterloo (3,671 results from 2009-2019, 3230 of which we could export to SciVal), and then limited by country to Canada (67,451 results from 2009-2019; only publications from 2014-2018 could be analyzed SciVal, i.e., 238,848 documents). These results were also exported to SciVal.

Because this study is limited to bibliometric search results found in the Scopus database, there are potentially publications by WICI members that are not included in the corpora we analyzed, especially books and book chapters (due to Scopus's emphasis on academic journals). For example, a total of 109 publications that were self-reported by WICI core or external core members were *not* found in Scopus. It is also possible that our research design privileges some fields over others. Much of our analysis is based on keyword frequency or prevalence, and fields like engineering, mathematics, and health sciences publish shorter articles in larger quantities than the social sciences or even environmental sciences. A White Paper produced by a University of Waterloo Working Group on Bibliometrics also points to the following limitations of citation-tracking databases (such as Scopus) and the research analytic tools that draw on them (such as SciVal): different databases use different methodologies, some of which can be especially limiting to some disciplines; not every type of publication is indexed and comprehensive coverage is not possible; citation practices in the academy are also gender biased; and the time it takes to assess "impact" differs across disciplines, which may lead to imbalanced results when considering cross-disciplinary research.[9] Future research could attempt to control for some of these limitations and field characteristics, as well as augment the research design to include more qualitative and expert-driven assessments.

After defining all our populations and corpora, we exported each of the Scopus search results (in the form of Saved Author Lists and Saved Document Lists) to SciVal for more detailed analysis (known on that platform as Publication Sets). There, we conducted bibliometric analysis on the various datasets and produced reports using SciVal's Reporting feature. However, at the time of analysis, SciVal had not yet

[9] University of Waterloo Working Group on Bibliometrics, 2016, "White Paper on Bibliometrics, Measuring Research Outputs through Bibliometrics," Waterloo, Ontario: University of Waterloo.

completed its' 2020 "rollover" (which occurs in June) where all 2019 publications were considered added, and the year "complete". So, while we could identify 2019 publications in our Scopus search, they could not be included in the SciVal publication sets for analysis. In future years, this SciVal analysis and reporting could be done over the summer to ensure a more up to date picture of the WICI community.

Finally, to gain deeper insight into the WICI Membership, the corpus *All Member' publications 2009-2018* was also imported into Gargantext and VOSviewer for further bibliometric network analysis (including co-occurrence analysis, citation analysis, and bibliographic coupling analysis). Due to researcher capacity, we limited the network analysis in this project to just this one corpus. Future research could easily replicate the VOSviewer and/or Gargantext analysis using additional publication sets.

Throughout the project, we maintained a relational database using Airtable[10] to keep track of WICI Members, associated institutions, publications, Scopus search queries, lists of keywords, project datasets, completed analyses on *populations* and *corpora,* and lists of reports. WICI will maintain ownership of this database. More detailed explanations of our datasets are included in each of the Community case studies below.

2.1 Analytical Tools

This report uses three tools for bibliometric and network analysis. We had already determined to use Scopus as our database of choice for this project, so using its companion bibliometrics platform was the best option. We chose VOSviewer over other bibliometric network software because of its accessibility and breadth of training and tutorials, as well as its ability to interface with many file types. Lastly, while Gargantext is not widely used, it is firmly grounded in a complex systems approach to research, and we were interested to see how its capabilities compare with other similar platforms.

First, SciVal[11] is a web-based analytics platform from Elsevier that enables research performance evaluation and visualization. It can help researchers identify and analyze emerging research trends, evaluate research activities at the individual-scholar and institution levels, create unique reports with a variety of key metrics, and more. One key limitation of SciVal is that it can only include publications for analysis that were published up to and including the previous year. Each June, SciVal performs a "rollover" and declares the previous year's publication set complete, and thus available for full analysis using their metrics/platform.

The second tool, **VOSviewer[12]**, is one that is increasingly used by university libraries or organizations seeking innovative ways to measure and analyze researcher activity. VOSviewer was developed by Nees Jan van Eck and Ludo Waltman at the Centre for Science and Technologies Studies, Leiden University, in The Netherlands. It is a free desktop tool for constructing and visualizing bibliometric networks and scientific landscapes. It can help researchers identify important scholars,

[10] https://www.airtable.com
[11] https://www.elsevier.com/solutions/scival
[12] https://www.vosviewer.com/

publications, and research hubs; visualize overviews of various organization's work; conduct various kinds of impact analysis; and more effectively communicate research and results. VOSviewer emphasizes bibliometric networks of journals, researchers, and individual publications; but it can be used for any kind of research positioning and trend identification where you have a collection of documents from which to work.

Finally, Gargantext[13] is an open-source online platform for exploring sets of unstructured documents and digital data. Developed by David Chavalrias and colleagues at the Complex Systems Institute of Paris (ISC-PIF), Gargantext combines natural language processing, text mining, complex network analysis, and interactive data visualization. Similar to VOSviewer, Gargantext can be used to build an adaptive representation of a question or problem related to a research field. But beyond that, Gargantext can handle large amounts of textual data of varying lengths, including things like tweets, news articles, or even entire digital books (upwards of 10,000 network nodes). It can help researchers identify thematic clusters and emerging research communities; identify trends and gaps and position our own work; and analyze social media conversations around climate change and energy transition issues. (See Appendix C for a full tour of Gargantext with screenshots.)

The table below compares these three tools based on criteria related to data, analysis, visualization, and special features.

Table 1: Comparison of Tools

[14]	SciVal	VOSviewer[15]	Gargantext
Types of data	Publications in Scopus Text files with publications IDs (DOI, PMID, or EID), up to 50,000 publications	Bibliographic or text data from databases (supported types: Web of Science, Scopus, Dimensions, and PubMed) Bibliographic or text data from reference manager files (supported: RIS, EndNote, and RefWorks) Bibliographic or text data through API (supported: Microsoft Academic, Crossref, Europe PMC,	Bibliographic data from databases (Web of Science, Jstor, Scopus, PubMed, SCOAP, REPEC) Bibliographic or text data in CSV format (including non-academic data like tweets, news articles, digital books, etc.) Zotero RIS files

[13] https://gargantext.org/

[14] Nees Jan van Eck and Ludo Waltman, "VOSviewer and CitNetExplorer Tutorial," 15th International Conference on Scientometrics & Informetrics, Istanbul, Turkey, June 29, 2015, available at: https://www.slideshare.net/NeesJanvanEck/issi2015-tutorial-vosviewerandcitnetexplorer

[15] We chose to use VOSviewer over its sister software, CitNetExplorer, because the latter can only produce direct citation networks of publications, whereas the former can produce any type of bibliometric network. However, CitNetExplorer also incorporates a time dimension, which explicitly interest some researchers. Further, the software can support millions of publications and tens of millions of citation relations (but it works best with Web of Science). Future versions of this report could also produce citation networks using CitNetExplorer. Available at: https://www.citnetexplorer.nl/

		Semantic Scholar, OCC, COCI, and Wikidata) Existing network data (supported: GML or Pajek)	ISTEX data (a French academic database)
Quantity of data	Up to 50,000 publications, depending on the desired analysis	Max. 10,000 publications	Upwards of 10,000 network nodes
Techniques	Various research metrics, citation impact assessment, benchmarking, ranking [16]	Text mining, Advanced mapping and clustering, Smart labeling algorithm	Natural language processing, text-mining, complex networks analysis and interactive data visualization Users can completely customize the "map list" (set of terms used in analytics) Users can group equivalent terms (e.g., children and kids)
Types of analysis	Basic citation analysis and performance Collaboration analysis Subject area analysis Topics and Topic Clusters analysis (including prevalence database-wide) Keyphrase analysis (including growing or declining relevance)	Several kinds of network analysis, including: Co-authorship analysis: The relatedness of items is determined based on their number of co-authored documents. Co-occurrence analysis: The relatedness of items is determined based on the number of documents in which they occur together. Citation analysis: The relatedness of items is determined based on the number of times they cite each other. Bibliographic coupling analysis: The relatedness of items is determined	Complex networks analysis Temporal evolution analysis (of the whole corpus, or filtered by individual or multi-terms), with customizable granularity (decade, year, month, hour). These analyses can also be compared across corpora and can be represented as bar graphs or area. Internal *n*-grams (or temporal term frequency graphs, similar to Google's Ngram Viewer[17]) also comparable between document sets) Co-word analysis

[16] "Scival Research Metrics Guidebook," 2019, Elsevier B.V., available at: https://p.widencdn.net/5pyfuk/ACAD_RL_EB_ElsevierResearchMetricsBook_WEB
[17] https://books.google.com/ngrams

		based on the number of references they share. Co-citation analysis: The relatedness of items is determined based on the number of times they are cited together.	
Types of data visualizations	Pie charts, bar graphs, “tree maps,” “wheel charts,” word clouds, geographical maps, tables	Network maps	Term network map Histograms
Types of outputs	SciVal Reports (shared online or as pdfs) CSV data exports Publication lists	Co-authorship networks of authors or organizations Co-citation networks of publications, journals, or authors Bibliographic coupling networks of publications, journals, authors, or organizations Co-occurrence networks of keywords or terms extracted from titles and abstracts of articles Density and overlay visualizations on most networks	Term network GEFX files Corpus histograms Term lists, document lists
Misc.	Research entities, research areas, and reports are shareable with other scholars (who have SciVal access)		Gargantext offers a high level of granularity when exploring a corpus. In document view, users can follow the instances of terms in the Map List (and modify the List from there). Users can import their own preexisting list of terms. In map view, users can see the list of publications where each term is used.

3.0 WICI Member Community

The **Core Members** included in this study are: Chris Bauch, Mark Crowley, Peter Deadman, Igor Grossmann, Keith Hipel, Thomas Homer-dixon, Sharon Kirkpatrick, Chrystopher Nehaniv, Dawn Parker, Stephen Quilley, Vanessa Schweizer, and Paul Thagard. The **External Core Members** are: Madhur Anand, Liane Gabora, Mary O'Connor, Raja Sengupta, and Roger White. When referring to both Core and External Core Members, we use the term "Extended Core Members."

The **Affiliate Members** (researchers and practitioners) are: Anna Klinkova, Anna Lawniczak, Chris Perlman,Dan McCarthy, David A. Petrie, Derek Robinson, Ed Jernigan, Edward Thommes, Eihab Abdel-Rahman, Hans De Sterck, Hassan Masum, Ilias Kotsireas, Isaac Tamblyn, Jeremy Pittman, Jessica Blythe, John McLevey, John Whalley, Jon MacKay, Kerstin Dautenhahn, Luis Sandoval, Manjana Milkoreit, Mark Hancock ,Mark Tovey, Mark Weber, Matteo Smerlak, Matthew Hoffmann, Matto Mildenberger, Mohamed Tawhid, Monica Cojocaru, Neil Craik ,Owen Gallupe, Paul Fieguth, Peter Carrington, Ponnu Kumaraswamy, Rebecca Saari, Reza Yousefi-Nooraie, Rob Robson, Robert Spekkens, Sarah Burch, Sarah Tolmie, Scott Heckbert, Sergio Rossi, Shreyas Sundaram, Simron J. Singh, Steven Mock, Tara Vinodrai, Tejal Patel, Virginia Capmourteres, William Sutherland, and Xiongbing Jin. We did not include Student Members in the current study.

3.1 Scopus & SciVal Analysis

3.1.1 Scopus & SciVal Data & Methods

After collecting publications using the Scopus IDs of the authors listed above, our initial dataset included the following corpora:

- *Extended Core Members' self-reported publications* 2009-2019 (233 documents—but 109 not found in Scopus),
- all *Core Members'* publications in Scopus 2009-2019 (588 documents),
- all *Extended Core Members'* publications in Scopus 2009-2019 (780 documents), and
- all *All Members'* publications in Scopus 2009-2019 (2,162 documents).

We then put our complex systems search query to use by "filtering" each of the corpora through the query. The filter gave us the following results:

- 31 of 124 documents (25 per cent) in *Extended Core Members' self-reported publications* in Scopus 2009-2019,
- 102 of 588 documents (17 per cent) in all *Core Members'* publications in Scopus 2009-2019,
- 150 of 780 documents (19 per cent) in all *Extended Core Members'* publications in Scopus 2009-2019,
- 356 of 2162 documents (16.5 per cent) in *All Members'* publications in Scopus 2009-2019.

The two tables below show a further breakdown of how many publications were captured by our complex systems query in Scopus, first sorted by corpus and second by the membership status of the author(s) Note that publications fitting in multiple corpora or membership statuses have been split, so the totals at the bottom reflect the actual number of publications, not the sums of the columns.

The results of our query suggest some possible preliminary insights: (1) that either the majority of publications by WICI members are *not* related to complex systems, (2) that our complex systems search query is missing some important terms, or (3) that WICI members do not always explicitly use complexity terminology in their work, even when their underlying ontology is grounded in a complex systems perspective.

Note that the number of publications by Extended Core Members that were captured by our query (150 documents) is not substantially different from the number of self-reported publications by the same authors that we identified in Scopus (124 publications, all of which were self-identified as complex systems related). This result suggests that our query picked up some publications by Core or External Core Members that they did not self-identify as being complex systems related. To validate this finding, future work could provide a list of those additional publications to the authors and confirm whether they are, indeed, complex systems research. We could do the same with the full WICI membership to determine whether our query missed publications that the authors consider to be complex systems research.

Table 2: Number of publications (till 2019 and in Scopus) captured by our complex-systems query (sorted by corpus):

Corpus	Captured	Not Captured	Total
Extended Core, self-reported	31	202 (109 not in Scopus)	233 (109 not in Scopus)
Core Members, all publications	102	583	685
Extended Core Members, all publications	150	739	889
All WICI All Publications	356	1806	2162
UW, CS query	3671	n/a	3671
Total	**3400**	**1753**	**5153**

Table 3: Number of publications (till 2019 and in Scopus) captured by our complex-systems query (sorted by member type):

Member Status	Captured	Not Captured	Total
Core	102	583	685
External Core	48	159	207
Member	138	804	942
External Member	73	415	488
Non-member	3671	n/a	3671
Total	**3858**	**1939**	**5797**

The following WICI Member Community analysis, as well as the University of Waterloo Community analysis, provide a more detailed look at the makeup of these corpora and our search query.

3.1.2 Findings: Keyword Analysis

To test our hypotheses regarding the complex systems query "filter" results, we took a closer look at the terms in our search query compared to publications' official Author Keywords. Taking the set of self-reported Extended Core Member publications (2010-2019)—which were author-identified as being complex systems-related—we see that our complex systems query captured only 31 of the 124 publications. If we take the subset that was *not* captured, and extract the Author Keywords (which are chosen by the author and included as bibliographic data in Scopus), we can potentially identify some keywords that are missing from our query. Of course, while the majority of the ~300 Author Keywords in this subset of publications are related to specific fields, topics, or applications, as many as 40 of them are or may be complex systems related.

These terms include: various types of models or modeling (e.g., cognitive modeling, disease-behaviour model, dynamic model, dynamic transmission model, epidemic modeling, graph model, graph model for conflict resolution, hierarchical graph model, mathematical modeling, network model, pair approximation models, stochastic model); keywords related to dynamics (e.g., dynamic analysis, infectious disease dynamics, metapopulation dynamics); keywords related to decisions and scenarios (e.g., decision analysis, Delphi, fuzzy preferences, multi-criteria decision analysis, multiple levels of preference, robust decision making, rough set theory, scenario discovery, scenario diversity analysis, vulnerability based scenario analysis, wicked dilemmas); keywords related to (dis)equilibrium (e.g., initial condition optimization, interval fuzzy equilibrium, interval fuzzy preference, interval fuzzy stability, stochastic optimization); and keywords related to systems or complexity more generally (e.g., complex landscape ecosystem, homophily, human-environment coupling, relational ontology, systems thinking).

Clearly, many of these keywords represent the research of one or two authors of the Extended Core Members population (e.g., Keith Hipel's graph model, or Vanessa Schweizer's scenario work). In fact, many of these terms emerge from the Gargantext and VOSviewer network analysis in the section 3.2. However, some of them capture areas of complex systems research that are not yet well-represented in our search query, and should be included in future bibliometric searches (e.g., dynamic analysis, robust decision making, vulnerability based scenario analysis, human-environment coupling, relational ontology, systems thinking).

Section 4.2 of this report presents additional keyword analysis of the University of Waterloo complex systems search results, including frequency analysis and a comparison to our search query terms.

3.1.3 Findings: SciVal Author Reports

Once we identified each WICI member in Scopus and created the Saved Author lists, we added each individual author *and* each group of authors as "Researcher entities" in SciVal. There, we used SciVal's Reporting feature to generate individual Author Reports, which included metrics related to:

- the author's overall research performance (number of publications in Scopus, h-index value),
- the top 15 keyphrases of the author (based on publications from 2009-2018 in Scopus),
- their publications organized by subject area,
- a list of their most cited publications,
- top five research topics,
- all topics and topic clusters the author has contributed to,

- scholarly output by Scopus source (or journal),
- the author's top collaborating institutions,
- and collaborating authors. [18]

We produced Author Reports for each of the WICI Core and External Core Members:

- Madhur Anand (external)
- Chris Bauch
- Mark Crowley
- Peter Deadman
- Liane Gabora (external)
- Igor Grossmann
- Keith Hipel
- Thomas Homer-Dixon
- Sharon Kirkpatrick
- Chrystopher Nehaniv
- Mary O'Connor (external)
- Dawn Parker
- Stephen Quilley
- Vanessa Schweizer
- Raja Sengupta (external)
- Paul Thagard
- Roger White (external)

Note that all SciVal author analyses are of publications from 2009 to 2018.

Taking Sharon Kirkpatrick's Author Report as an example, we see that (according to the Scopus database) her scholarly output for the years 2009 to 2018 was 67 publications with a field-weighted citation impact of 3.67. Those publications have a total of 3,913 citations and an average of 58.4 citations per publication. Kirkpatrick has an *h*-index of 28 and an *h*5-index of 12.

In terms of subject areas, 34.1% of Kirkpatrick's publications are categorized as "Medicine," 32.5% as "Nursing," etc. Each of these subject areas are broken down further as well. On page 103 of the Supplemental Report, we see the list of her top five most cited publications.

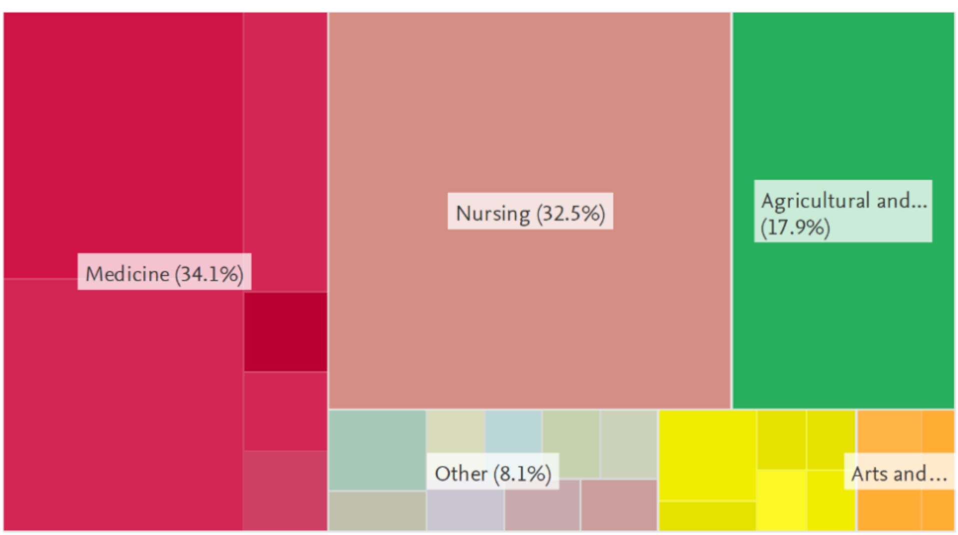


Figure 2: Tree Map: Publications by subject area for Sharon Kirkpatrick (produced in SciVal)

[18] See Appendix D for SciVal's explanation of topics, topic clusters, and keyphrases.

Next, we see Kirkpatrick's Top 5 Research Topics (defined by SciVal), which include: food and food questionnaires; food supply and food nutrition assistance; diet and food; food deserts and residence characteristics related to food; and specifically, carbohydrates and beverages. In the list of Kirkpatrick's Topic Clusters, we see that the vast majority of her publications are in the "obesity, motor activity, child" cluster.

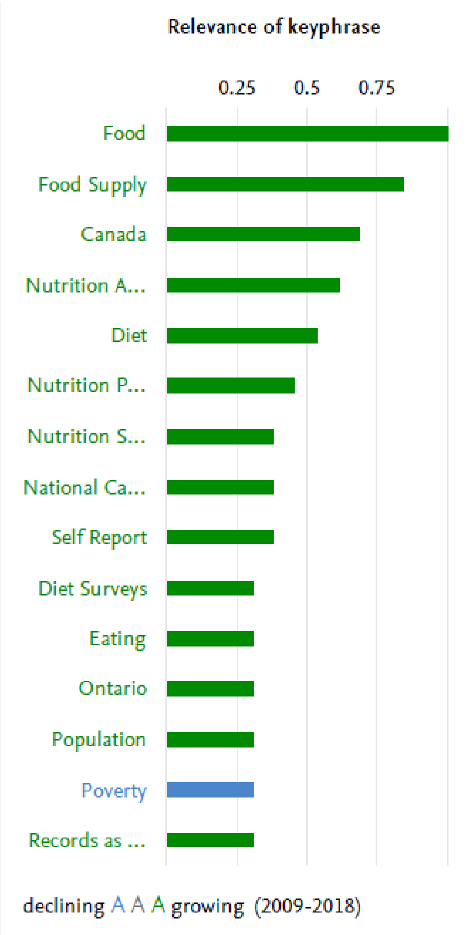
She publishes most frequently in the Journal of the Academy of Nutrition and Dietetics, and she has co-authored most frequently with researchers from the National Institutes of Health (USA), the University of Waterloo, and the University of Toronto. However, she has disproportionate citations of publications co-authored with researchers at the National Institutes of Health, Information Management Services, Inc., and the United States Department of Agriculture.

Table 4: Top Collaborating Institutions for Sharon Kirkpatrick (produced in SciVal)

	Institution	Co-authored publications	Citations
1	National Institutes of Health	24	2,714
2	University of Waterloo	15	193
3	University of Toronto	13	668
4	Information Management Services, Inc.	7	1,219
5	University of Calgary	7	351
6	United States Department of Agriculture	5	1,727
7	Alberta Health Services	4	39
8	Cancer Care Ontario	4	39
9	University of Newcastle	4	39
10	The Gertner Institute	3	323

Beginning on page 112 of the Supplemental Report, we see an extensive list of Kirkpatrick's co-authors (ranked by number of co-authored publications), their institutional affiliations, and other metrics.

Finally, we see Kirkpatrick's top 15 keyphrases, with those in green growing in frequency of use and those in blue declining (for the period 2009-2018) (see Figure 3).



