

Analysing Scientific Collaborations of New Zealand Institutions using Scopus Bibliometric Data

Samin Aref

Department of Computer Science and
Te Pūnaha Matatini
University of Auckland, New Zealand
sare618@aucklanduni.ac.nz

David Friggens

Ministry of Business Innovation
& Employment, Wellington
New Zealand
david.friggens@mbie.govt.nz

Shaun Hendy

Department of Physics and
Te Pūnaha Matatini
University of Auckland, New Zealand
shaun.hendy@auckland.ac.nz

ABSTRACT

Scientific collaboration is one of the main enablers of development in small national science systems. Although analysing scientific collaborations is a well-established subject in scientometrics, evaluations of scientific collaborations within a country remains speculative with studies based on a limited number of fields or using data too inadequate to be representative of collaborations at a national level. This study represents a unique view on the collaborative aspect of scientific activities in New Zealand.

We perform a quantitative study based on all Scopus publications in all subjects for more than 1500 New Zealand institutions over a period of 6 years to generate an extensive mapping of scientific collaboration at a national level. The comparative results reveal the level of collaboration between New Zealand institutions and business enterprises, government institutions, higher education providers, and private not for profit organisations in 2010-2015.

Constructing a collaboration network of institutions, we observe a power-law distribution indicating that a small number of New Zealand institutions account for a large proportion of national collaborations. Network centrality concepts are deployed to identify the most influential institutions of the country in terms of collaboration. We also provide comparative results on 15 universities and Crown research institutes based on 27 subject classifications.

KEYWORDS

Big data modelling, Scientific collaboration, Scientometrics, Network analysis, Scopus, New Zealand

Reference format:

S. Aref, D. Friggens, and S. Hendy. 2017. Analysing Scientific Collaborations of New Zealand Institutions using Scopus Bibliometric Data. ArXiv preprint.

1 INTRODUCTION

There is a growing body of literature that recognises the importance of scientific collaboration in economic development [1]. The scientific collaborations can be analysed based on bibliometric data using network analysis tools and techniques

[2]. The main objective of analysing scientific collaboration is to gain an understanding of how knowledge flows between authors [3, 4], institutions [5] and countries [6, 7]. It also helps quantifying research performance measures with a focus on the collaborative aspect of research [8].

Scientific collaboration is seen not only as a performance measure, but also a representation an entity outreach and connections to other entities. Some studies focus on collaborations within a country to compare researchers or institutions and facilitate national research policy development. Perc analysed collaboration at the level of individuals in Slovenia [9] and similar study has been undertaken for Turkey [10]. Collaborations can be investigated between different countries. Park et al. investigated collaborations between China and South Korea using bibliometric data [11]. Nguyen et al. analysed collaborations of Vietnam with several other countries [12].

The university-industry collaboration has been investigated extensively as an essential connection between institutions of a science system. Abramo et al. investigated the university-industry collaboration in Italy [13] and found that university researchers collaborating with industry have a higher research performance. Investigating collaborations between specific types of institutions in a country usually requires an analysis of research outputs that represent a collaboration tie between the two types of institutions [13]. Yoon and Park investigated the collaboration between South Korea universities, industry, and government using network analysis tools and techniques on patent data [14]. The intermediate step of using network analysis to study scientific collaboration is evaluating joint outputs of authors affiliated with different types of institutions. Each bibliometric record of such nature represents a visible research connection that can be aggregated for evaluating collaboration at a national level.

It is important to point out that three types of collaborations have long existed in economic development literature and are relevant to the role of scientific collaboration; the triadic relationship between academia, industry and, the government is referred to as The Triple Helix. The term was coined by Etzkowitz and Leydesdorff [15] to refer to the shift from a dyadic industry-government relationship in an industrial society to a complex hybridisation of elements from academia, industry and government in a knowledge society. This shift is a result of innovation dynamics that support economic development.

The main contribution of this study is quantifying different types of scientific collaboration in New Zealand (NZ). This requires studying co-publications of all pairs of New Zealand

institutions to evaluate the current engagement level between them. Research collaboration among various institutions is critical to policy development as it facilitates evaluating the current state of collaboration and helps identifying capacities for improvement in different fields of research. While a few global studies exist that provide some general observations on New Zealand scientific collaboration, a specific study on New Zealand academia, government and corporations' collaboration has never been undertaken. Following the triple helix concept, we investigate collaboration among all New Zealand institutions that have a publication in a scientific database within a specified time range.