Overall, these Author Reports give us a range of insights regarding the research areas, topics, output, and collaboration of individual scholars. These insights can be valuable to the scholars themselves, their institutions, or a group of scholars looking for new opportunities, new collaborators, and ways to leverage their collective expertise. Of course, big data and quantitative assessments like these have their limitations, and more qualitative methods are needed to add nuance to these findings. In future research, it would be valuable to present these results to the individual scholars themselves—especially around topics and keyphrases—to incorporate their own views about what is important in and about their work.

Lastly, while the Author Reports we produced in SciVal include *all* the author's publications between 2009 and 2018, future research could take a self-identified complex systems-related subset of an author's publications and define and analyze them as a corpus. While this would require more manual intervention, it would enable a more direct look at a "guaranteed" list of complexity publications; and would help us better understand how databases like Scopus categorize and label that work.

Figure 3: Top 15 Keyphrases of Sharon Kirkpatrick (produced in SciVal)

3.1.4 Findings: SciVal Corpus Reports

The bulk of our analysis is in our **Corpus or Publication Set Reports,** which we produced for each sub-population within the WICI community (see Appendix B for full list; see full reports starting on page 291 of the Supplemental Report). These include the following analyses:

- overall research performance analytics of the set,
- a summary of outputs in Top 10% Citation Percentiles,
- publications in Top 10% Journal Percentiles by CiteScore Percentile,
- international collaboration summary and academic-corporate collaboration summary,
- publications by Subject Area,
- most published-in topics (and Topic Clusters),
- keyphrase analysis[19],
- scholarly Output by Scopus Source (i.e., top sources),
- a detailed collaboration list,
- a list of all institutions where authors/co-authors are based, and

[19] See Appendix D for SciVal's explanation of keyphrases.

- the top authors in Canada (by number of citations, publications, and/or h-index).

Taking the “All WICI Members’ Publications 2009-2018, filtered by CS query” corpus as an example, we can see that this subset of WICI members’ publications captured by the complex systems search query amounts to 357 publications, written by a total of 816 authors (including all co-authors not affiliated with WICI). The total citation count for this corpus is 9,132 and the average citations per publication is 25.6. Twenty-one per cent of these publications are in the top 10% most cited worldwide, and 48% of them are published in the top 10% of journals (by CiteScore percentile). In terms of collaboration, over 52% of publications are co-authored with institutions outside of Canada.

Perhaps unsurprisingly, the subject area that accounts for the most publications in this set is computer science (with approximately 130 publications), followed by environmental science and mathematics (each with approximately 80 publications), and then the social sciences, agricultural and biological sciences, engineering, biochemistry genetics, physics and astronomy, earth and planetary sciences. With less than 20 publications each, the second half of the list includes medicine, “multidisciplinary,” psychology, neuroscience, decision sciences, materials science, arts and humanities, chemistry, immunology and microbiology, business and management, economics and econometrics, energy, chemical engineering, and nursing (see Figure 4below).

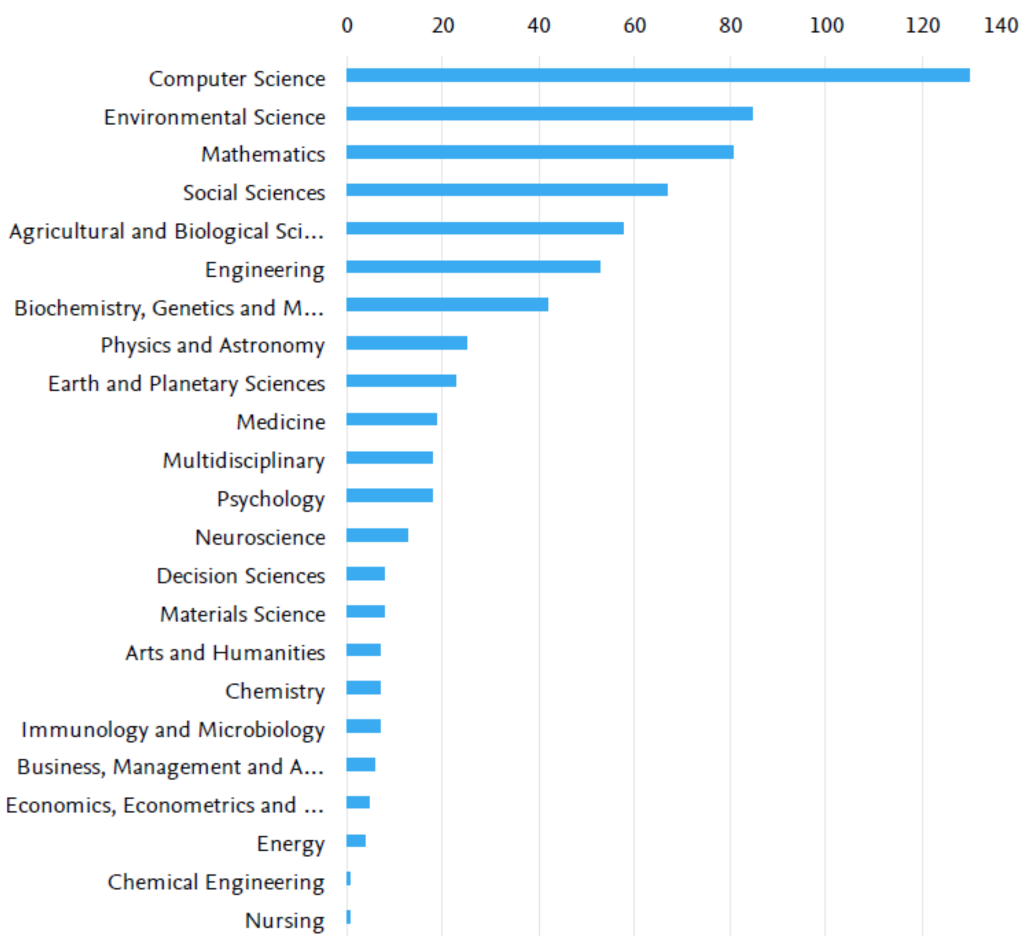


Figure 4: Publications by Subject Area, All WICI All Publications, filtered by CS query (produced in SciVal)

If we compare this subject area breakdown to the one for the non-filtered corpus (see Figure 5 below)—all publications 2009-2018 by all WICI members—we see that while computer science maintains first place, both environmental science and social science account for a somewhat larger proportion of the complex systems subset of publications than they do in the larger corpus. Medicine also ranks lower in the complex systems subset than in the full corpus. Psychology ranks somewhat lower in the full corpus than in the subset, suggesting that the field is not as well represented in the complexity work by WICI membership. Finally, multidisciplinary research ranks a full six positions higher in the subset, a key indication that complex systems related research is more often multidisciplinary than non-complex systems work.

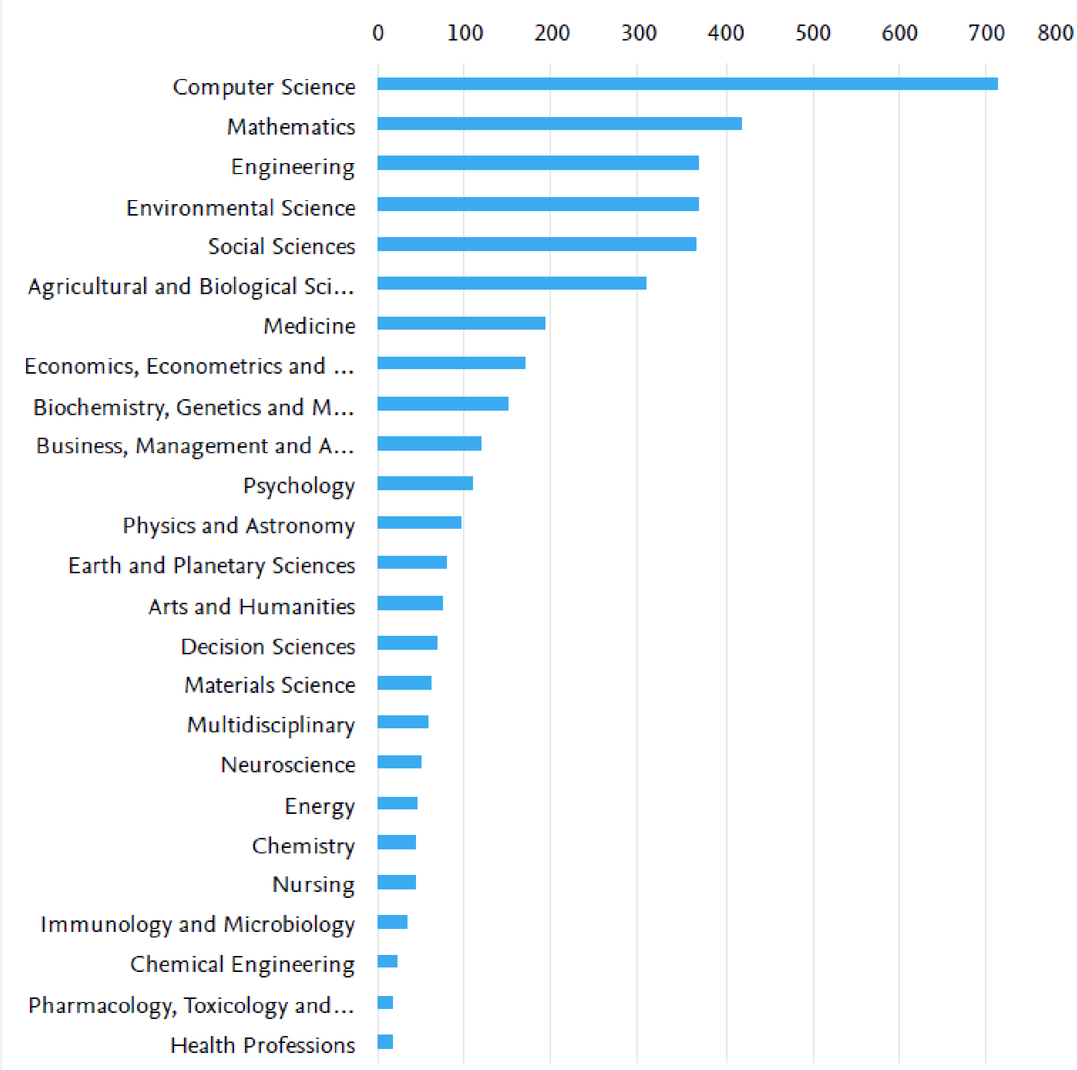


Figure 5: Publications by Subject Area, All WICI All Publications (produced in SciVal)

In the keyphrase analysis for this corpus—All WICI Members' Publications from 2009 to 2018, filtered by our complex systems query (see page 558 of the Supplemental Report)—we can see that many, but not nearly all, of the terms in our complex systems search query are represented. Some of the terms that emerged from our manual keyword analysis (see section 3.1.2) of WICI Extended Core Members' self-reported publications (that were identified in Scopus but *not* captured by our complex systems query) are found in the figure below as well (e.g., dynamics, models, decision making, and vulnerability), strengthening the case for including them in future complex systems related research queries.

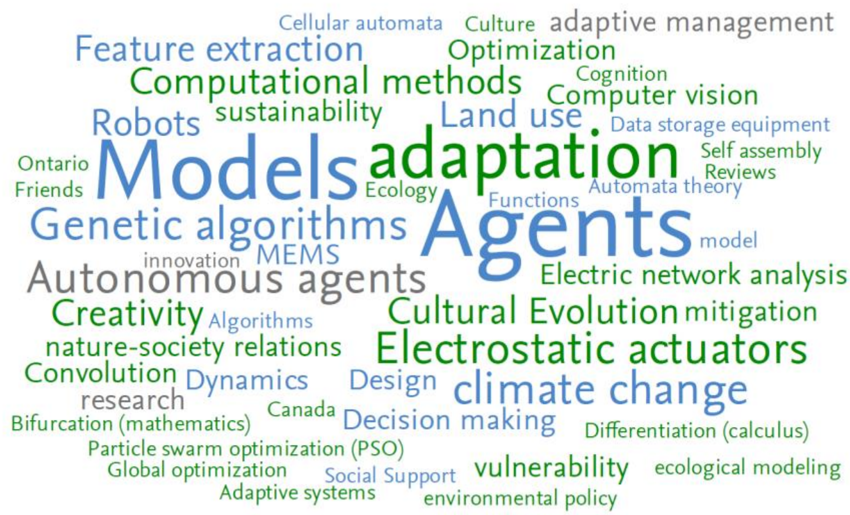


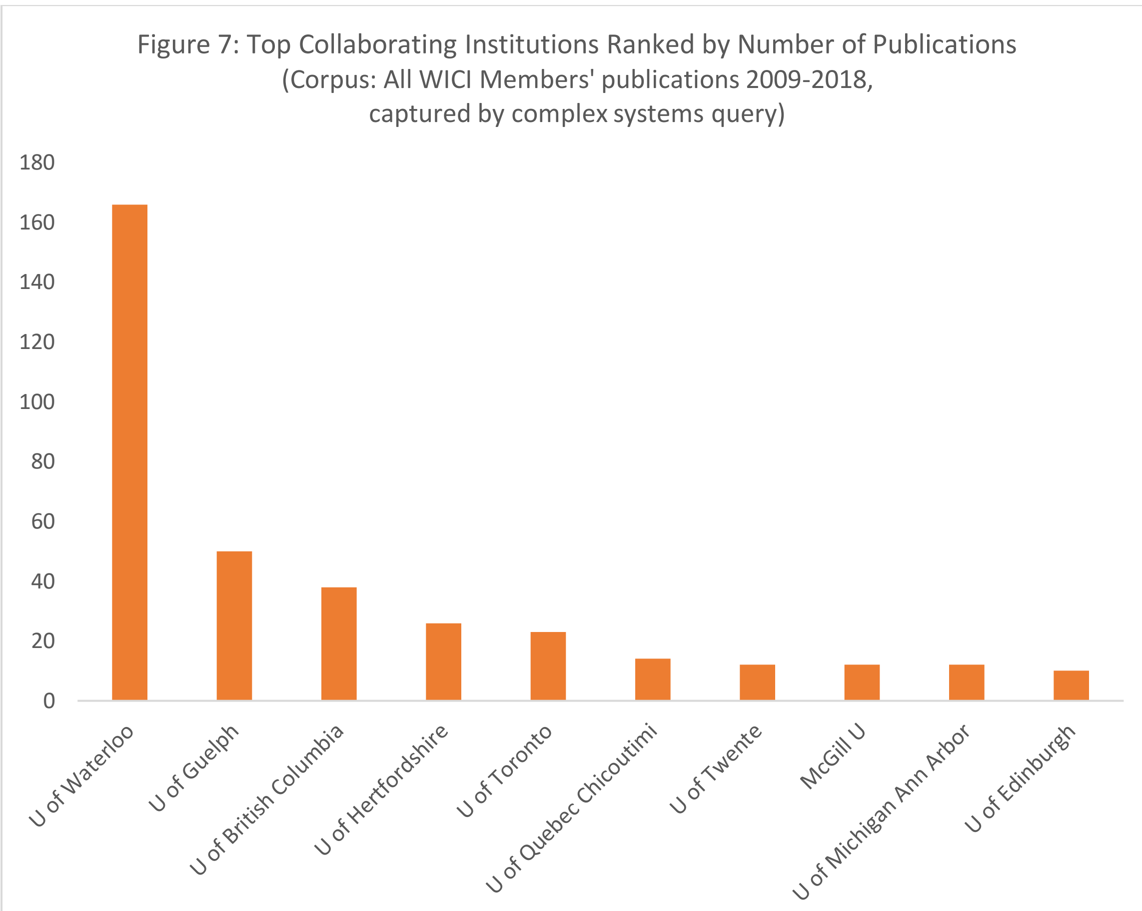
Figure 6: Keyphrase analysis (produced in SciVal)

Next, we see a list of journals (Scopus Sources) where these publications were published, ranked from most to least. For example, 12 publications in this corpus were published in *Lecture Notes in Computer Science,* 11 publications were published in *Environmental Modelling and Software,* and so on. As this list of the ten most represented journals accounts for only 61 of the 357 publications in the corpus, we can conclude that most of the publications are spread across dozens more journals. In other words, WICI members publish very broadly. However, another method—perhaps simply a comparison with the larger UW or Canada complex systems datasets—is needed to get an accurate sense of which journals publish more complex systems research.

Table 5: Sources ranked by scholarly output (produced in SciVal)

Scopus Source	Scholarly Output ↓	Citations	Authors	CiteScore 2018
Lecture Notes in Computer Science	12 ▼	58	27 ▼	1.06
Environmental Modelling and Software	11 ▼	484	59 ▼	5.08
Journal of Theoretical Biology	7 ▲	29	11 ▲	2.04
PLoS ONE	7 ▲	97	34 ▲	3.02
Climatic Change	5 ▲	120	35 ▲	4.25
Proceedings - International Conference on Image Processing, ICIP	5 ▲	20	13 ▲	0.00
Ecology and Society	4 ▲	71	16 ▲	4.81
Proceedings of the ASME Design Engineering Technical Conference	4 ▼	9	7 ▼	-
Journal of Computational Physics	3 ▲	53	6 ▲	3.31
Physics of Life Reviews	3 ▲	222	6 ▲	1.29

SciVal also enables us to identify the home institutions where co-authors in this corpus are based, and may provide leads on potential future nodes for the Canadian Network of Complex Systems, as well as international partners. As a reminder, this corpus includes All WICI Members' publications from 2009-2018 that were captured by our complex systems query. The institutions (including University of Waterloo) that are most frequently represented in the corpus authorship (see Figure 5 below) are: University of Waterloo, University of Guelph, University of British Columbia, University of Hertfordshire, University of Toronto, Universite du Quebec a Chicoutimi, University of Twente, McGill University, University of Michigan Ann Arbor, and University of Edinburgh. This figure only shows the top ten institutions, but we can generate larger lists (e.g., the top 50) in SciVal quite easily in the future.



In the last section of the Corpus Report, the list of Top Collaborating Canadian Authors in our Supplemental Report (see page 563 to 598) may provide leads on potential new WICI members (at UW or elsewhere in Canada). This list shows the most well represented co-authors in the corpus (All WICI Members' publications from 2009-2018 captured by our complex systems query) ranked by the number of publications they have co-authored in the corpus. We can also see each author's affiliation, their views count for these publications in Scopus, and their citation count for these publications. This list (and similar ones in the Supplemental Report) would be a good place to start to assess interest in being a part of or at least connected to a more formal Canadian Network of Complex Systems Scholars. In future research, we could also expand the list of collaborating authors to include those based internationally.

Overall, these Corpus Reports (as well as the Author Reports) are useful for assessment, benchmarking, and getting a clearer picture of scholarly output and research activities of WICI members. But they also offer valuable direction for developing a Canadian Network of Complex Systems, and for tapping into existing domestic and international complexity communities, as shown in the lists of collaborating institutions and co-authors. Finally, the reports give us a sense of the main topics and topic clusters that WICI members contribute to, which can help us identify research trends, gaps, and opportunities within the community.

3.2 Network Analysis

In the next part of the research project, we took some our datasets from the SciVal analysis stage and imported them into two network analysis software programs, Gargantext and VOSviewer. While our SciVal Reports are highly detailed and present many different bibliometrics of individual authors and publication sets, network analysis allows us to study the relationships between authors and publications. Understanding these relationships yields different insights, especially around identifying "communities" of authors, journals, or keywords/themes.

In this section, we first discuss the Gargantext data and methods, present our network visualization results, and highlight findings. We then do the same for our VOSviewer analysis.

3.2.1 Gargantext Data & Methods

As a reminder, Gargantext is a software platform that combines natural language processing, digital text mining, network analysis, and graph visualization. Users can analyze and interact with text at the level of individual documents, the level of the database of mined text, and the level of the graph itself, which makes it unique compared to other text mining or network analysis software. (See Appendix C for a screenshot tour and explanation of Gargantex's features.)

For our analysis, we used the corpus of All WICI Member's publications from 2009 to 2018 (a total of 2168 publications—i.e., not just the subset captured by our complex systems search query).[20] We exported the corpus, formatted it for Gargantext, and uploaded it to our private researcher account. Unfortunately, Gargantext did not recognize published conference proceedings; but the counts for journals were the same as SciVal.

Initially, the Gargantext text-mining software automatically extracted 349 terms (which might be single words or multi-word phrases) from the corpus—i.e., from publication titles, abstracts, and official keywords. However, many of these were generic to academic research and writing (e.g., "introduction," "first," "second," "the author argues," etc.). We refined the list of terms by manually deleting those generic words and ended up with a total of 244 terms (we also exported this list as a CSV file for our records). Gargantext refers to this final list of terms as the "map list."

[20] We decided to use the same year range as the SciVal reports for easier and more accurate comparison, despite having 2019 data available.

Once the map list was finalized, we generated a map (also referred to as a network or graph) of the relationships between all the terms in the map list. Gargantext computes the maps based on a semantic proximity measure called "conditional distance," which is a formula that "reflects the interaction between two terms in the corpus [and makes] it possible to identify the communities of use of the terms, or to synthesize the interactions that have been demonstrated in the literature between several entities (for example drugs)."[21] Each term is visualized as a node in the map (a non-directed graph), and the links between the nodes indicate "which terms co-occur in the same meaning unit and with what probability."[22] Toggling the map's "proximity measure" button allowed us to maximize modularity and more clearly view communities and clusters. Gargantext has limited export and/or saving features, so we took a snapshot of the full map, and then zoomed in to each cluster and took additional snapshots (each of which are included in section 3.2.2). We also exported the map as a GEXF file (a standard network file type) for future network analysis.

3.2.2 Findings: Gargantext Maps

With Gargantext's clustering algorithm, we identified eight thematic clusters based on the co-occurrence of terms analysis described above. These clusters consisted of terms related to: (1) the graph model of conflict resolution, (2) tree research, (3) disease dynamics and vaccination, (4) ecosystem services and management, (5) children and health interventions, (6) diet, food, and assessment design, (7) Canada and development, (8) and climate change and adaptation. The clusters do not necessarily align perfectly with one author's or a group of authors' work because the terms are analyzed across the entire dataset of publications. However, because some authors have more publications in the dataset or their research is highly concentrated in one topic area, several of the map clusters represent them more heavily. To aid in viewing, we took a zoomed-in screenshot of each cluster, each of which are included below with a full list of terms in the caption.[23]

[21] "About semantic proximity measures used in Gargantext," translated from French by Google Translate, https://iscpif.fr/gargantext/mesures-utilisees-dans-gargantext/.

[22] Rosa Vicari et al., Supplement of "Climate risks, digital media, and big data: following communication trails to investigate urban communities' resilience," 2019, p. 1486.

[23] The GEFX file is also available in the project database.