2 OVERVIEW OF NEW ZEALAND COLLABORATIONS

Scopus and Web of Science (WoS) are two competing bibliometric databases of academic publications. Scopus is owned by Elsevier and accompanied by SciVal¹ an analytics service for Scopus data. WoS is maintained by Clarivate Analytics (formerly Thomson Reuters).

In what follows, basic results from Scopus and WoS that are related to New Zealand in comparison to small advanced economies initiative countries² are discussed:

1. University-Industry Research Connections (UIRC) series of studies [16] based on the WoS data
2. An implementation of a snowball metric [17] in SciVal based on Scopus data

2.1 University-Industry Research Connections Report 2014

The most recent UIRC report [16] uses University-Industry Co-publication (UIC) as an indicator of scientific performance in various countries. Using 2009-2012 WoS data, the report shows that three New Zealand universities (University of Auckland, University of Canterbury, and Massey University) have an average level of overall UIC while the other two investigated (Victoria University of Wellington and University of Otago) have a medium-low level of overall UIC. These evaluations are based on a scale of high, medium-high, average, medium-low, and low UIC in comparison with other universities listed in 2014 edition of Leiden University Ranking³. The overall scores of the five New Zealand universities across seven disciplines can be found in UIRC 2014 report [16]. The report also demonstrates that New Zealand has an average overall UIC score which is above small advanced economies initiative countries like Singapore and Israel, but below Denmark, Switzerland, Finland and Ireland. Note that, UIRC 2014 does not include Crown Research Institutes (CRIs), which are government owned research laboratories accounting

for a considerable amount of New Zealand scientific collaborations. Methodological details of UIRC 2014 can be found in a study of public-private collaboration [18].

2.2 Academic-Corporate Collaboration Snowball metric in SciVal

SciVal is a bibliometric analysis service based on the Scopus data providing research performance of 7500 research institutions worldwide (at the time of access). SciVal implements the Snowball Metrics [17]; The Academic-Corporate Collaboration Snowball metric is calculated based on Scopus data in 2011-2016. Numerical values for 15 New Zealand universities and CRIs range from 0.4% to 1.8% suggesting that they are comparably active in collaboration with corporations. Note that SciVal's institutional mapping and classification is not complete. It has better coverage of the larger organisations and misses many small ones. Notably it only has one commercial NZ institution (Fonterra). The overall Academic-Corporate Collaboration metric for New Zealand is the lowest among small advanced economies initiative countries.

3 MATERIALS AND METHODS

In order to quantify New Zealand research collaborations at a national level, we consider all New Zealand institutions that have a publication in Scopus within the six-year time window of 2010-2015. We standardise and classify thousands records of collaborations based on 2010-2015 publications to a list of institution pairs and their reciprocal number of joint publications.

We used a full extract of Scopus [19] limiting to NZ publications between 2010 and 2015. This was combined with Ministry of Business Innovation & Employment (MBIE) internally developed mapping of Scopus Affiliation IDs to NZ institutions. Scopus uses an automated process to cluster unstructured affiliation text and apply its internal Affiliation IDs, but this is conservative and results in multiple IDs per institution. MBIE's manually compiled mapping groups together all Affiliation IDs for each NZ institution (for example there are 81 Affiliation IDs for University of Auckland) and assigns a category (see below). Some affiliations are missing or are not identifiable as specific institutions (eg independent researchers), and in some cases there are data errors incorrectly identifying an affiliation as being from NZ; these were excluded. Data analysis was performed using the R language. VisNetwork and Gephi are used for network visualisations.

While UIRC 2014 contains 5 NZ universities and there are only 37 NZ institutions listed in SciVal (at the time of access), our study comprises of over 1500 New Zealand institutions.

¹ SciVal Elsevier Research Intelligence Solution www.scival.com (accessed on 17/02/2017)

² Advanced economies with 5-10 million population including Denmark, Finland, Ireland, Israel, New Zealand, Singapore, and Switzerland

³ CWTS Leiden Ranking 2014 leidenranking.com/ranking/2014 (accessed on 17/02/2017)

This study mainly focuses on collaboration measures of 15 New Zealand universities and CRIs listed below:

1. AgResearch (New Zealand Pastoral Agriculture Research Institute)
2. AUT University (Auckland University of Technology)
3. ESR (Environmental Sciences Research)
4. GNS Science (the Institute of Geological and Nuclear Sciences)
5. Landcare Research
6. Lincoln University
7. Massey University
8. NIWA (National Institute of Water and Atmospheric Research)
9. Plant and Food Research (New Zealand Institute for Plant and Food Research)
10. Scion (New Zealand Forest Research Institute Limited)
11. University of Auckland
12. University of Canterbury
13. University of Otago
14. University of Waikato
15. Victoria University of Wellington

We use the numbers in the list above to refer to the 15 institutions later in the paper. The key measure to be used is the collaboration record which is the number of joint publications for pairs of institutions in Scopus within the specified time range. An author with affiliations to two or more institutions is not counted as a collaboration.

We adopt the triple helix concept [15] and use a classification system with four institution classes. The classification is a mapping between Scopus affiliation IDs, standard institution names, and one of the four categories:

1. Business enterprise
2. Private not for profit (PNP)
3. Government
4. Higher education.

Business enterprises are institutions registered in New Zealand Companies Office Register [20]. Private not for profit institutions include institutions classified as building society, charitable trust, contributory mortgage broker, credit union, friendly society, incorporated society, industrial & provident society, limited partnerships, other bodies, overseas issuer, participatory security, retirement villages, superannuation scheme, or unit trust in New Zealand Companies Office other Registers [21]. Government class comprises of Crown Research Institutes, central government institutions, local government institutions, other government institutions, schools, public hospitals, and district health boards. Finally, the higher education class includes private training establishments, universities, polytechnics, institutes of technology, independent research organisations, and education providers classified as wānanga.

For pairs of collaborating institutions, the collaboration records measure equals total number of joint publications. However, when collaboration records are aggregated for a specific institution, this equality does not hold. Consider a publication that has AUT, ESR, and GNS as the affiliations of its three distinct authors. Such a publication will be counted three times: once as a collaboration record of AUT-ESR, once for ESR-GNS, and once for AUT-GNS. Recalling that CRIs and

universities belong to government and higher education categories respectively, when we aggregate the collaborations of AUT with the government sector, the publication contributes two to the collaboration records between AUT and the government sector.

4 RESULTS AND DISCUSSION

4.1 General Network Properties

We construct a scientific collaboration network in which nodes are institutions and edges represent collaborations between them. Four institution categories define node types and the collaboration records values are used as weights on the edges. In this section, we discuss overall network properties including size, order, and degrees. The network has roughly 1500 nodes and 4200 edges (network density is 0.004). Degrees of the nodes follow a power-law distribution with many institutions having less than 30 collaborators while in the tail distribution a few institutions have hundreds of collaborators. Fig. 1 shows a visualization of the network in which size of the nodes are proportional to their weighted degrees. Eight universities and seven CRIs are shown in blue and green colour respectively.

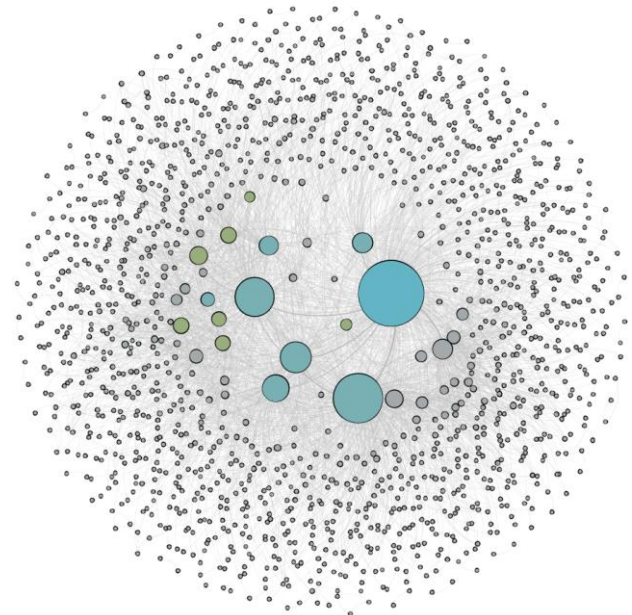


Figure 1: New Zealand scientific collaboration network

The network in Fig. 1 is made of one relatively large component as well as 21 two-node components and 2 three-node components representing 23 groups of collaborating institutions isolated from the rest of network. Like other collaboration networks [3,4], NZ collaboration network has the small world property. It has an average clustering coefficient of 0.53. For the giant component, the average path length (degree of separation) is 2.75 which is much shorter than the network diameter which is 6.