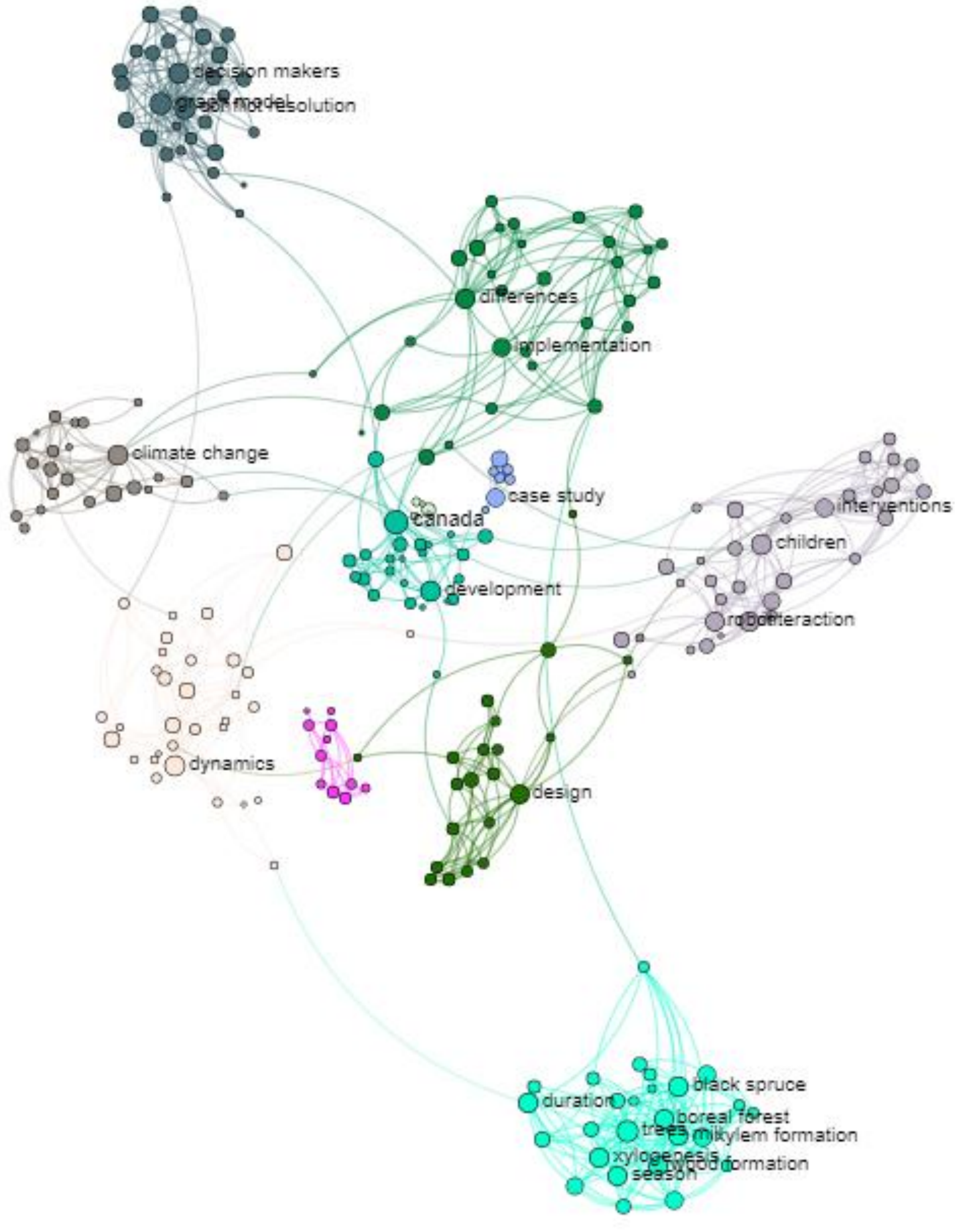


__Figure 8:__ Co-occurrence Map of Terms using WICI Member Publications, 2009-2018 (entire map)

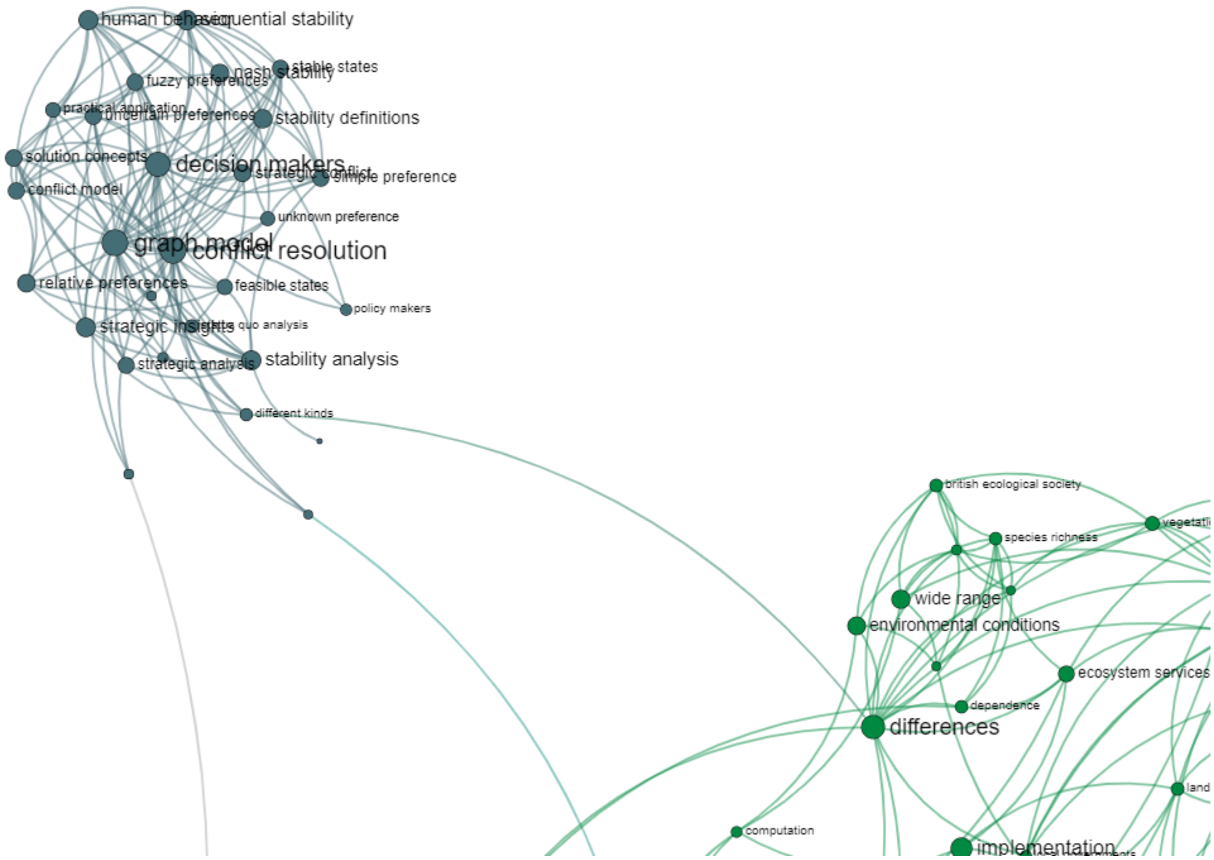


Figure 9: *Co-occurrence Map of Terms using WICI Member Publications, 2009-2018 - Cluster 1. Terms included in cluster, listed somewhat according to node size: conflict resolution, decision makers, graph model, stability analysis, strategic insights, fuzzy preferences, strategic conflict, conflict model, sequential stability, stability definitions, nash stability, simple preference, feasible states, uncertain preferences, relative preferences, solution concepts, stable states, middle east, human behavior, policy makers.*

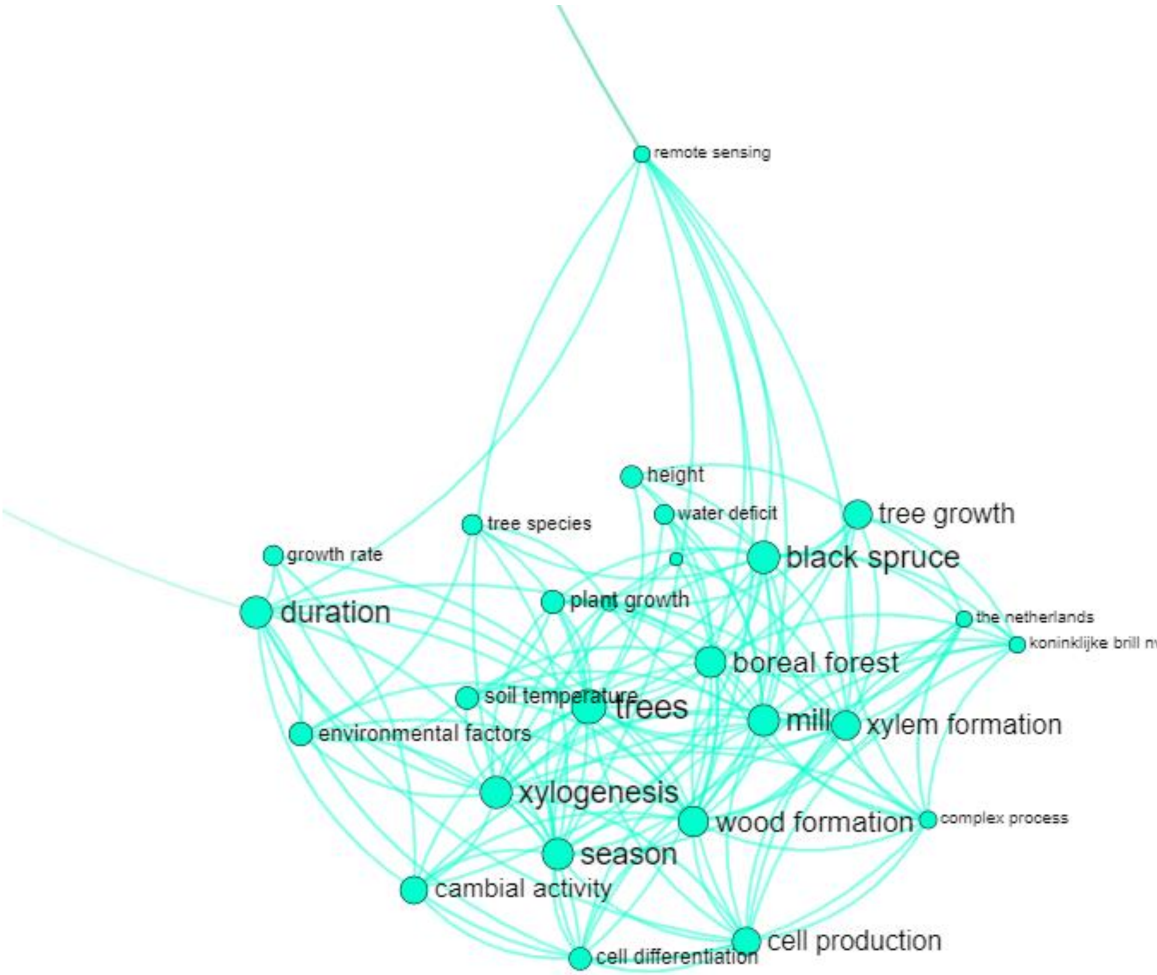


Figure 10: *Co-occurrence Map of Terms using WICI Member Publications, 2009-2018 - Cluster 2. Terms included in cluster, listed somewhat in order of node size: trees, black spruce, boreal forest, xylogenesis, season, duration, xylem formation, cell production, wood formation, cambial activity, mill, tree growth, plant growth, soil temperature, tree species, height, cell differentiation, water deficit, remote sensing, the Netherlands, growth rate, environmental factors.*

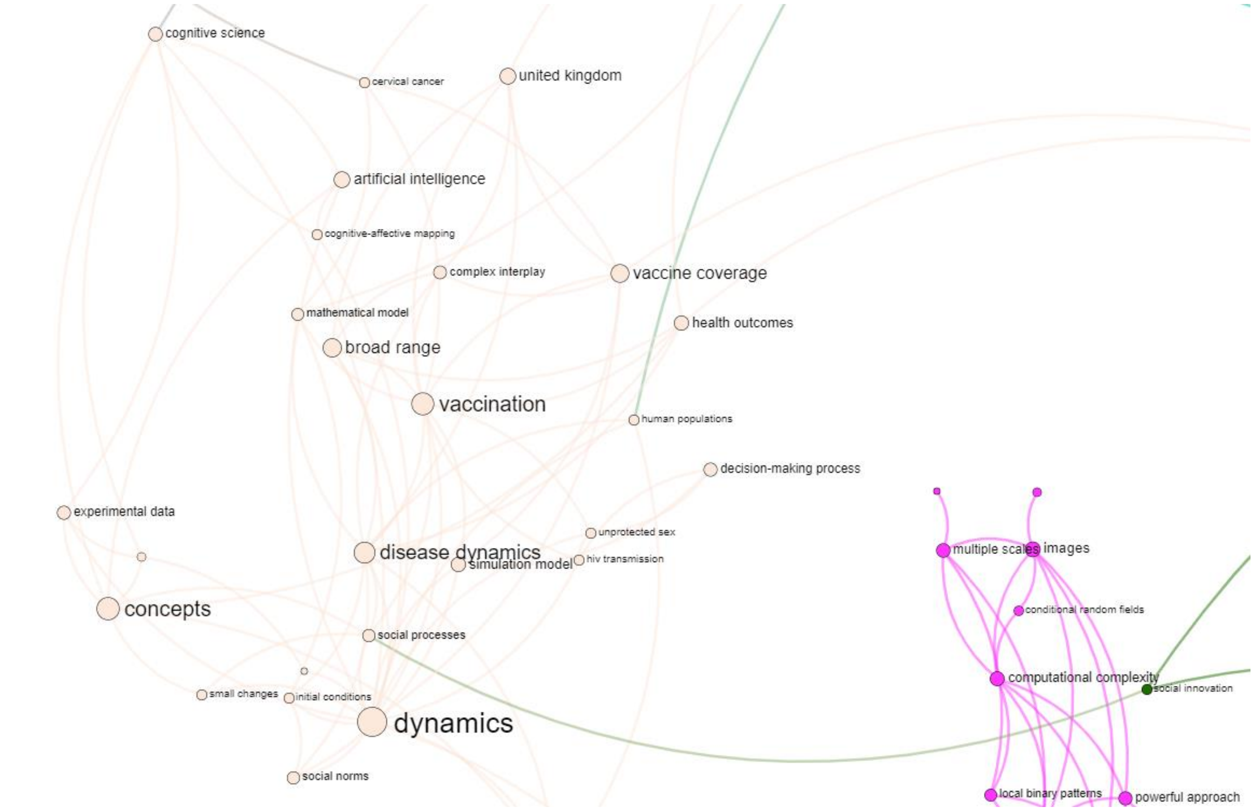


Figure 11: *Co-occurrence Map of Terms using WICI Member Publications, 2009-2018 - Cluster 3. Terms included in cluster, listed somewhat in order of node size: dynamics, disease dynamics, vaccination, concepts, vaccine coverage, social processes, health outcomes, social norms, simulation model, united kingdom, complex interplay, cultural differences, human populations, initial conditions, cognitive science, artificial intelligence, cognitive-affective mapping, unprotected sex, hiv transmission, cervical cancer, decision-making processes, mathematical model, small changes, experimental data.*

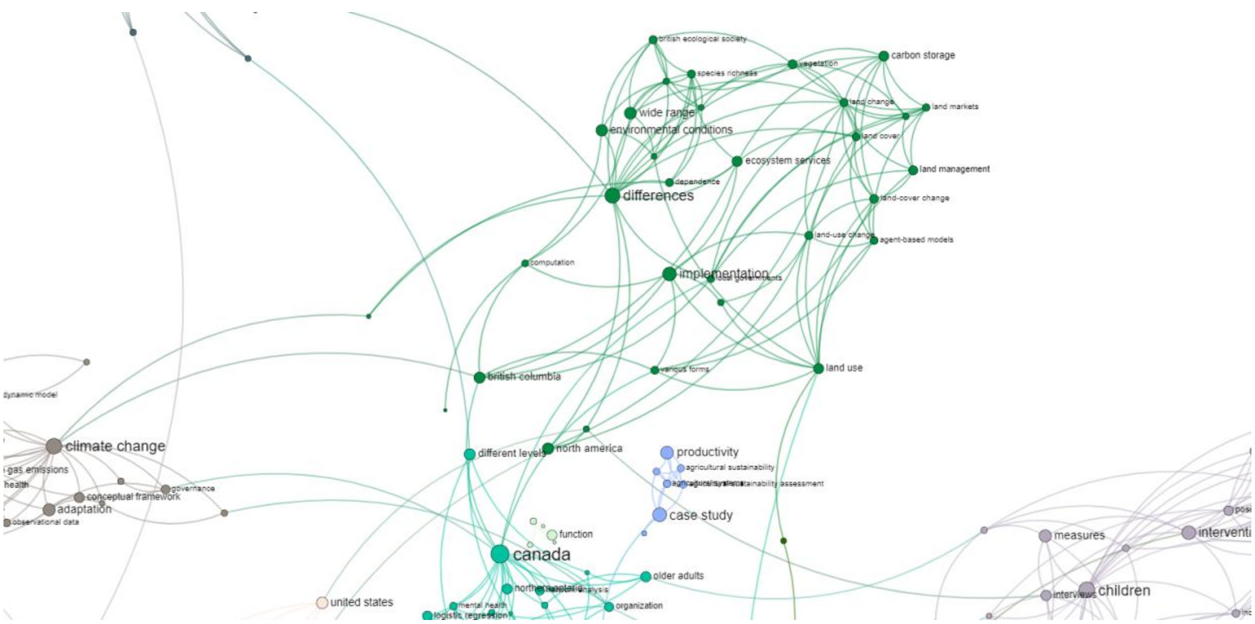


Figure 12: *Co-occurrence Map of Terms using WICI Member Publications, 2009-2018 - Cluster 4. Terms included in cluster, listed somewhat in order of node size: differences, implementation, environmental conditions, ecosystem services, wide range, British*

Columbia, north America, land use, carbon storage, vegetation, land change, land cover, land markets, land management, land-cover change, land-use change, agent-based models, local governments, computation, dependence, species richness.

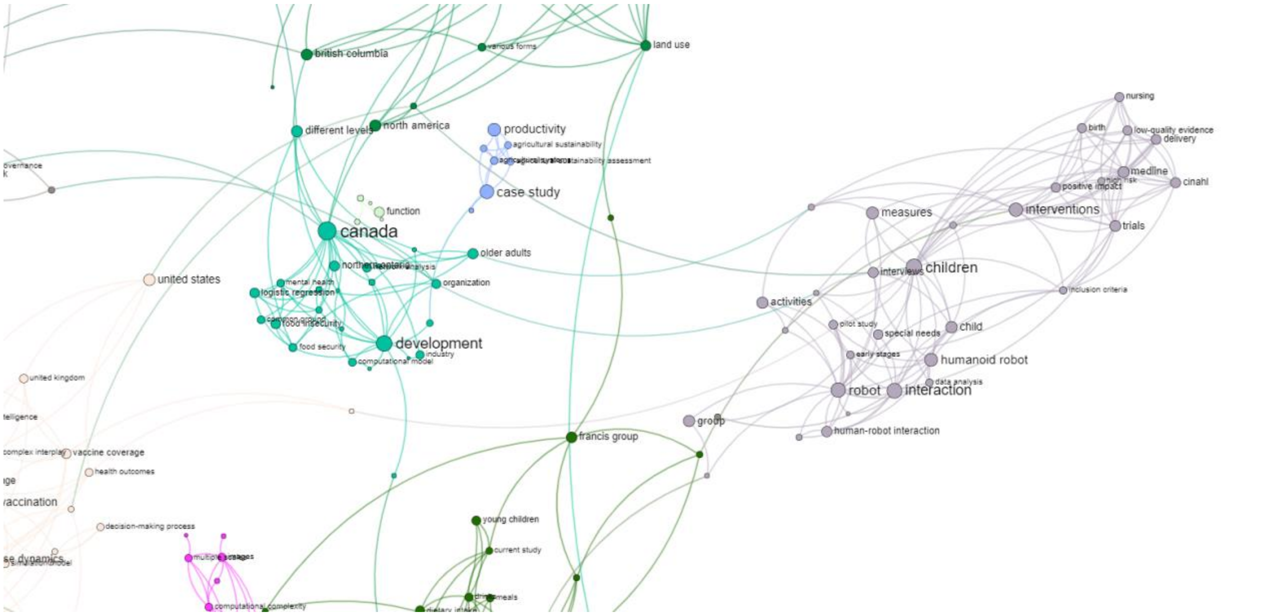


Figure 13: *Co-occurrence Map of Terms using of WICI Member Publications, 2009-2018 - Cluster 5. Terms included in cluster, listed somewhat in order of node size: children, child, interaction, interventions, measures, interviews, humanoid robot, robot, human-robot interaction, activities, group, pilot study, early stages, special needs, data analysis, inclusion criteria, trials, positive impact, high risk, medline, birth, nursing, low-quality evidence, delivery.*

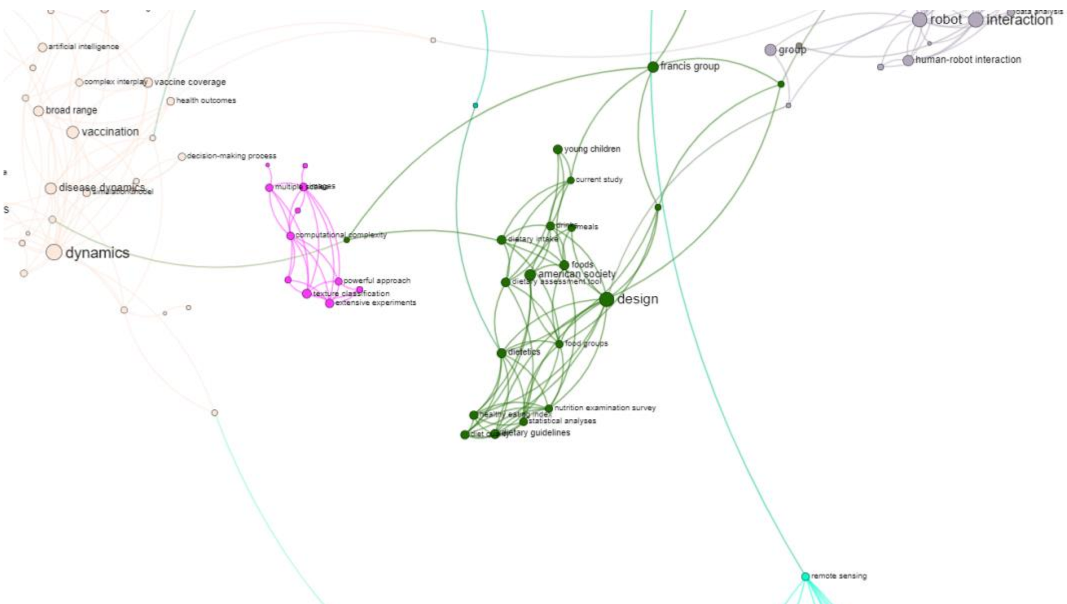


Figure 14: *Co-occurrence Map of Terms using WICI Member Publications, 2009-2018 - Cluster 6. Terms included in cluster, listed somewhat in order of node size: design, american society, foods, dietetics, nutrition examination survey, statistical analyses, healthy eating index, diet quality, dietary guidelines, food groups, dietary assessment tool, dietary intake, drinks, meals, current study, young children, francis group.*

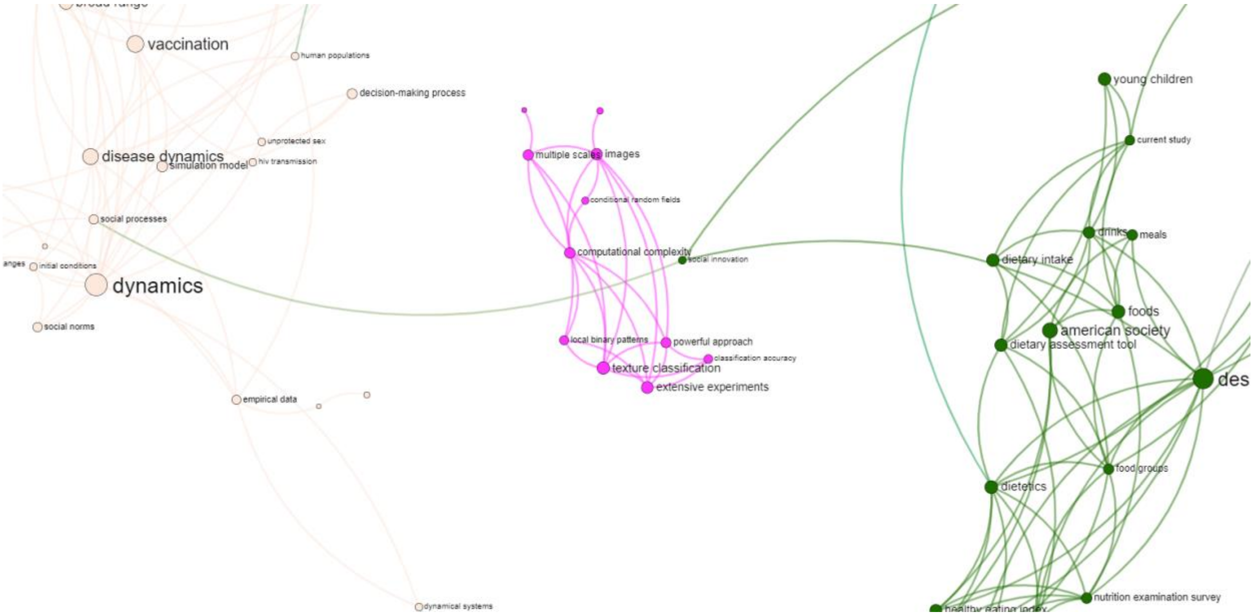


Figure 15: *Co-occurrence Map of Terms using WICI Member Publications, 2009-2018 - Cluster 7. Terms included in cluster, listed somewhat in order of node size: texture classification, extensive experiments, computational complexity, multiple scales, images, conditional random fields, local binary patterns, powerful approach, classification accuracy.*

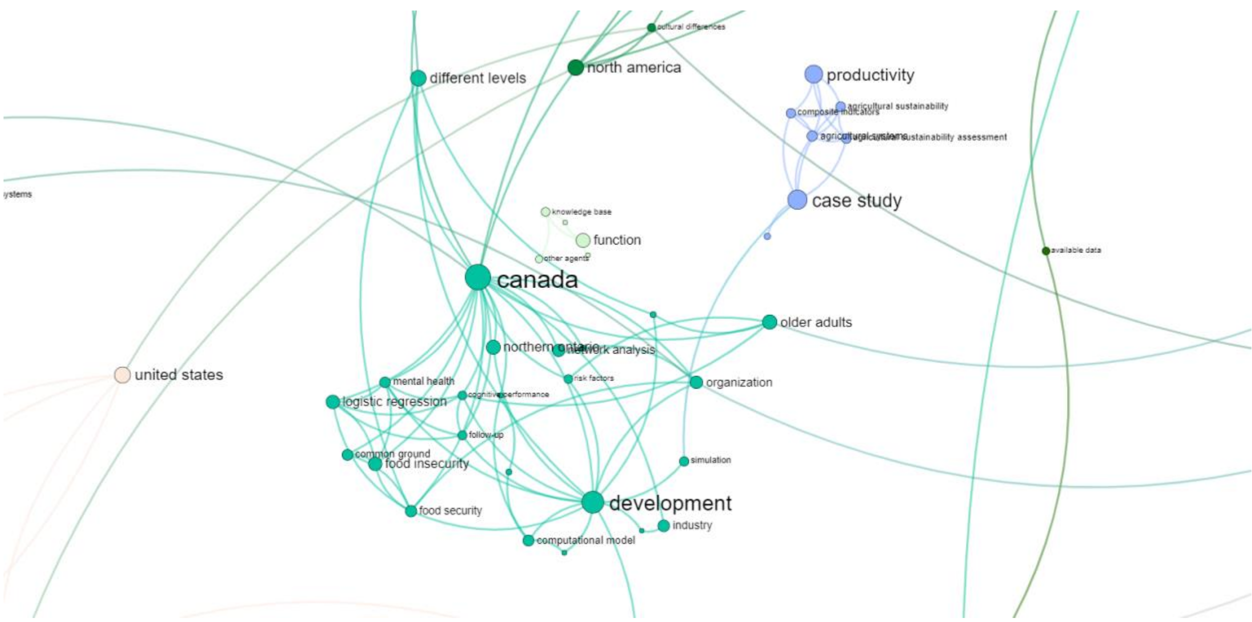


Figure 16: *Co-occurrence Map of Terms using WICI Member Publications, 2009-2018 - Cluster 8. Terms included in cluster, listed somewhat in order of node size: Canada, development, different levels, northern Ontario, older adults, food insecurity, food security, logistic regression, common ground, mental health, cognitive performance, organization, network analysis, risk factors, computational model, simulation, social innovation, follow-up, alcohol use, wise reasoning, social network analysis, new ideas, neural network.*

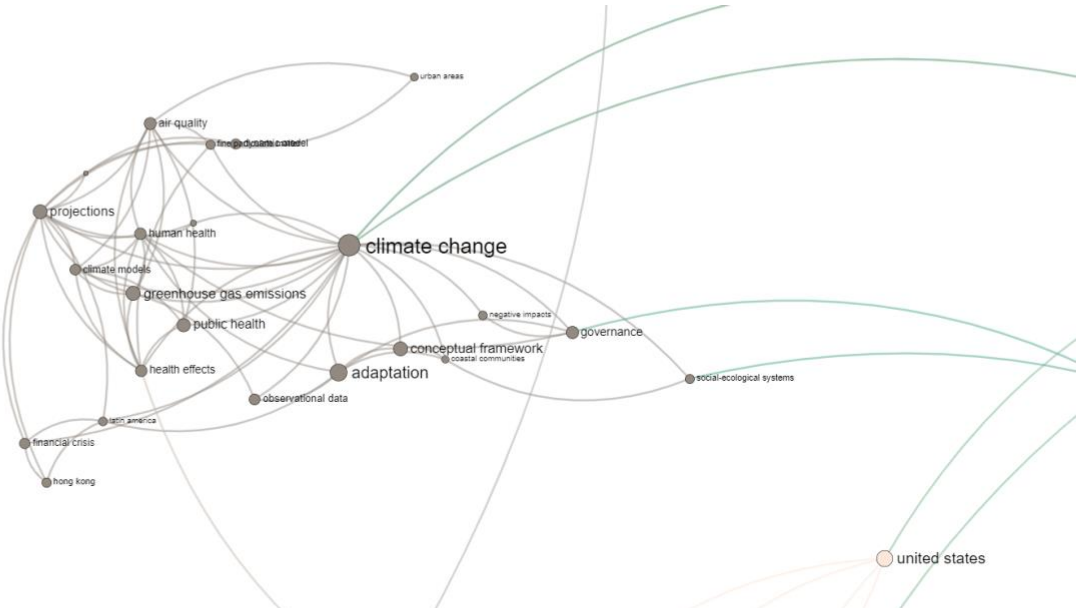


Figure 17: *Co-occurrence Map of Terms using WICI Member Publications, 2009-2018 - Cluster 9. Terms included in cluster, listed somewhat in order of node size: climate change, adaptation, conceptual framework, governance, greenhouse gas emissions, public health, climate models, human health, projections, health effects, air quality, urban areas, financial crisis, hong kong, latin America, observational data, coastal communities, negative impacts, social-ecological systems.*

After familiarizing ourselves with WICI member's research areas and producing the SciVal Author Reports, it is clear from looking at the sets of terms that some of these thematic clusters represent the work of just a few scholars (e.g., cluster 1 obviously captures Keith Hipel and colleague's work on the graph model of conflict resolution; cluster 2 represents Madhur Anand, Chris Bauch, and Dawn Parker's research related to trees). Others, however, cover a broader set of work despite still clustering around key terms (e.g., cluster number nine on climate change is clearly quite multi-disciplinary, including terms related to environmental issues, social concerns, data and methods, and different geographies). In these cases, there may be several authors represented in a single cluster who are not actually aware of each other's work or who do not collaborate yet.

We find that the more author-homogenous clusters also tend to have fewer outward connections to other clusters. Cluster 1 links out only three times, and cluster 2 only twice, whereas cluster 3 has seven links to six other clusters, and cluster 4 has ten links to five other clusters. This makes sense, because a greater diversity of authors means an increased likelihood that one or more author's research makes its way into more than one cluster. This observation suggests that either (1) the terms in those more homogenous clusters are overrepresented because of a high publication count of one or two authors, or that (2) the authors in those clusters do not collaborate with many others outside their main research topics. The first possibility points to a need to control for publication counts in future research, especially when using a dataset that spans multiple disciplines (with potentially differing publishing rates). The second possibility points to potential untapped collaboration opportunities that would strengthen the overall coherence of complex systems research.

Here, a limitation of Gargantext prevents us from digging further into the full publication data underlying the clusters. While the data obviously exists in the platform, it is only possible for users to see a list of the five "most related" publications represented by a single node or groups of nodes (i.e., a cluster). If we could access the entire list of publications related to each node, we could generate an author list for each cluster, as well as note which authors are represented in multiple clusters. Future research could involve finding a way around this software limitation using several tools.

In addition to insights about the clusters themselves, some of the links *between* clusters also point to possible bridges from one body of work to another. For example, the term "social innovation" is the bridge between "social processes" in cluster 3 and "dietary intake" in cluster 6. Clusters 5 and 6 are linked via "interventions" and "physical activity." Clusters 1 and 3 are linked via "middle east" and "cognitive science."

Of course, some bridging terms are more generic, like "differences" and "different kinds", are geographical (names of countries or provinces), or specify the population under study (e.g., "older adults" or "children). Most of these terms sit on the periphery of a cluster, demonstrating what network science has—for decades—called "the strength of weak ties." As Granovetter's original piece argues, "Weak ties are more likely to link members of *different* small groups than are strong ones, which tend to be concentrated within particular groups."[24] These weak ties, or bridges, could point to potentially novel areas of collaboration, and even offer empirical support to multidisciplinary funding applications.

3.2.3 VOSviewer Data & Methods

While Gargantext has deep and highly granular analytical capability, VOSviewer enables more *types* of analysis. In Gargantext, the units of analysis (i.e., the network nodes) are terms from the map list, while in VOSviewer, the nodes can be publications, journals, authors, or keywords. Furthermore, users can perform multiple kinds of analysis on those units/nodes, including co-occurrence relations (as in Gargantext), co-authorship relations, direct citation relations, co-citation relations, and bibliographic coupling relations. Each of these analyses provide different ways to examine all the relationships that exist in a set of publications.

For our VOSviewer analysis, we used the same ALL WICI Publications 2009-2018 corpus that we used in Gargantext. We conducted the following analyses:

- a co-occurrence analysis of terms (where terms are automatically extracted from the titles and abstracts of publications in the corpus, and are linked and positioned based on whether they occur together more or less frequently);
- a co-occurrence analysis of author keywords and of all keywords (similar to the first analysis, but using official keywords instead of automatically extracted terms);
- a co-authorship network analysis (where authors are linked based on the number of publications they co-author together);
- a bibliographic coupling of sources, weighted by citations (where journals are linked based on the number of references they share); and
- a co-citation of sources analysis (where journals are linked based on how frequently they are cited together).

---

[24] Mark Granovetter, 1973, "The Strength of Weak Ties," *American Journal of Sociology,* Volume 78, Issue 6, p. 1376.

3.2.4 Findings: VOSviewer Maps

We created our first map—a co-occurrence analysis of terms (see Figure 16)—using the following steps and parameters: after downloading and installing VOSviewer, (1) create a map based on text data; (2) choose data source, from reference manager file; (3) select RIS file; (4) select Title and Abstract fields; (5) wait while it extracts terms; (6) choose binary of full counting (I chose binary); (7) choose minimum number of occurrences. I stayed with the default of 10, and in this corpus, 868 terms meet this threshold; (8) choose what percentage of the most relevant terms to display. I stayed with the default of 60 per cent, and in this corpus, 521 terms meet this threshold; (9) manually verify terms (sort by relevance, occurrences, or alphabetically); and (10) click finish.

To interpret this map, first note that the horizontal and vertical axes are meaningless; it can be flipped or rotated without changing the meaning of the results. The colours in the map indicate different clusters of terms that are closely related (or often co-occur), and the larger the term node, the higher the number of publications that use it.[25]

In this map of terms, we identified six thematic clusters. Three of the clusters were similar to the Gargantext results: (1) one related to the graph model and conflict resolution, (2) another related to health more generally, and (3) yet another related to environment more generally. The other three clusters were related to (4) international trade, (5) engineering problems and solutions, and (6) behaviour (of systems, people, study participants, governments, and robots).

If we compare the overall structure of this VOSviewer map to the Gargantext map of terms in section 3.2.2, we first notice a difference in modularity (the platforms each use a different algorithm). Because of the density of the network, it is difficult to see *how* the clusters are connected together in the VOSviewer visualization. In other words, the "strength of weak ties" is less obvious because of VOSviewer's clustering algorithm (this might be remedied by raising some of the thresholds for inclusion). However, we can still see that the purple cluster (1) on the far left related to the graph model of conflict resolution, and a portion of the green cluster on the far right related to tree research (3), are less related to the whole network – the same thing we observe in the Gargantext map. This result suggests a relative lack of overlap in terms between cluster 1 and cluster 3. However, we know from our WICI Core Member analysis that some of the conflict resolution research is related to water and other environmental disputes; so the distance between clusters in the map does not necessarily indicate a lack of collaboration between subject areas.

If we cross-reference the six identified clusters in the network below with the SciVal analysis of subject areas for this dataset, we find a significant, but not complete alignment of clusters with SciVal-defined subject areas. The chart on page 487 of the Supplemental Report shows the top subject areas in this order: computer science, mathematics, engineering, environmental science, social sciences, agricultural and biological sciences, medicine, and economics. Environmental science, social science, and medicine/health are more strongly represented in the network analysis in both software platforms than in the SciVal subject area analysis. This suggests, as UW's White Paper on Bibliometrics points out,[26] that

[25] Nees Jan van Eck and Ludo Waltman, 2015, "VOSviewer and CitNetExplorer Tutorial," presented at the 15th International Conference on Scientometrics & Informetrics, Istanbul, Turkey, June 29, 2015. Available at: https://www.slideshare.net/NeesJanvanEck/issi2015-tutorial-vosviewerandcitnetexplorer.

[26] University of Waterloo Working Group on Bibliometrics, 2016, "White Paper on Bibliometrics, Measuring Research Outputs through Bibliometrics," Waterloo, Ontario: University of Waterloo.

citation-tracking databases (e.g., Scopus) and their analytics counterparts (e.g., SciVal) are not always accurate assessors of the impact, prevalence, or epistemic boundaries of a research field.

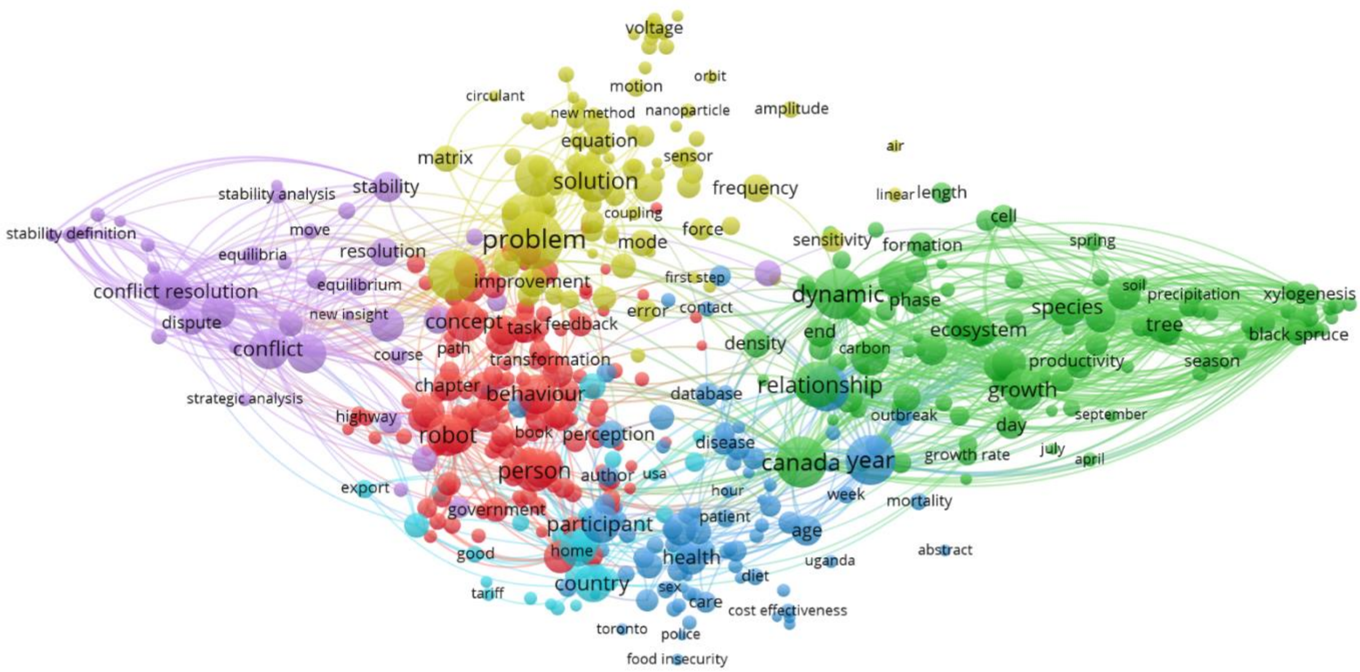


Figure 18: Co-occurrence Analysis of Terms

VOSviewer

In addition to thematic clusters, these bibliometric analyses reveal the relatedness of members' research, possible future author collaborations, and potential publication targets for complex systems research, among other insights. They also validate many of the results from the SciVal Author Reports and Publication Set Reports. Future versions of this report should conduct network analysis on multiple publication sets in an attempt to narrow and/or nuance some of these findings.

The following two networks also analyze the co-occurrence of terms but use author keywords and "all" keywords (i.e., all keywords that Scopus has indexed) instead of terms that VOSviewer automatically extracts (i.e., Figure 16). Our parameters and steps for the author keyword analysis (Figure 17) were similar to the first VOSviewer map, but because the overall number of keywords was significantly lower than automatically extracted terms (4831 keywords)—and because we still wanted to be able to clearly identify clusters—we reduced the minimum occurrence threshold to four. This step gave us 236 keywords, 217 of which were connected and visible in the map (i.e., this graph does not show the keywords that are completely isolated at the edges of the network). We followed the same parameters for the analysis of both author *and* indexed keywords (Figure 18), which identified 1008 out of 13, 581 keywords. The map shows the strongest 1000 keywords and their co-occurrence connections.

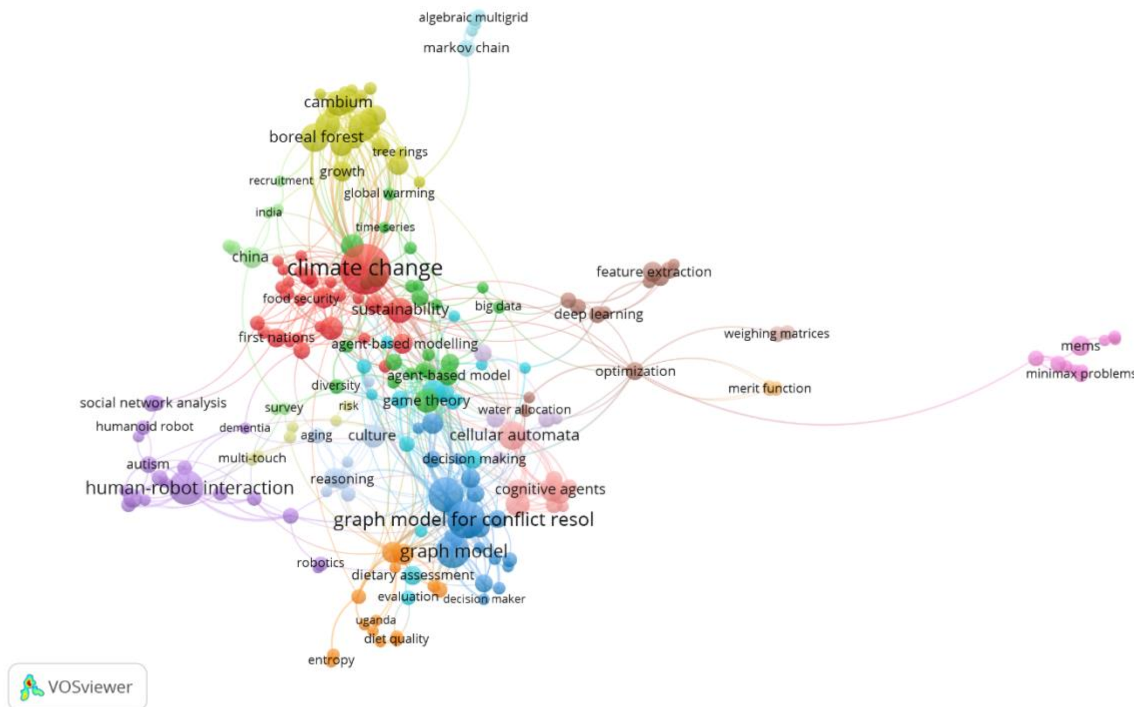


Figure 19: Co-occurrence of Author Keywords

While the colours are not consistent across the maps, we can see that quite a few additional clusters are identified when using keywords (either author keywords of Scopus indexed keywords) versus just using terms from a text mining technique. In the first case (Figure 17), we see about a dozen keyword clusters that are much more granular than the first co-occurrence of terms map. In fact, we see more of the terms here that are included in our Scopus complex systems search query (e.g., markov chain, agent-based model or modeling, social network analysis, and cellular automata) (see Appendix A).

In the second case (Figure 18), the clusters track more closely with the first co-occurrence of terms network, but show an even greater representation of keywords related to medicine/health, robotics, artificial intelligence, and algorithms. Interestingly, the graph shows more relatedness between keywords about the graph model of conflict resolution and decision making, and climate change and environmental policy—a link that other analyses did not capture very well.

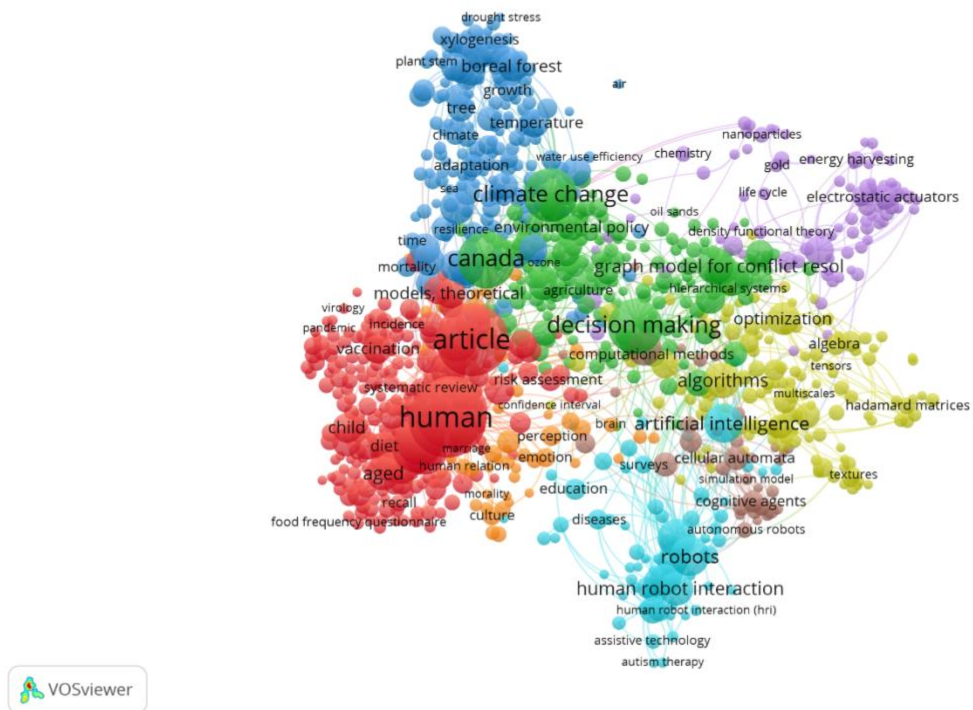


Figure 20: Co-occurrence of All Keywords (Author and Indexed).

Another way to analyze a publication sett in VOSviewer is through co-authorship. To build our co-authorship map, we used the same dataset as in previous maps, and used VOSviewer's "full counting" method to ensure each co-authorship link has the same weight. We reduced the minimum number of documents of an author to 1 (down from the default of 5), which gave us a total of 3550 authors. To aid in identifying clusters, we selected 1000 with the strongest link weight calculation, all of which are visualized in the map. We were unable to export the lists of authors in each cluster, but this data is available and navigable in VOSviewer itself. In the resultant map below (Figure 19a), we see clusters of authors who frequently publish together (based on the dataset we imported). The size of the nodes is weighted by the number of documents an author has in the dataset. These same results are largely reflected in the Collaborating Authors analysis in the SciVal Author Reports, though with some differences because of the thresholds of the VOSviewer analysis. It was difficult to extract legible screenshots from the VOSviewer analysis, due to the size and spread of the network, so we included several zoomed-in views. Again, the analysis is more navigable in the platform itself. While the VOSviewer maps do not have the "granularity" that Gargantext has (i.e., the ability to drill down to see lists of neighbouring nodes and relevant publications), we can use some of the data and analyses from SciVal to see details about the authors named in this graph.[27]

[27] Simply use the search tool in the Supplemental Report pdf to find individual authors (they may appear in Collaborating Author lists in the Author Reports or in Top Author lists in the Corpus Reports). Further, if the author is named in the Supplemental Report, then they "exist" in Scopus and SciVal; one could easily find their affiliation(s), publications, and other metrics.

There are many insights we can draw from this graph about strong co-author relationships, the breadth of collaboration by an author, the number of authors publishing in similar areas, key authors that link two different publishing communities, and more. For example, note the closeness between Anand, M. and Bauch, C.T. (see Figure 19b below), since they publish very frequently together, though not as frequently with others (as indicated by relatively few out-links). Also, note the high number of authors in the green cluster at the top-left dominated by Dautenhahn, K. and Nehaniv, C. (see Figure 19c below), both of whom are in the Faculty of Engineering at the University of Waterloo and publish frequently on the subjects of robots and human-robot interactions. This cluster is highly connected, suggesting a rather cohesive research community. Lastly, as with the "bridging terms" in the Gargantext map, here, we can see "bridging authors." For example, the cluster at the bottom-left dominated by Hipel, K. is far from the cluster near the top-centre dominated by Fieguth, P., but they share as a bridge node the author Liu, Y. (see Figure 19d).

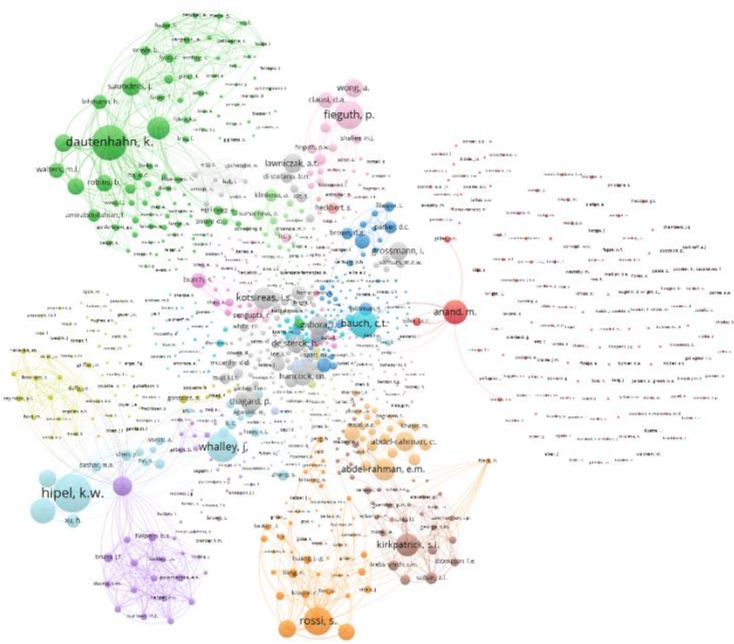


Figure 21a: Co-authorship Network.

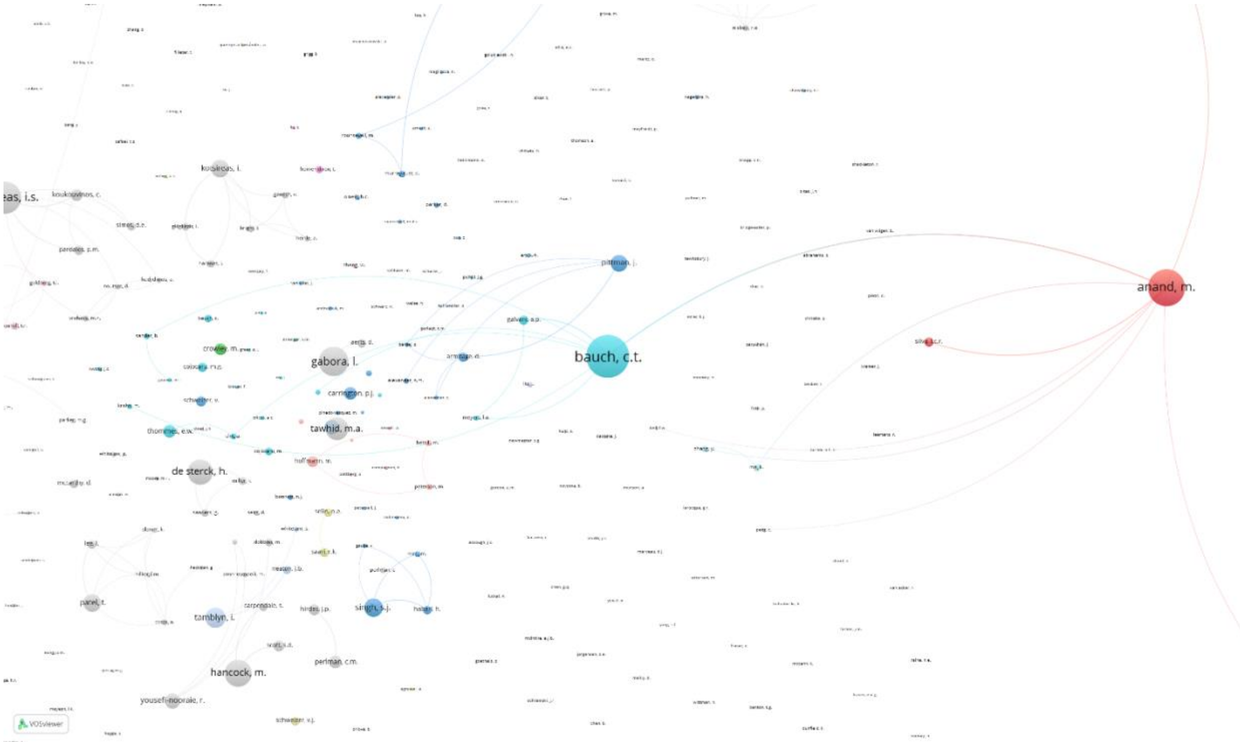


Figure 21b: Co-authorship Network, zoomed in on Bauch, C.T. and Anand, M..

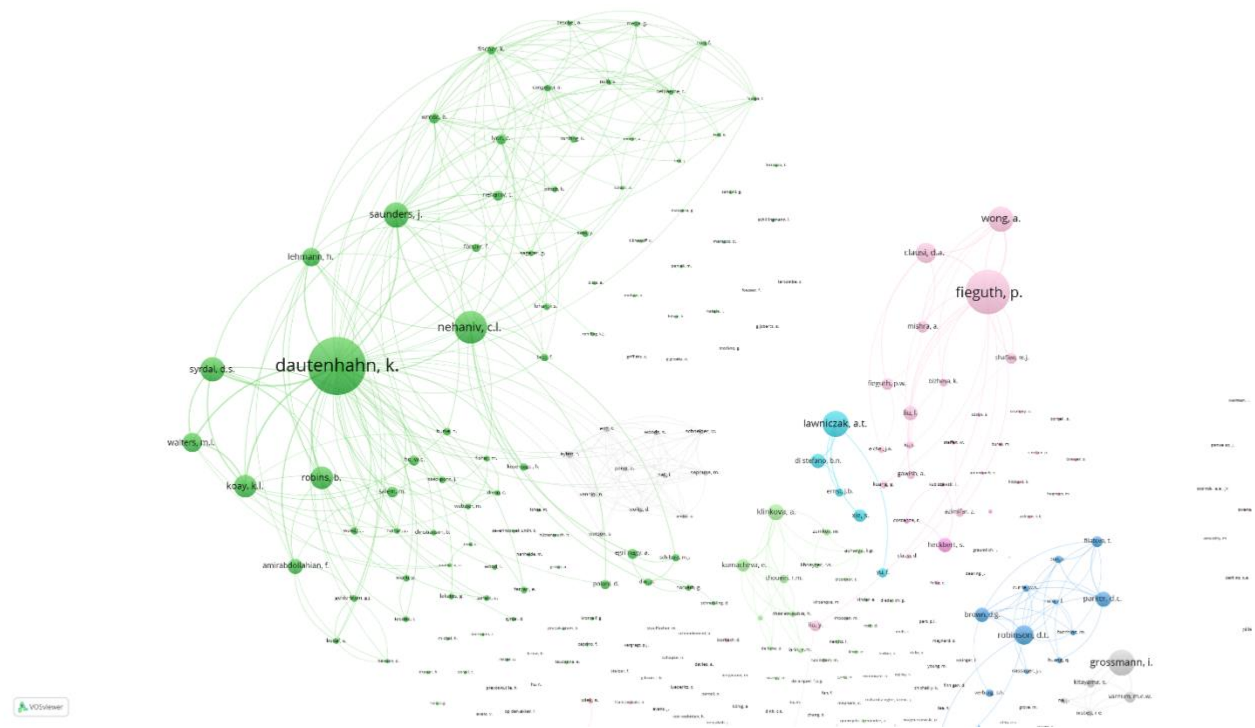


Figure 21c: Co-authorship Network, zoomed in on Dautenhahn, K. and Nehaniv, C.

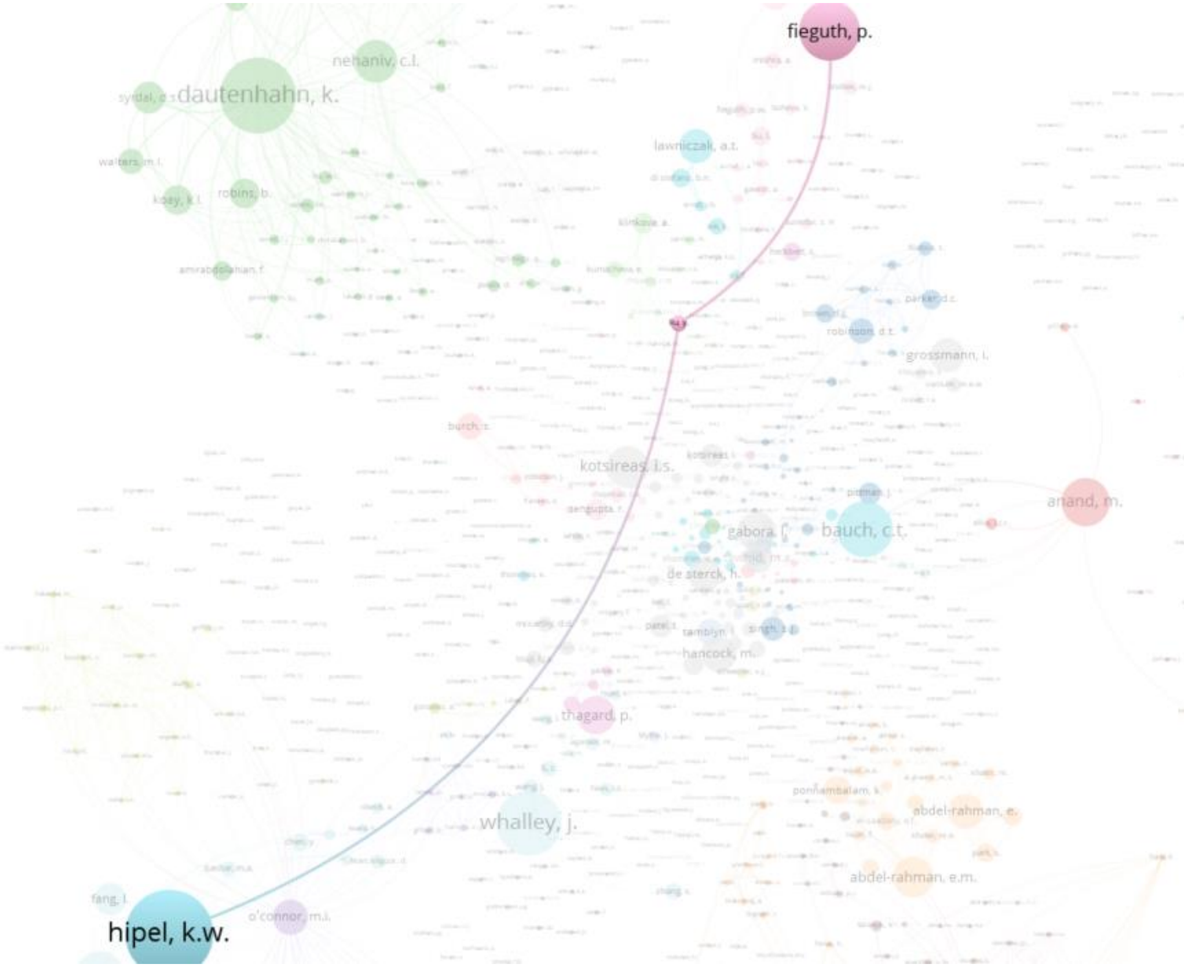


Figure 21d: Co-authorship Network, zoomed in on Hipel, K. and Fieguth, P.

In the final two network analyses, each node represents an academic journal or publication source, and the larger the node, the greater number of citations the journal has received in this dataset.

For the bibliographic coupling of journals analysis, we followed the guidance by van Eck and Waltman (2014) and selected fractional counting and—given the relatively small size of our publication dataset—a minimum of three documents per source. Of the 1028 sources in the corpus, 198 met the citation threshold and Figure 20 shows the 176 sources that are connected. In Figure 20, two journals are linked if there is a third journal that is cited by both journals, and distance indicates how often that happens. Put another way, "bibliographic coupling is about the overlap in the reference lists of publications."[28]

For the co-citation of journals analysis, the initial results following the same parameters were too large to offer much meaning (over 27,000 journals). Instead, we raised the citation minimum to 20, which left us with 393 sources. In Figure 21, two journals are linked if there is a third journal that cites them both,

[28] Ibid.

and journals are positioned closely together when they have more co-citations. Closeness indicates relatedness, based on co-citations.

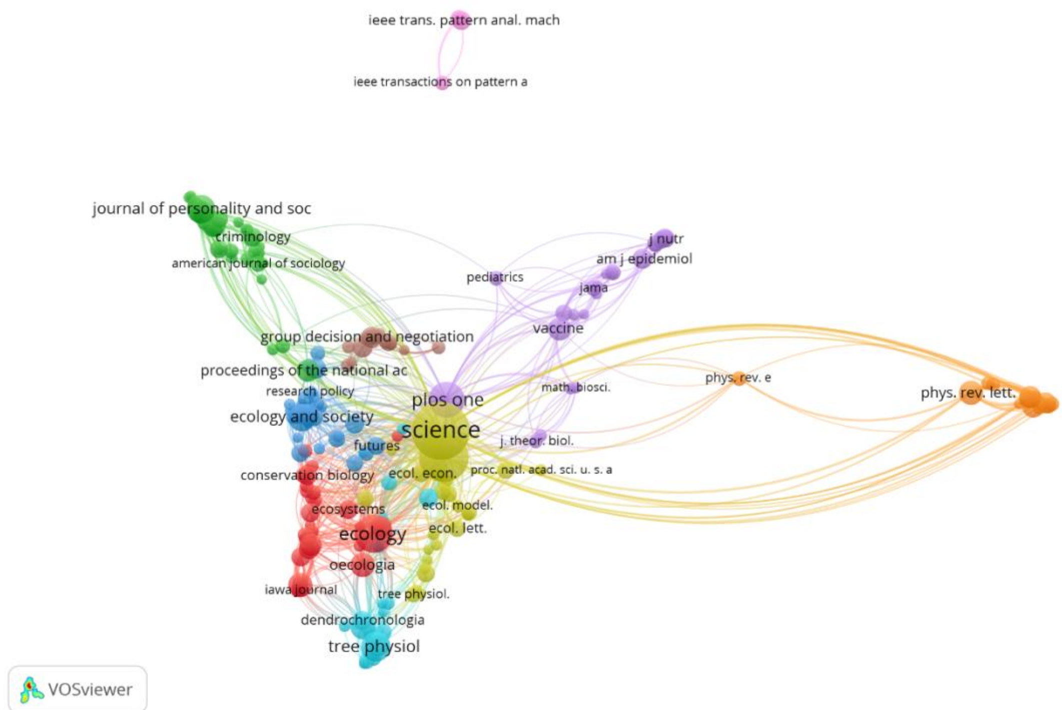


Figure 22: Bibliographic Coupling of Sources, weighted by citations.

Looking at the two maps, while the cluster colours are not consistent, we see similar clusters emerge. Some of these include journals related to ecology, ecological economics, medicine and nutrition, physics, sociology and criminology, tree research, decision making and negotiation, biology, psychology, computer science, and science more generally. Here, we can again use the SciVal data in the Supplemental Report to see further details about each source (in connection with this or another corpus), and cross-reference with the Author Reports' lists of top publication sources (i.e., "Scholarly Output by Scopus Source").

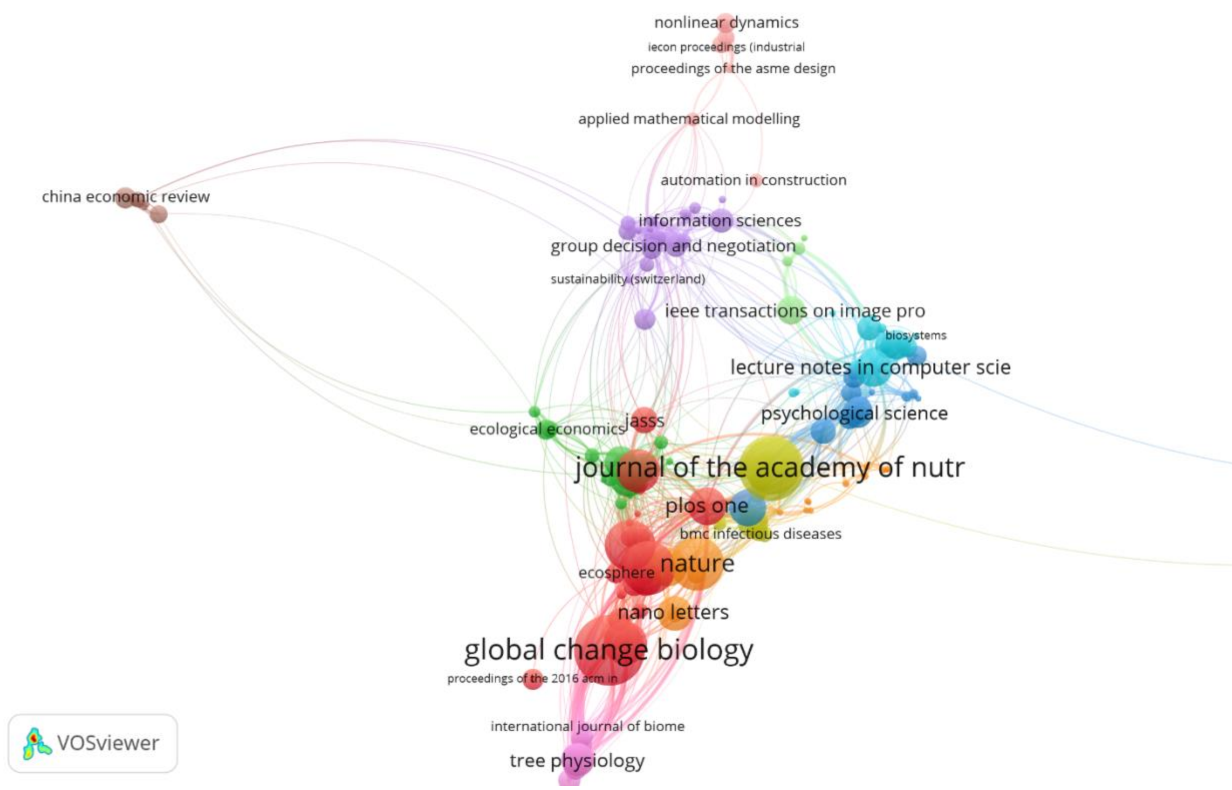


Figure 23: Co-citation of Sources.

Our findings in this section merely scratch the surface of what is possible with bibliometric analysis using Gargantext, VOSviewer, and other tools not yet employed. But even at this stage, our analysis network provides great insight into the thematic clusters of research across the WICI member community; underlines the importance of author keywords in getting a more granular and accurate view of topic clusters; highlights areas of work that are very interdisciplinary and others that are more siloed; points to opportunities for cross-cluster collaboration; visualizes the relevant academic journal landscape; and identifies potential new WICI members at both University of Waterloo and across Canada.

In terms of evaluating the tools, both Gargantext and VOSviewer provide a lot of analytical power. The greatest benefit of Gargantext is the ability to drill down and interact with the network map. However, a key limitation is the inability to export the underlying data in multiple formats, and especially the lack of full lists of publications related to specific nodes. Furthermore, adding capacity to run various co-author and citation network analyses in the same platform would make it much stronger. VOSviewer has very broad capability, and so provides several ways of looking at and understanding the same corpus of publications. Still, some sort of interaction capacity (e.g., the ability to quickly view the list of publications by an author in a co-authorship network, or view the list of publications that are represented by a keyword node in a co-occurrence of terms network) would enable researchers to work with their bibliometric data more iteratively and at multiple levels.

4.0 University of Waterloo Community

The WICI Member analysis required us to first (1) define *who* made up the community by using author's Scopus IDs, then (2) gather all of the community's publications including non-complex systems research, and (3) identify the subset of complex systems publications using our self-developed search query. In contrast, because Scopus and SciVal attach institutional affiliations to individual authors, we could skip right to the third step to identify the subset of complex systems publications by scholars based at the University of Waterloo (including WICI members and students based at UW).

4.1 Data & Methods

To build this corpus, we again used our search query and limited the results by publication year range 2009-2019 and by institutional affiliation to the University of Waterloo. We exported the results to SciVal for analysis, the full results of which can be found in the Supplemental Report (see pages 618 to 686). Section 4.2 discusses several highlights and key findings from SciVal. We also conducted a manual keyword analysis of this corpus—which is discussed in section 4.3—and compared the results to the WICI Core Member corpus, and to our complex systems search query terms.

Future research could easily import these data into VOSviewer, Gargantext, or other software for various network analyses as well.

4.2 Findings: SciVal Reports

Overall metrics

The University of Waterloo query filtered by the complex systems query produced 3,245 publications by 7,553 authors (each publication has at least one (co)author based at UW). On average, each publication has 20.5 citations; the total citation count for the corpus is 66,637 at the time of analysis. Over 20 per cent of those publications are in the Top 10% Citation Percentile, and nearly 43% are published in the Top 10% Journal Percentiles. Furthermore, over 50% of the publications involved international collaboration.

Subject areas

The most common subject areas track very closely with those in the "All WICI Members All Publications, filtered by CS query" corpus, with Computer Science and Engineering leading strongly followed by Mathematics, Physics and Astronomy, Materials Science, Environmental Science, and Social Sciences. Chemistry, Biochemistry and Genetics, and Medicine are just over the 200 publications level (see Figure 22 below). A further breakdown of sub-subjects is visualized in the Tree Map on page 621 of the Supplemental Report (but only the online version is interactive).

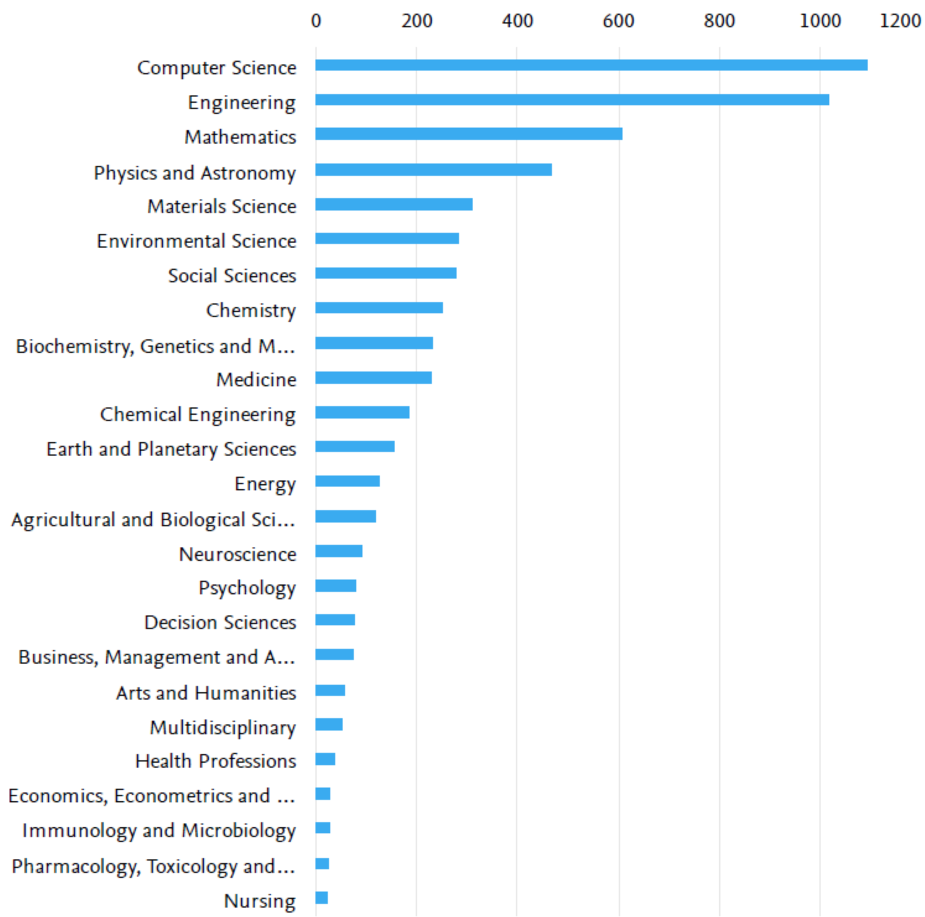


Figure 24: Publications by Subject Area, UW CS query

These results are similar to the WICI member community (publications by all members, filtered by our complex systems query) with a few key exceptions (see Figure 25 below for comparison). First, environmental science and social science rank higher in the WICI member complex systems publications than in the UW-wide complex systems search results *and* the set of *all* WICI member publications. Therefore, we surmise that complexity related research in the environmental and social sciences is better represented among WICI members' work than in complexity research more broadly. On the other hand, we see in Figure 24 that there is a significant amount of complex systems research being conducted in engineering, mathematics, physics and astronomy, and materials science that is *not* well represented in the WICI membership. Lastly, note that most of the complexity publications in the arts and humanities, immunology and microbiology, and multidisciplinary research shown in the UW-wide graph are captured in the WICI members' complex systems subset, suggesting that WICI members do, in fact, account for the majority of this work across campus.

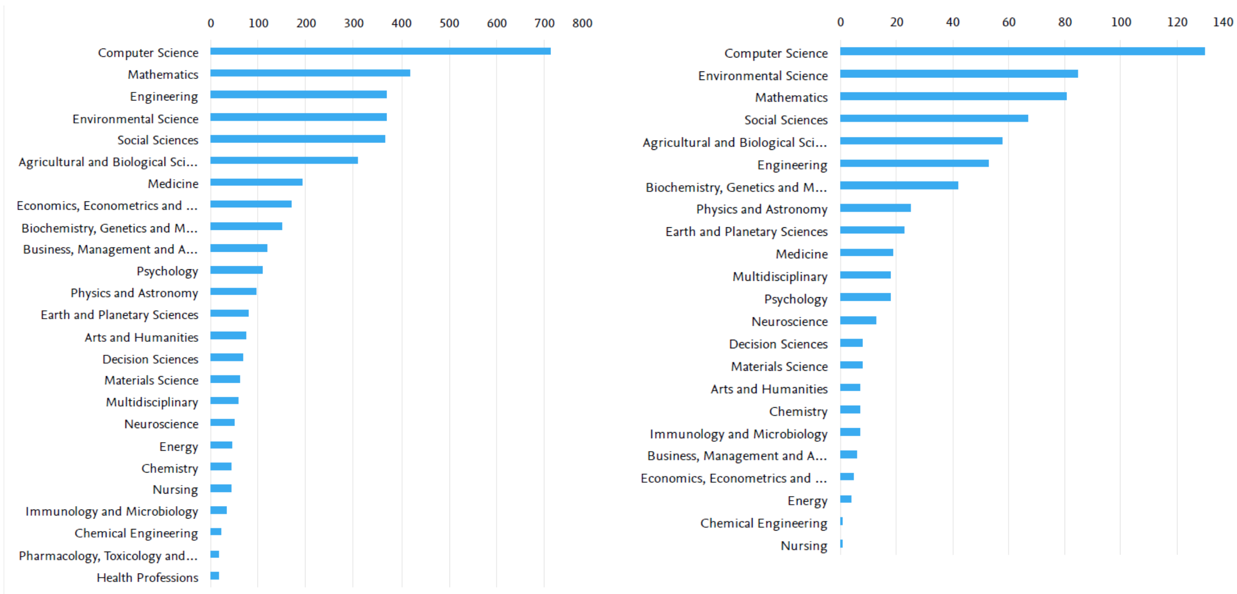


Figure 25: Publications by Subject Area, comparison between “All WICI All Publications” and “All WICI All Publications, filtered by CS query”

The wheel of Topics (see Figure 26) shows the distribution of research in a different way (again, refer to SciVal's definition and methodology behind Topics and Topic Clusters in Appendix D). Each circle or bubble represents a Topic, and the size of the bubble represents the number of publications on this topic in this specific dataset. The location of the bubbles in the wheel is based the journals where the research is published, and the disciplinary categories of those journals (according to the All Science Journal Classification (ASJC) categories). Topic position relates to the whole Topic (i.e., everything in the Scopus database) and not just to the specific dataset under study. Topics closer to the centre of the wheel are more likely to be multidisciplinary than those near the edge.[29] The online version of the SciVal Report is interactive, and users can hover over each individual circle to see the name of the topic. Unfortunately, the static image does not offer much in the way of insights beyond what we could determine from the subjects breakdown above.

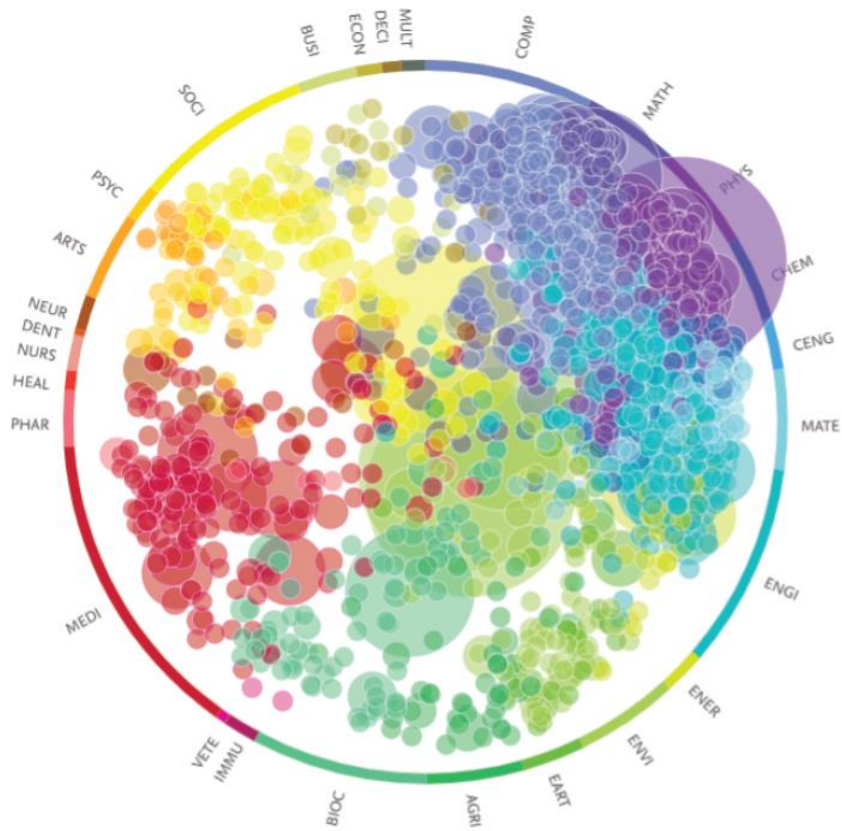


Figure 26: Topics Wheel (produced in SciVal). See the Supplemental Report for the legend of field/discipline names.

[29] "Topic Prominence in Science FAQs," Elsevier B.V. Available at: https://service-elsevier-com.proxy.lib.uwaterloo.ca/app/answers/detail/a_id/28428/supporthub/scival/.

Collaboration, top institutions

In the Collaboration details, we see that publications with international collaboration have higher average citations per publication than those with only Canadian collaboration (24.7 compared to 19.9). The same is also true for Canadian collaboration compared with UW-only collaboration.

The SciVal analysis also showed the top ten institutions where (co)authors of this set of publications are based. Since we limited the search to University of Waterloo affiliation, UW expectedly tops the list. The next three highest ranked in terms of scholarly output are: University of Toronto, Perimeter Institute for Theoretical Physics, and University of Guelph. The remaining institutions are quite close in scholarly output and include: McMaster University, Ryerson University, Nanyang Technological University, the French National Centre for Scientific Research (CNRS), Western University, and University of Ottawa.

Table 6: Top 10 Institutions ranked by publications in this dataset (produced in SciVal)

	Name	Scholarly Output	Authors	Citations per Publication
1.	**University of Waterloo**	3,218 ▲	2,712 ▲	20.5
2.	University of Toronto	127	231 ▲	39.0
3.	Perimeter Institute for Theoretical Physics	103 ▲	75 ▲	41.3
4.	University of Guelph	73 ▲	105	23.2
5.	McMaster University	54 ▲	96 ▲	25.8
6.	Ryerson University	51 ▲	33 ▲	21.9
7.	Nanyang Technological University	49 ▲	62 ▲	33.8
8.	CNRS	47	85	21.6
9.	Western University	47 ▲	85 ▲	23.7
10.	University of Ottawa	41 ▲	62	34.9

To clarify, this list comprises the institutions where (co)authors of our Scopus search results (which were filtered by our complex systems query) are affiliated. As such, it is likely that several of these institutions have enough complex systems research taking place to warrant a “node” designation in a future Canadian Network of Complex Systems. Note that this list does not capture institutions where there is *interest* (emerging or otherwise) in complex systems research but few or no related publications during 2009-2018.

Top authors

Finally, our SciVal analysis of this publication set shows the top 500 authors (by scholarly output) in this publications set, limited to the University of Waterloo (with some exceptions – some co-authors are based at institutions elsewhere in Canada or internationally) (see pages 630-686 of the Supplemental Report). To clarify, this list comprises scholars who are (co)authors on publications captured by our Scopus search, limited by UW affiliation and filtered through our complex systems query. After importing this list to our Airtable database for ease of sorting and analysis, we found that a total of 355 authors in the top 500 (i.e., those in our dataset with the largest scholarly output) are based at the University of Waterloo (see list below), 16 of whom are WICI members (see list below). Two additional external WICI members—Madhur Anand and Shreyas Sundaram—were identified in the top 500 as well.

- Abdel-Rahman, Eihab
- Bauch, Chris T.
- Burch, Sarah Lynne M.
- De Sterck, Hans
- Deadman, Peter J.
- Fieguth, Paul W.
- Gallupe, Owen
- Hipel, Keith William
- Homer-Dixon, Thomas F.
- Parker, Dawn Cassandra
- Pittman, Jeremy B.
- Ricardez-Sandoval, Luis Alberto
- Robinson, Derek T.
- Schweizer, Vanessa Jine
- Thagard, Paul R.
- Zhang, Haotian

Note, here, the absence of Core members Mark Crowley, Igor Grossmann, Sharon Kirkpatrick, Chrystopher Nehaniv, Stephen Quilley, and of External Core members Liane Gabora, Mary O’Connor, Raja Sengupta, and Roger white. The absence may be explained by the inability of our complex systems search query to capture all of WICI Core members’ complexity publications, either because of missing keywords in our list of terms, or because of too-sparse use of complex systems terms in those publications.

The remaining list of approximately 337 authors based at UW is the first place WICI should look for complex systems researchers not yet connected to WICI (again, see pages 630-686 of the Supplemental Report).

4.3 Findings: Keyword Analysis

One type of analysis that was not available in SciVal for either the WICI or UW communities was an analysis of official Author Keywords (though we did conduct *keyphrase* analysis in SciVal). After filtering all the WICI community corpora by our complex systems query and identifying complex systems research at the University of Waterloo, we imported all of the bibliometric results into our Airtable database and extracted the Author Keywords used in the publications (approximately 950 unique terms or phrases). After merging terms with slightly different spelling, we calculated the frequency of each keyword across the whole publication dataset. While many keywords were used only once or twice, 58 were used ten or more times (see Table 4 for the full ranked list). Several keywords were very well represented, including adaptation, climate change, genetic algorithm(s), robustness, optimization, self-assembly, multi-agent system(s), social network(s), black holes, distributed generation, agent-based modelling/models, and machine learning.

Table 7: Author Keywords by frequency (min. 10 occurrences)

Adaptation	99	Stability	15
Climate Change	90	Synchronization	15
Genetic Algorithm(s)	79	Distribution System(s)	15
Robustness	50	Multiscale Mode(l)ling	14
Optimization	47	Vulnerability	14
Self-Assembly	45	Resource Allocation	14
Multi-Agent System(s)	45	Uncertainty Analysis	13
Social Network(s)	40	Sustainability	13
Black Holes	29	Multi-Agent	13
Distributed Generation	29	Complex Network(s)	13
Agent-Based Mode(l)ling	27	Model(l)ing	13
Agent-Based Model(s)	23	Classical Theories Of Gravity	12
Machine Learning	22	Reinforcement Learning	12
Governance	19	Mobile Social Network(s)	12
Smart Grid(s)	18	Decision(-)Making	12
Social Network Analysis	18	Evolution	11
Simulation	18	Complexity	11
Multi-Scale	18	Policy	11
Boolean Function(s)	17	Adaptive Control	11
Nanoparticle(s)	17	Robust Control	11
Resilience	16	Artificial Neural Network(s)	11
Phase Transition(s)	16	Time Delay	10
Social/Socio-Ecological System(s)	16	Particle Swarm Optimization	10
Climate Change Adaptation	15	Security	10
Chaos	15	Uncertainty	10
Canada	15	Multi-Objective Optimization	10
Neural Networks	15	Fractal(s)	10
Deep Learning	15	Trust Model(l)ing	10
Bifurcation	15	Hopf Bifurcation	10

By comparing this list of Author Keywords to the list of terms in our complex systems search query, we can see which of our search terms showed up (or not) as keywords, and what additional complex systems-related keywords that were present in the dataset but missing in our search query (see Table 7). Approximately 35 of our search terms were present as Author Keywords in the dataset, meaning 44 were not. Note that some of our search terms may still have been present in the abstracts of publications. A review of the full set of Author Keywords showed that approximately 70 terms/phrases could be considered relevant to complex systems research. Future bibliometric research on complex systems should consult these lists to further refine a complexity search query.

Table 8: Complex Systems Keyword Comparison

Terms from our CS query absent as Author Keywords	**Terms from our CS query present as Author Keywords**	**Additional CS-related keywords present**
Adaptive walk	Adaptation	Adaptive capacity
Artificial societ*	Agent based/agent-based	Adaptive control
Autopoe*	Attractor	Adaptive management
Basin of attraction	Bifurcation	Adaptive networks
Cascading failure	Bipartite graph	Artificial neural networks
Catastrophe theory	Boolean function	Biologically inspired computing
Clustering coefficient	Chaos	Bistability
Complexity science	Coevolution	Chaos synchronization
Connectionis*	Complex adaptive system	Chaotic attractor
Convergent evolution	Complex network	Chaotic maps
Critical transition	Complexity theory	Community detection
Degree distribution	Criticality	Complexity
Dissipative structure	Embodied cognition	Complex systems
Edge of chaos	Emergence	Composite nonlinear feedback
Effective complexity	Evolutionary game theory	Computational complexity
Emergent behavior/our	Fractal	Cultural evolution
Evolution of cooperation	Genetic algorithm	Decentralized control Distributed (control, generation, model predictive control, parameter systems, power generation)
Fitness landscape	Hysteresis	Dynamic (behavior, optimization, programming, systems)
Historical contingency	Interaction netw*	Embedded systems
Hypercycle	Multi agent/multiagent/multi-agent	Evolution
Hypernetw*	Multi scale/multiscale/multi-scale	Fuzzy logic
Infinite loop	Network analysis	Fuzzy systems
Limit cycle	NP-complete	Mobile social networks
Long-tailed distribution	Opinion dynamics	Nonlinear systems

Morphodynamics	Path dependence/path-dependence	Planning under uncertainty
Morphogenesis	Phase transition	Regime shifts
Nature-inspired computing	Power law	Resilience
SNA	Robustness	Socio/social-ecological systems
Network dynamics	Saddle point	Stability
Nonequilibrium statistical physics	Self assembly	Structurual similarity
Open-ended volution	Self organis/self organiz/self-organis/self-organiz	Trust modelling
Preferential attachment	Social networks	Uncertainty / uncertainty analysis
Punctuated equilibrium	Social simulation	Vulnerability
Scale invariance	Spatial netw*	
Scale-free distribution	Swarm intelligence	
Scale-free network		
Sensitive dependence on initial conditions		
Sensitivity to initial conditions		
Small-world network		
Small-world property		
Socio-semantic network		
Sociophysics		
Synergetics		
Tipping point		

5.0 Canada Community

5.1 Data & Methods (Scopus & SciVal)

For the Canada-wide search, we initially used the same protocol as the UW-wide search. However, the results numbered over 50,000 publications and was therefore too large to export to SciVal (without doing it manually, 2000 articles at a time). Instead, we identified the Canada-wide dataset endogenously in SciVal using their “research area” feature. To define a research area, we started with our complex systems query, and then refined the definition to include only Canadian institutions. Because of the limitations of that feature, we also had to limit the results to only the last five years (2014-2018). However, the results were still too large to analyze in SciVal, so we decided to divide the research area into two, manually separated by “medical” and “non-medical” subject areas. In the end, the “medical” dataset included 28,848 publications and the “non-medical” dataset comprised 60,790 publications. We were then able to conduct most of the same SciVal analyses as in the UW and WICI communities.[30] The full Scival results can be found in the Supplemental Report (see pages 687 to 840).

5.2 Findings (SciVal Reports)

Keyphrase analysis

One additional analysis we could do with the research areas we defined, but not with the other publication sets, was a keyphrase analysis. Below, two word clouds show the most common words/phrases in the dataset, first in the medical subject areas, and second in the non-medical subject areas.

Figure 27: Keyphrase Analysis, medical subject areas (produced in SciVal)

[30] Because we had to treat this population differently than the others due to its size (i.e., define it as a research area in SciVal), we were unable to perform all of the same analyses as the other datasets (including publications by subject area, topics, and topic clusters).

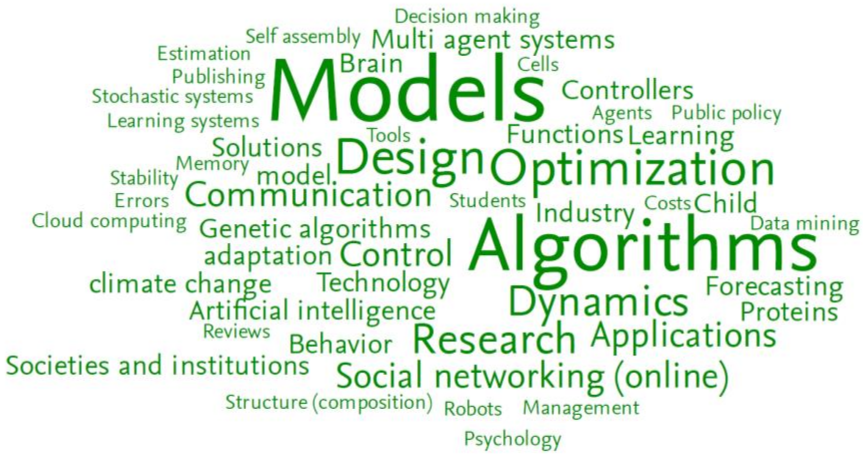


Figure 28: Keyphrase Analysis, non-medical subject areas (produced in SciVal)

Many complex systems keywords (that we used in our Scopus query) show up in the second keyphrase analysis but not the first, likely because the frequency of use is lower than other important keyphrases in the medical subject areas. However, if we compare Figure 26 to the co-occurrence analyses of terms and keywords in the WICI member dataset, we see evidence of similar thematic clusters. Terms such as algorithms, artificial intelligence, and behavior; climate change, adaptation, technology, and societies and institutions; decision making, models, multi agent systems, and public policy. Furthermore, we see many of these terms in the UW keyword analysis (see section 4.3) including algorithms, optimization, multi-agent systems, dynamics, decision making, and others—some of which were not (but perhaps should be in the future) included in our complex systems search query.

Collaboration, top institutions

Across Canada, the top five institutions with the most publications in this complex systems plus medical subjects research area are: University of Toronto, University of British Columbia, McGill University, University of Alberta, and University of Montreal. All of these institutions have medical schools, so unsurprisingly, the University of Waterloo ranks seventeenth on this list. Note that while the graphs below show zero publications for the years 2009 to 2013, this is only because the dataset itself is limited to the years 2014 to 2018.

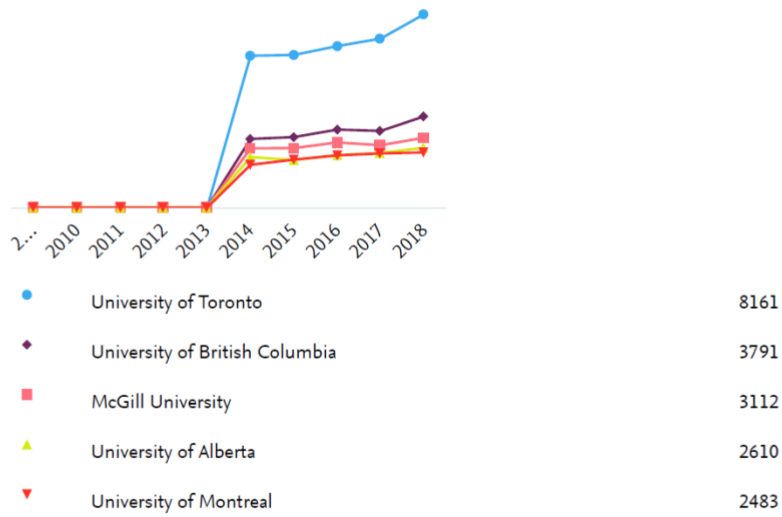


Figure 29: Top 5 Institutions (by # of complex systems publications), medical subject areas, 2014-2018 (produced in SciVal)

For the complex systems plus non-medical subjects research area, the top five Canadian institutions with the most publications are: University of Toronto, University of British Columbia, McGill University, University of Alberta, and University of Waterloo.

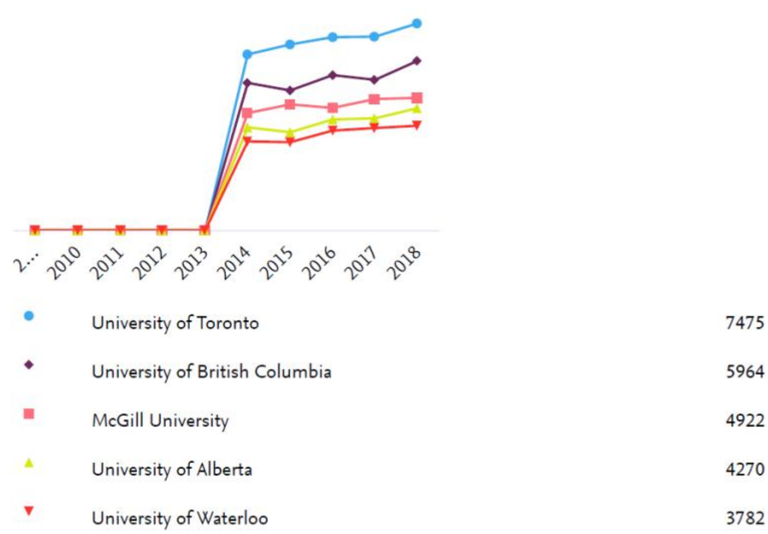


Figure 30: Top 5 Institutions (by # of complex systems publications), non-medical subject areas, 2014-2018 (produced in SciVal)

Top authors

The Supplemental Report also lists—for both the medical and non-medical subject area searches—the top 500 authors (ranked by scholarly output) and their institutional affiliations (see pages 702-751 for the medical subject areas list, and pages 777-829 for the non-medical subject areas list). As with the UW community, we imported the entire 1000-author list into Airtable for easier analysis and found a total of 34 authors who are based at UW, just one of whom—Keith Hipel—is a WICI member). This list of 33 authors (as well as the whole list of 1000) is an excellent starting place to identify high-output potential “complex systems scholars” (strictly or loosely defined) for either WICI membership or collaboration.

- (Sherman) Shen X., Xuemin
- Armitage, Derek R.
- Azad, N. L.
- Boutaba, Raouf
- Budman, Hector M.
- Chen, Pu
- Chen, Zhongwei
- Cohen, Robin S.
- El-Saadany, Ehab
- Elkamel, A.
- Fidan, Baris
- Fischmeister, Sebastian
- Haas, Carl T.M.
- Hassan, Fathy Mohamed Bayoumi
- Hipel, Keith William
- Khajepour, Amir
- Liu, Juewen
- Liu, Xinzhi
- Mann, Robert B.
- Melko, Roger G.
- Mourtzakis, Marina
- Mozaffari, Ahmad
- Pawliszyn, Janusz B.
- Ricardez-Sandoval, Luis Alberto
- Risko, Evan F.
- Salama, Magdy M.A.
- Tam, Michael Kam Chiu
- Varatharajan, Sharanya
- Wang, Xiaosong
- Wang, Zhou
- Wong, Alexander
- Yeow, John T.W.
- Zhou, Y. Norman
- Zhuang, Weihua

Top sources/journals

Finally, our analysis shows the top 100 journals or sources where these publications are found. For the medical subject areas in this dataset, these include: Canadian Journal of Emergency Medicine, BMJ Open, Canadian Journal of Cardiology, Critical Care Medicine, BMC Public Health, Pediatric Critical Care Medicine, Critical Care, Progress in Biomedical Optics and Imaging (Proceedings of SPIE), and Frontiers in Microbiology (see pages 752-761 of the Supplemental Report). For the non-medical subject areas in this dataset, these include: PLoS ONE, Lecture Notes in Computer Science, Scientific Reports, Nature Communications, Proceedings of the National Academy of Sciences of the United States of America, Astrophysical Journal, Proceedings of SPIE (The International Society for Optical Engineering), Physical Review B, Canadian Conference on Electrical and Computer Engineering, and Monthly Notices of the Royal Astronomical Society (see pages 830-840 of the Supplemental Report).

Obviously, these lists of most-published-in journals is a direct reflection of the distribution of subject areas in the dataset. In this case, lead by computer science, engineering, and physics. Future research could parse the larger dataset into individual subject areas and repeat the analyses regarding top collaborative institutions and top journals, in order to get a more targeted understanding.

6.0 Conclusion & Recommendations

This study represents the first comprehensive bibliometric and network analysis of the WICI member community and the broader complex systems community at the University of Waterloo and across Canada. Through its design and integration of analytical tools/platforms, we lay the groundwork for more regular assessment of the community’s research activities, which will enable and strengthen complex systems research at the University of Waterloo.

Our SciVal analysis and reporting, specifically, also offers many insights and leads for the establishment of a Canadian Network of Complex Systems by identifying potential future partners, collaborations, and co-authors. The Top Institutions and Top Collaborators and Authors lists are an excellent place to start. By gaining a fuller picture of the WICI member community—individually and as a group—WICI can better represent its members interests, and communicate more effectively with the University, the members, and potential members. Further, this analysis can offer potential collaborators or funders a more detailed, robust and accurate representation of WICI's research activities.

Using Gargantext and VOSviewer, we identified the emergence of several thematic clusters across WICI members' research, including: (1) conflict resolution (specifically Keith Hipel's work using his graph model), (2) health (with subclusters around disease dynamics and vaccination; child health and health interventions; and diet, food, and assessment design), (3) environment (with subclusters around forestry research; ecosystem services and management, and climate change and adaptation), (4) trade and development, (5) engineering problems and solutions, and (6) behaviour (of systems, people, study participants, governments, and robots). Understanding these clusters may contribute to setting up various thematic working groups, journal special issues, and events that will engage WICI members more effectively. Within WICI, we see especially strong representation of complex systems and health research, as well as environment and climate change related research.

A key challenge we discovered in the course of this research, was the difficulty of actually identifying "complex systems" research and scholars. This difficulty arises in part due to a lack of self-identification by researchers (i.e., there is usually no "complex systems" discipline by which to identify), and a relative lack of complex systems concepts being used as keywords (either by authors or by databases like Scopus). However, even seemingly precise complex systems keywords can pick up a lot of "chaff" in a database search because of the non-technical use of terms such as "non-linearity," "tipping points," or "systems." This broader challenge is one that a more formal Canadian Network of Complex Systems could work to solve (see specific recommendations below).

This report also gave us an opportunity to evaluate several analytical tools. First, SciVal proved to be a highly valuable platform, despite some of its technical shortcomings and glitches. Most valuable, perhaps, is the integration with Scopus, which allows for quicker and deeper analysis of search results. SciVal also enables powerful analysis of several *types* of entities, including individual authors, author groups, publication sets, pre-defined or user-defined research areas, and whole institutions. This breadth gives users many angles from which to analyze their problem or answer their bibliometric research question.

Second, VOSViewer has much greater analytical capacity than Gargantext because of its ability to use each element in a citation as a data point. Users can create visualizations of a publication set based on co-authorship, co-citation, co-occurrence of terms or keywords, citations, shared references, shared publication source (i.e., journal), and more. However, the maps that VOSviewer produces are largely static. They provide an excellent birds-eye-view of the set, but users can only drill down so far in their analysis. VOSviewer results are also not as easily shareable with other researchers as SciVal or Gargantext, because it is a desktop-based platform—one would need to export the network data and then reproduce it in VOSviewer, Gephi, or other network visualization and analysis software. The visualizations themselves, however, are more easily exported for use in presentations or publications.

Third, Gargantext allows users to "look under the hood", so to speak, by easily seeing which publications use which terms, which terms are more frequently connected to others, and several other data interactions. Being able to interact with the documents, analyze individual or sets of terms, and compare across corpora makes the platform unique. However, while Gargantext effectively identified thematic clusters and allowed for meaningful customizability of the "term maps", we would suggest that the tool would be more useful for large sets of short texts, such as social media posts or news articles. Another valuable aspect of Gargantext is its collaborative nature. Because the platform is web-based, a group of researchers could collectively build a map list of terms, and iteratively build a more nuanced visual representation of a field, subject, problem, or question.

Finally, we offer several recommendations for how to leverage the data and findings of this report, as well as some suggestions for future research.

Name the discipline. We see complex systems as an emerging scientific paradigm, characterized by a particular vocabulary, whose application is supported by a wide variety of methods. Funding agencies should include "complex systems science" as a disciplinary category, as currently in place in the UK ESRC funding portal, allowing scholars to self-identify.

Establish a Canadian Network for Complex Systems. Scholars within the network could work to develop standardized keywords, and then could subsequently strive to include these keywords in publications. If they could also include a "Canadian Network for Complex Systems" affiliation of acknowledgement, they could use that tag to self-identify publications they consider to be complex-systems themed or supported. CNCS should collaborate with academic databases, publishers, and other international complex systems institutes to identify complexity science thematic areas.

WICI: Treat the Supplemental SciVal report as a list of "leads." WICI can look to the lists of Top Collaborating authors and institutions (in all the corpora, but especially the full WICI community and UW community) for potential WICI members, co-authors, partners, and future complex systems hubs/nodes.

WICI: Strategically engage with themes. The thematic clusters we identified using Gargantext and VOSviewer could be the basis of forming several WICI working groups, programming streams, or more targeted funding efforts. Initial match between these automatically identified and our self-identified clusters indicates this approach could be productive. Important keywords and terms identified in this report could also be drawn out to support WICI's current thematic focal areas.

WICI: Share data and mobilize knowledge. WICI can make the data and findings of this report accessible to and searchable by WICI members, perhaps through a shared Airtable database and/or Kumu network map connected to the WICI website. The raw data files could also be made accessible via WICI's Dropbox. WICI can also share the online interactive SciVal reports and various research entities with WICI members. Mobilizing knowledge this way could add value to membership by helping members better understand how their work is represented and categorized in one of the largest academic databases, stay on top of their subject areas in a more comprehensive way, and identify potential collaborators, "competitors" and opportunities.

WICI: continue annual bibliometric analysis. In the long term, WICI should perform SciVal analyses and reporting each year as part of the annual report, to support a longer-term strategic plan of growth and collaboration. Now that these have been set up, the Saved Author lists and Saved Search protocols can be easily replicated in the future.

Expand future research. Additional citation and network analyses of the other publication sets identified in this report could also provide more insight about emerging epistemic communities. Future research could compare WICI member's research activity to that of another complexity hub (e.g., the group based at the University of Calgary or the new Complex Systems Lab at Western University), to see how the thematic clusters differ. This research could also be extended to databases other than Scopus.

Acknowledgements

Special thanks to the Complex Systems Institute of Paris Ile-de-France (ISC-PIF) for granting us access to and providing technical support for Gargantext.

References

"About semantic proximity measures used in Gargantext," translated from French by Google Translate, Gargantext. Available at: https://iscpif.fr/gargantext/mesures-utilisees-dans-gargantext/.

"People: Categories of Membership." Waterloo Institute for Complexity and Innovation (WICI), University of Waterloo. Available at: https://uwaterloo.ca/complexity-innovation/people.

Chavalarias, David, Paul Bourgine, Edith Perrier, Fréderic Amblard, François Arlabosse, et al.. 2009. "French Roadmap for complex Systems 2008-2009." HAL. Available at: https://hal.archives-ouvertes.fr/hal-00392486.

"Make Your First Map In 10 Min." Gargantext. Available at: https://iscpif.fr/gargantext/your-first-map/

"Glossary." 2020. Complexity Explorer, Santa Fe Institute. Available at: https://www.complexityexplorer.org/explore/glossary.

Granovetter, Mark. 1973. "The Strength of Weak Ties." *American Journal of Sociology,* Volume 78, Issue 6, pp 1360-1380.

Mitchell, Melanie. 2009. *Complexity: A Guided Tour*. Oxford University Press: New York, NY.

Parker, Dawn, Vanessa Schweizer, Peter Deadman, Chris Bauch, Chrystopher Nehaniv, Sharon Kirkpatrick, and Igor Grossmann. 2019. "Waterloo Institute of Complexity and Innovation 2019 Annual Report." WICI, University of Waterloo. Available at: https://uwaterloo.ca/complexity-innovation/sites/ca.complexity-innovation/files/uploads/files/2019_wici_annual_report_final_0.pdf.

Ranjbar-Sahraei, B. and R.R. Negenborn. 2017. "Research Positioning & Trend Identification." AIDA, TU Delft Library & Centre for Science and Technology Studies, Leiden University. Available at: https://d1rkab7tlqy5f1.cloudfront.net/Library/Themaportalen/Research%20Analytics/AIDA-Booklet.pdf

"SciVal Quick Reference Guide." Elsevier B.V. Available at: https://www-scival-com.proxy.lib.uwaterloo.ca/help/downloads/sv_quick_reference_guide.pdf

"Scival Research Metrics Guidebook." 2019. Elsevier B.V. Available at: https://p.widencdn.net/5pyfuk/ACAD_RL_EB_ElsevierResearchMetricsBook_WEB

"Topic Prominence in Science FAQs." Elsevier B.V. Available at: https://service-elsevier-com.proxy.lib.uwaterloo.ca/app/answers/detail/a_id/28428/supporthub/scival/ .

van Eck, Nees Jan and Ludo Waltman. 2014."Visualizing bibliometric networks." In Y. Ding, R. Rousseau, & D. Wolfram (Eds.), *Measuring scholarly impact: Methods and practice* (pp. 285-320). Springer. Available at: http://dx.doi.org/10.1007/s11192-009-0146-3 van Eck, Nees Jan and Ludo Waltman. 2015. “VOSviewer and CitNetExplorer Tutorial,” presented at the 15th International Conference on Scientometrics & Informetrics, Istanbul, Turkey, June 29, 2015. Available at: https://www.slideshare.net/NeesJanvanEck/issi2015-tutorial-vosviewerandcitnetexplorer

Vicari, Rosa et al. 2019. Supplement of “Climate risks, digital media, and big data: following communication trails to investigate urban communities’ resilience.” *Natural Hazards and Earth System Sciences,* 19, pp. 1485-98.

University of Waterloo Working Group on Bibliometrics. 2016. “White Paper on Bibliometrics, Measuring Research Outputs through Bibliometrics.” Waterloo, Ontario: University of Waterloo.

Appendix

A) Scopus Complex Systems Search Query:

TITLE-ABS-KEY ("adaptation" OR "adaptive walk" OR "agent based" OR "agent-based" OR "artificial societ*" OR "attractor" OR "autopoe*" OR "basin of attraction" OR "bifurcation" OR "bipartite graph" OR "Boolean function" OR "cascading failure" OR "catastrophe theory" OR "chaos" OR "clustering coefficient" OR "coevolution" OR "complex adaptive system" OR "complex network" OR "complexity science" OR "complexity theory" OR "connectionis*" OR "convergent evolution" OR "critical transition" OR "criticality" OR "degree distribution" OR "dissipative structure" OR "edge of chaos" OR "effective complexity" OR "embodied cognition" OR "emergence" OR "emergent behavior" OR "emergent behaviour" OR "evolution of cooperation" OR "evolutionary game theory" OR "fitness landscape" OR "fractal" OR "genetic algorithm" OR "historical contingency" OR "hypercycle" OR "hypernetw*" OR "hysteresis" OR "infinite loop" OR "interaction netw*" OR "limit cycle" OR "long-tailed distribution" OR "morphodynamics" OR "morphogenesis" OR "multi agent" OR "multiagent" OR "multi-agent" OR "multi scale" OR "multiscale" OR "multi-scale" OR "nature-inspired computing" OR "network analysis" OR "SNA" OR "network dynamics" OR "nonequilibrium statistical physics" OR "NP-complete" OR "open-ended evolution" OR "opinion dynamics" OR "path dependence" OR "path-dependence" OR "phase transition" OR "power law" OR "preferential attachment" OR "punctuated equilibrium" OR "robustness" OR "saddle point" OR "scale invariance" OR "scale-free distribution" OR "scale-free network" OR "self assembly" OR "self organis*" OR "self organiz*" OR "self-organis*" OR "self-organiz*" OR "sensitive dependence on initial conditions" OR "sensitivity to initial conditions" OR "small-world network" OR "small-world property" OR "social networks" OR "social simulation" OR "socio-semantic network" OR "sociophysics" OR "spatial netw*" OR "swarm intelligence" OR "synergetics" OR "tipping point")

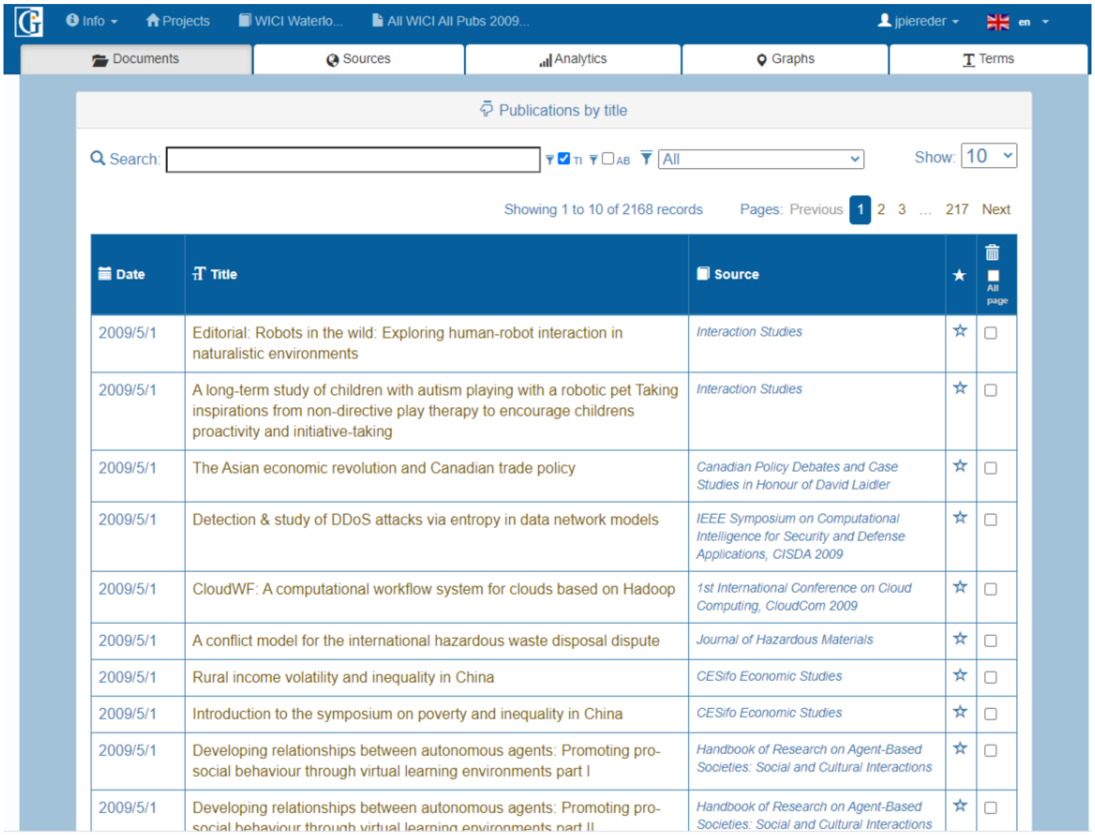
B) List of Analyses/Reports

	SciVal	Gargantext	VOSviewer
Core Members	Author Reports		
Core Members – Self-reported publications	Publication Set Report		
Core Members – Self-reported publications, filtered by CS query	Publication Set Report		
Core Members – All publications 2009-2018	Publication Set Report		
Core Members – All publications, filtered by CS query	Publication Set Report		

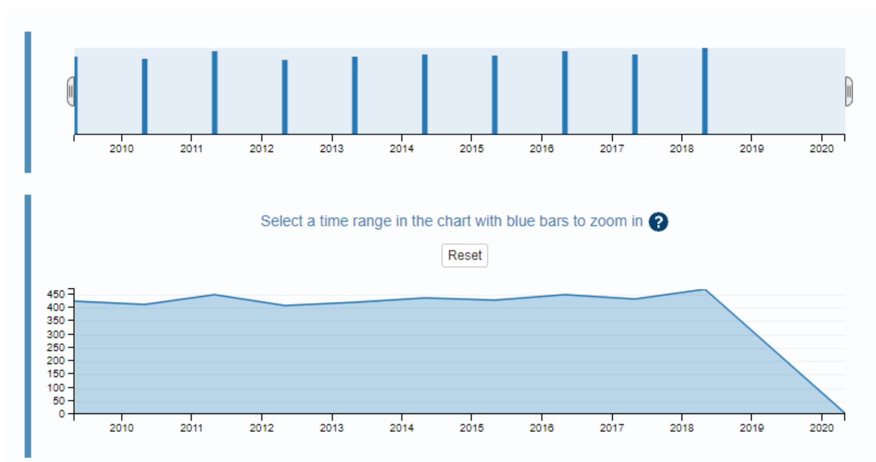
Extended Core Members (Internal + External)	Author Reports		
Extended Core Members – All publications 2009-2018	Publication Set Report		
Extended Core Members – All publications 2009-2018, filtered by CS query	Publication Set Report		
All WICI Members	No Author Reports yet		
All WICI Members – All publications 2009-2018	Publication Set Report	Co-word/term network analysis	Co-occurrence of terms analysis; Co-authorship network analysis; Co-occurrence analyses of author keywords and all keywords; Co-citation of sources analysis; Bibliographic coupling of sources;
All WICI Members – Publications 2009-2018, filtered by CS query	Publication Set Report	Future research	Future research
All WICI Members – Publications 2009-2018 (health and medical)	Publication Set Report		
University of Waterloo – Publications 2009-2018, filtered by CS query	Publication Set Report	Future research	Future research
Canada – Publications 2014-2018, filtered by CS query (medical subject areas)	Publication Set Report		
Canada – Publications 2014-2018, filtered by CS query (non-medical subject areas)	Publication Set Report		

C) Tour of Gargantext

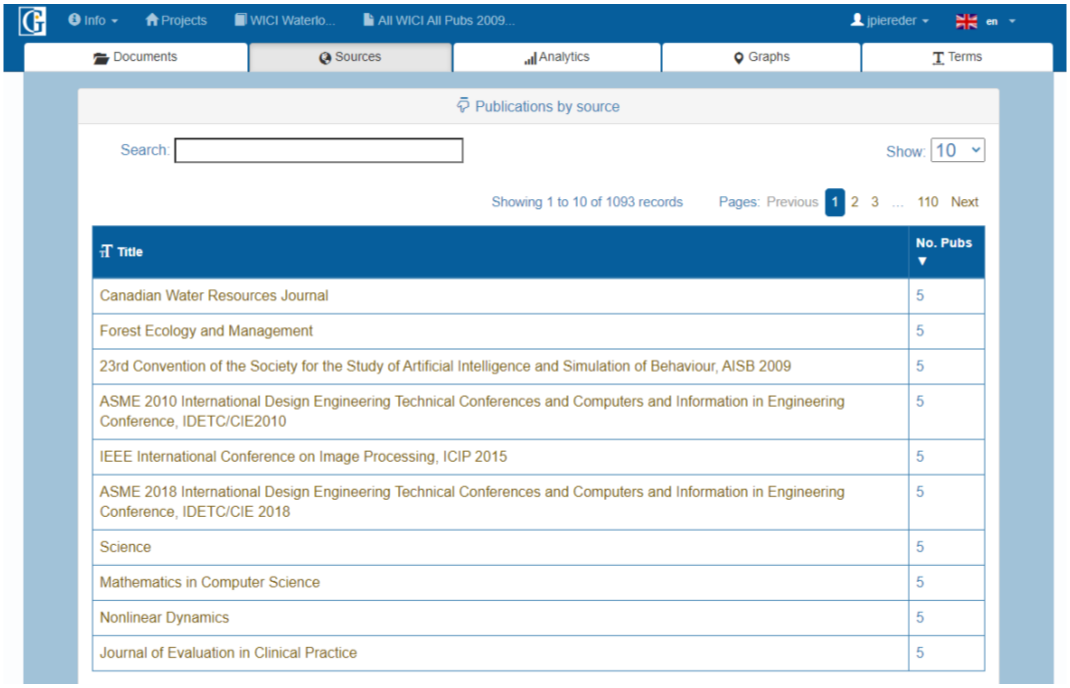
The first step in using Gargantext is to upload a text corpus (e.g., publication set, dataset of tweets, collection of news articles, etc.). After the upload and initial analysis is complete, the “documents” view shows the full list of imported records, including publication date, title, and source. Here, users can search for specific documents, sort by different fields, mark documents as favourites, delete documents from the corpus, show the list of favourites, and show duplicates.



There is also a fully adjustable temporal histogram:

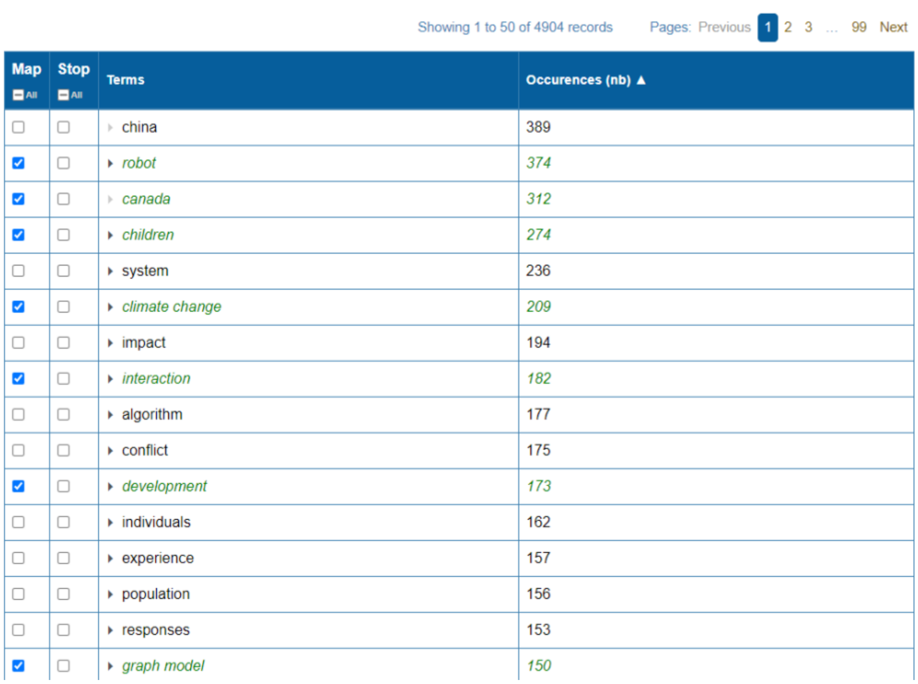


Users can also search for journals, or see a list of sources ranked by number of publications in the "sources" view:



Title	No. Pubs ▼
Canadian Water Resources Journal	5
Forest Ecology and Management	5
23rd Convention of the Society for the Study of Artificial Intelligence and Simulation of Behaviour, AISB 2009	5
ASME 2010 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference, IDETC/CIE2010	5
IEEE International Conference on Image Processing, ICIP 2015	5
ASME 2018 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference, IDETC/CIE 2018	5
Science	5
Mathematics in Computer Science	5
Nonlinear Dynamics	5
Journal of Evaluation in Clinical Practice	5

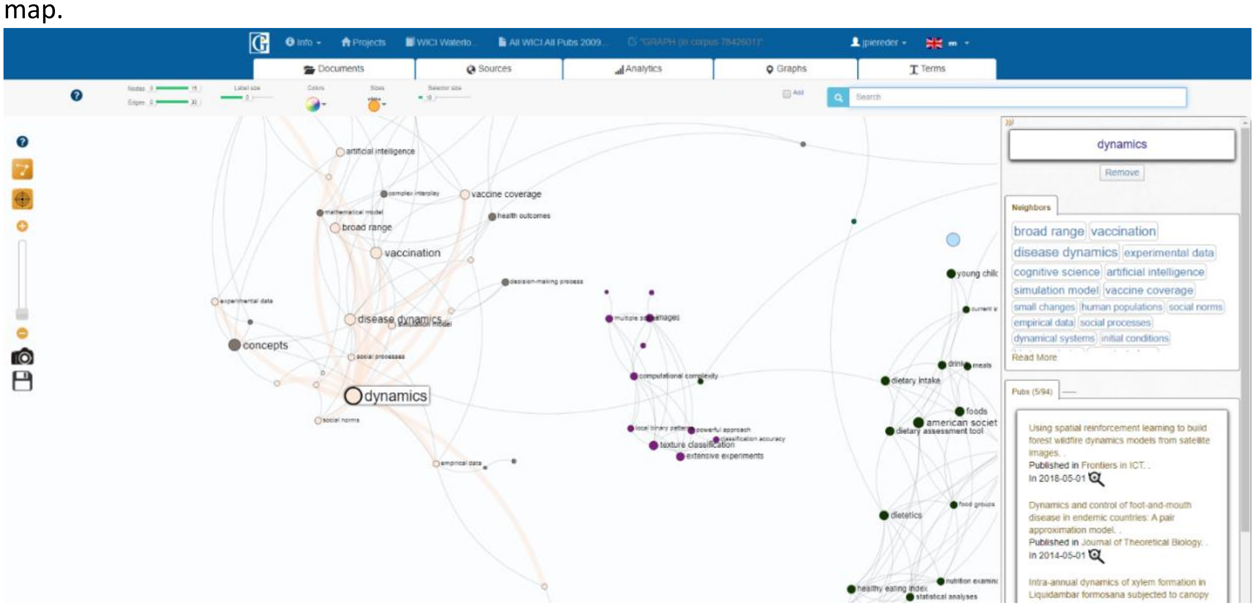
After a user has uploaded their corpus and Gargantext runs the initial text mining, the "terms" view shows the full list of extracted terms and number of occurrences. Using natural language processing, the software pre-identifies potential Maplist terms and "stop" terms (e.g., and, is, there, etc.), and groups plural or similar terms (e.g., robot, robotics, robots). However, users can manually edit the groups and entire list as needed.



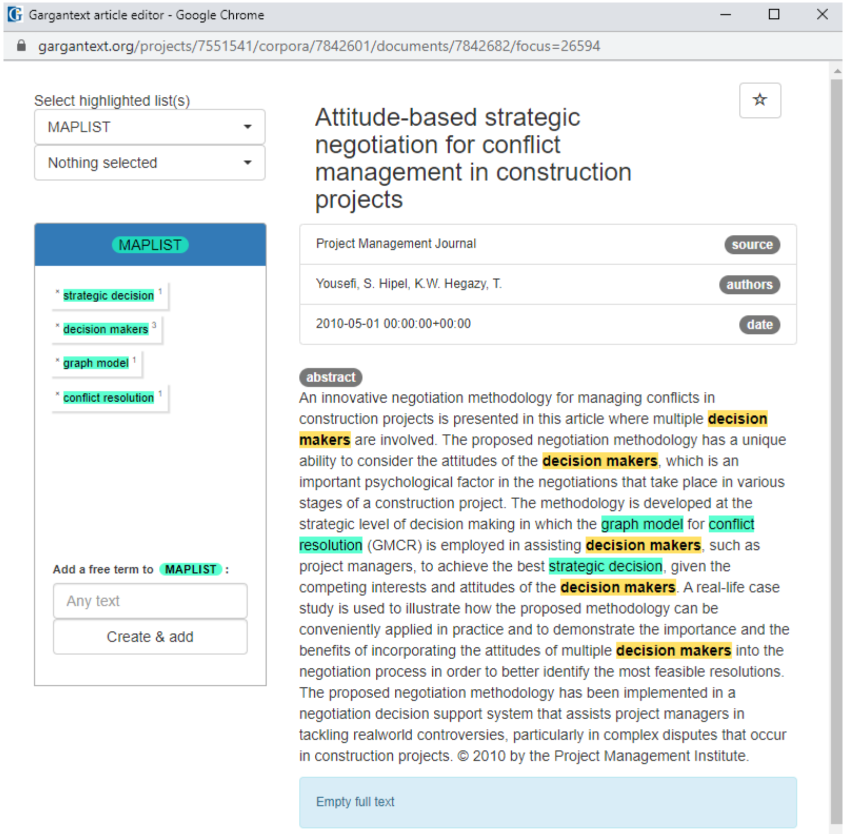
Map All	Stop All	Terms	Occurences (nb) ▲
☐	☐	china	389
☑	☐	*robot*	*374*
☑	☐	*canada*	*312*
☑	☐	*children*	*274*
☐	☐	system	236
☑	☐	*climate change*	*209*
☐	☐	impact	194
☑	☐	*interaction*	*182*
☐	☐	algorithm	177
☐	☐	conflict	175
☑	☐	*development*	*173*
☐	☐	individuals	162
☐	☐	experience	157
☐	☐	population	156
☐	☐	responses	153
☑	☐	*graph model*	*150*

Once a user has finalized their Maplist of terms, they can generate network maps in the "graphs" view (this process can take a while). See https://iscpif.fr/gargantext/mesures-utilisees-dans-gargantext/ for an explanation of the semantic proximity measures used in Gargantext.

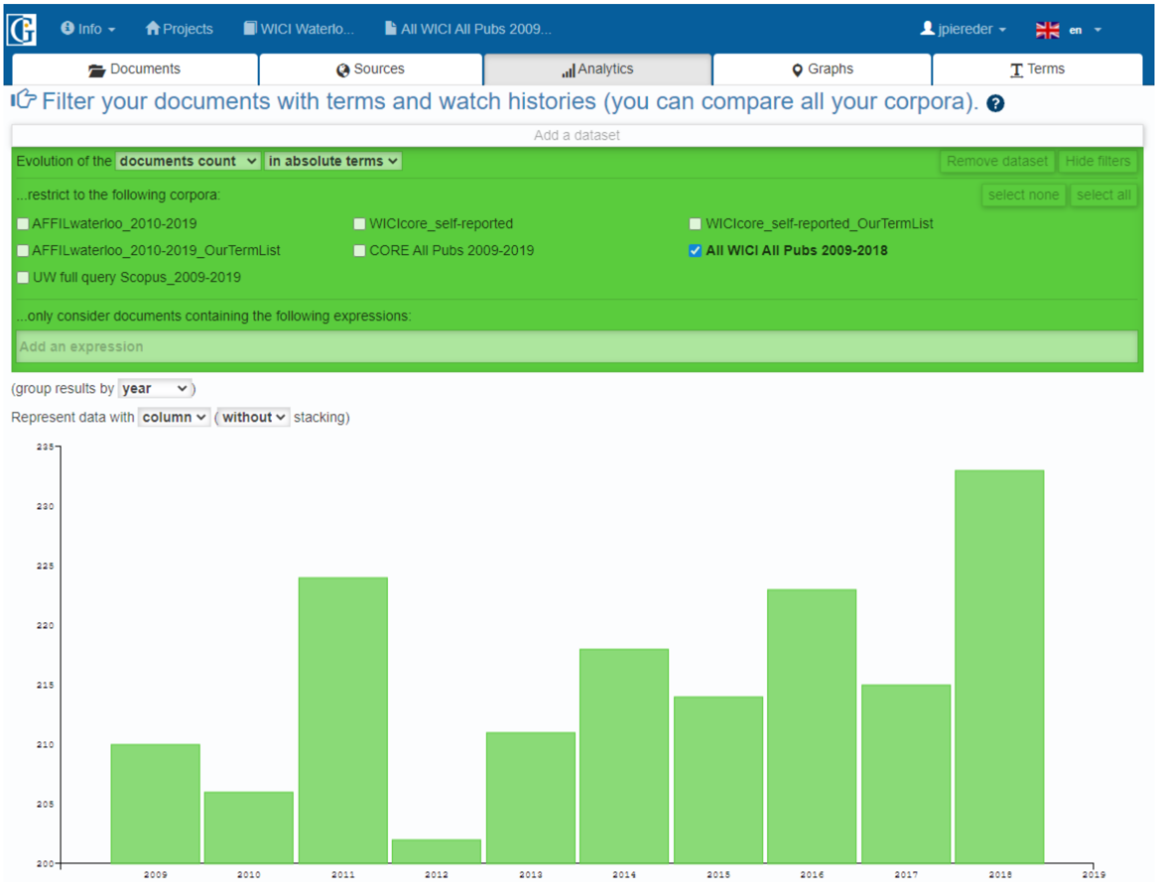
After the analysis is finished, users can take full advantage of the platform's granularity. The screenshot below shows the user interface of the map screen. On the right, the panel shows which term(s) is/are selected (hold down SHIFT to select more than one term), its "neighbour" terms, and the list of publications where the term is used. On the left, users can zoom in and out of the map, save the network data as a GEFX file, export an image, and toggle on/off the clustering algorithm. Users can also search for individual terms in the search bar at the top right. Also at the top, users can adjust the size of the nodes, the cluster colours, and limit the overall number of nodes and edges that are visible in the map.



If a user selects one of the publications listed in the right-hand panel, a "document view" screen will open (see screenshot below). Here, a user can see the full title, journal/source, authors, publication date, abstract, and the Maplist terms that are found in the abstract plus their frequency. The terms are highlighted right in the text It is also possible to add additional Maplist terms from this view. Lastly, users can “favourite” individual publications by clicking the star icon.



Finally, the “analytics” view allows users to track the history of individual or sets of terms over time, and compare across different corpora. Results can be grouped by decade, year, month, day, minute, and even down to the second (which would be more useful for social media analysis).



For more screenshots and tutorials of the software, visit https://gargantext.org/.

D) SciVal Definitions

Topics & Topic Clusters

The following definitions are excerpts from SciVal's Support Center article, "Topic Prominence in Science FAQs." For the full list of FAQs and definitions, see https://service-elsevier-com.proxy.lib.uwaterloo.ca/app/answers/detail/a_id/28428/supporthub/scival/ .

What is a topic?
"A Topic is a collection of documents with a common focused intellectual interest and can be large or small, new or old, growing or declining. Over time, new Topics will surface, and as Topics are dynamic they will evolve.

As with the nature of today's research landscape many Topics are multidisciplinary, and old Topics may be dormant, but they still exist. In addition, researchers themselves are mobile, and work in various different research areas, and thereby contribute to multiple Topics.

Scopus publications are clustered into Topics based upon a direct citation analysis**.** Where there is a weak citation link, there is a break and a new Topic is formed."

How are topics created?
"We take the entire citation network – over 1 billion citation links between 48+ million Scopus-indexed documents from 1996 forward and an additional 20+ million non-indexed documents that are cited at least twice – and break that network into roughly 96,000 Topics. A Topic is created where the linkages within the Topic are strong and the linkages outside the Topic are weak. Only the indexed documents are included in Topics."

What are Topic Clusters?
"Topic Clusters are formed by aggregating Topics with similar research interest together to form a broader, higher-level area of research. These Topic Clusters can be used to get a broader understanding of the research being done by a country, institution (or group) or researcher (or group), before drilling into the more niche underlying Topics.

Each of the 96,000 Topics have been matched with one of the 1,500 Topic Clusters. As with Topics, a researcher or institution can contribute to multiple Topic Clusters, but a Topic can only belong to one Topic Cluster and a publication can only belong to one Topic (and therefore one Topic Cluster).

Topic Clusters are formed using the same direct citation algorithm that creates the Topics. When the strength of the citation links between Topics reaches a threshold, a Topic Cluster is formed."

What do the bubbles on the wheel of science represent?

- "Each bubble represents a Topic or Topic Cluster (depending on the view you've chosen).
- The size of the bubble indicates the output of the entity in the Topic or Topic Cluster.
    - This means the same Topic can be different sizes for different entities, but positioned in the same place on the Wheel (see below).
- The position of the bubble is based upon the ASJC categories of the journals in which the Scholarly Output is published.

- The position relates to the Topic as a whole and is not affected by the entity you are looking at.
- The more influence an ASJC has over a Topic, the closer it will bring the Topic to its side of the Wheel of Science. As a result, the Topics closer to the center of the Wheel are more likely to be multidisciplinary, compared to the Topics towards the edge of the Wheel.
- Note that a Topic may be placed at the edge of the Wheel, but still be considered multidisciplinary because it is equally influenced by a number of ASJCs that are located on the same side of the Wheel."

Keyphrases

The following definition comes from SciVal's Support Center article, "How are keyphrases calculated?" See https://service-elsevier-com.proxy.lib.uwaterloo.ca/app/answers/detail/a_id/27763/supporthub/scival/.

How are keyphrases calculated?

"SciVal uses the Elsevier Fingerprint Engine to extract distinctive keyphrases within the Research Area. The Elsevier Fingerprint Engine uses text mining and applies a variety of Natural Language Processing techniques to the titles, abstracts and keywords of the documents in the Research Area, Publication Set, Topic or Topic Cluster in order to identify important keyphrases.

Keyphrases are matched against a set of thesauri spanning all major disciplines to create a list of standardized keyphrases. For each document we take the list of standardized keyphrases and select which ones are important based on Inverse Document Frequency (IDF). This technique incorporates a factor that diminishes the weight of words that occur frequently in the set of documents and increases the importance of words that occur rarely. Each keyphrase is then given a relevance between 0 and 1 with 1 given to the most frequently occurring keyphrase. Remaining keyphrases are given a value based on their relative frequency.

In SciVal we take a weighted list of keyphrases per publication and aggregate that up to different entity levels (i.e., a Research Area or Topic)."