66.8% of the institutions in Fig.1 are business enterprises. Proportions for PNPs, government institutions, and higher education institutions are 18.5%, 11.1%, and 3.6% respectively.

The average unweighted degree of the network is 5.66 which represents the average number of collaborators for a given institution. The average number of collaboration records per institution (average weighted degree) is 38.13.

4.2 Ego Networks

In what follows, smaller parts of the network are illustrated. These parts are referred to as ego networks which represent the scientific collaboration network from the perspective of a specific node. Figs. 2-16 represent the ego networks of 15 universities and CRIs. Note that locations of the nodes in Figs. 2-16 are totally random except for the node in the centre which represent the main institution of the ego network.

Each ego network represents the main institution located in the centre and its collaborators as the peripheral nodes coloured respective to the classification. Thickness of the edges is proportional to the number of joint publications between the central institution and its collaborators. Business enterprises are shown in red, government institutions in green, higher education institutions in blue, and PNP organisations in purple.

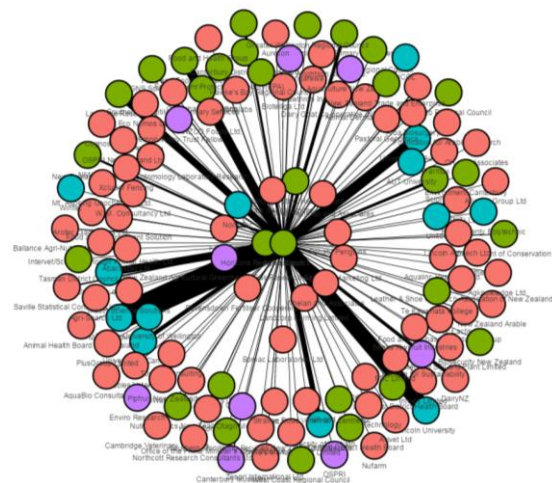


Figure 2: AgResearch Ego Network

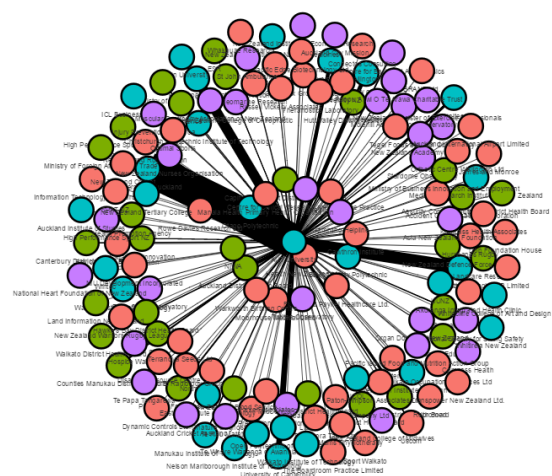


Figure 3: AUT University Ego Network



Figure 4: ESR Ego Network

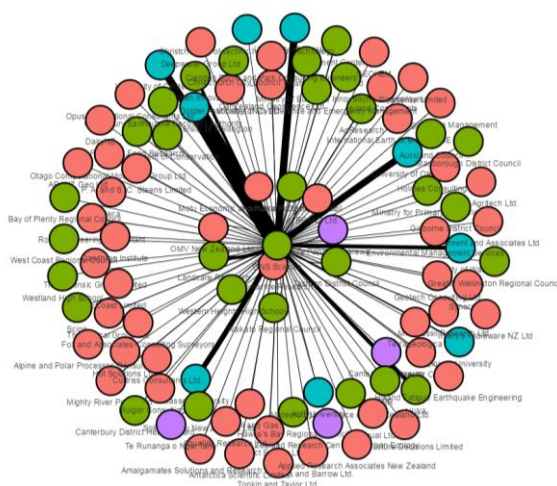


Figure 5: GNS Science Ego Network

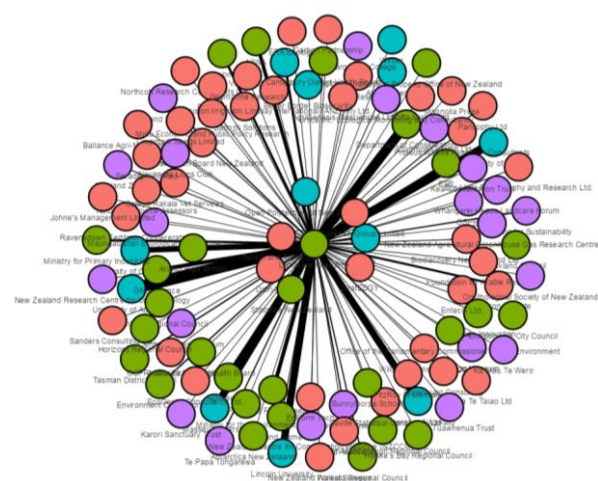


Figure 6: Landcare Research Ego Network



Figure 7: Lincoln University Ego Network

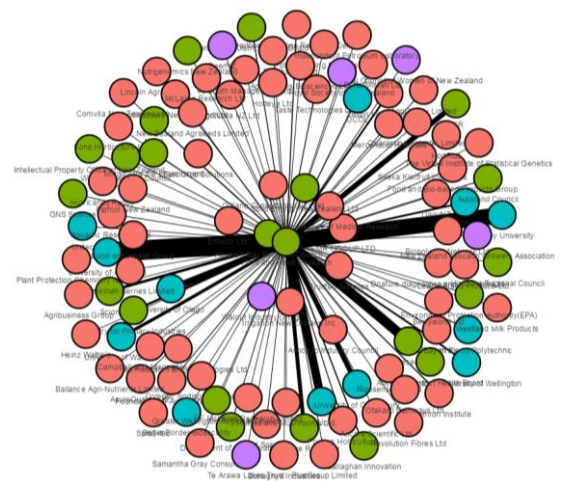


Figure 10: Plant and Food Research Ego Network

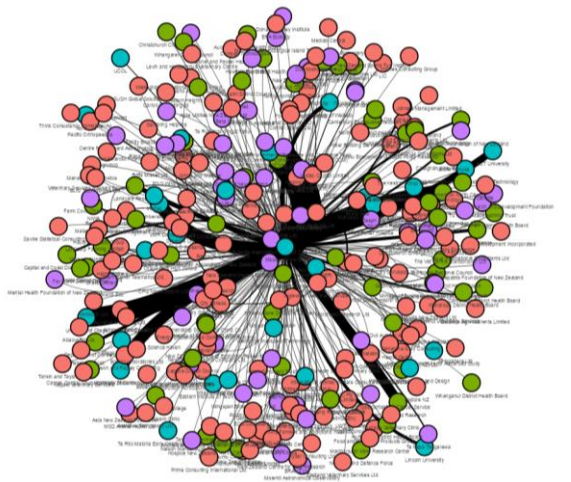


Figure 8: Massey University Ego Network

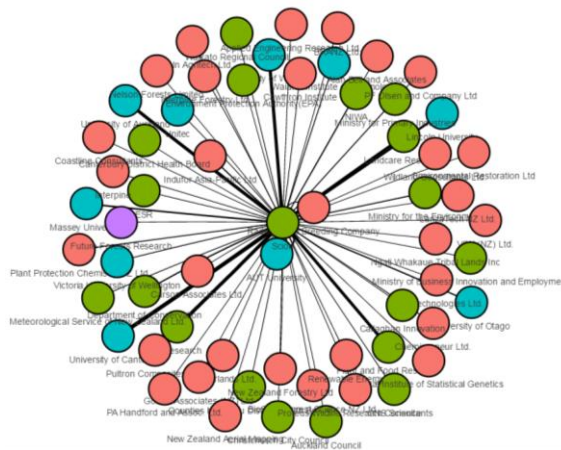


Figure 11: Scion Ego Network

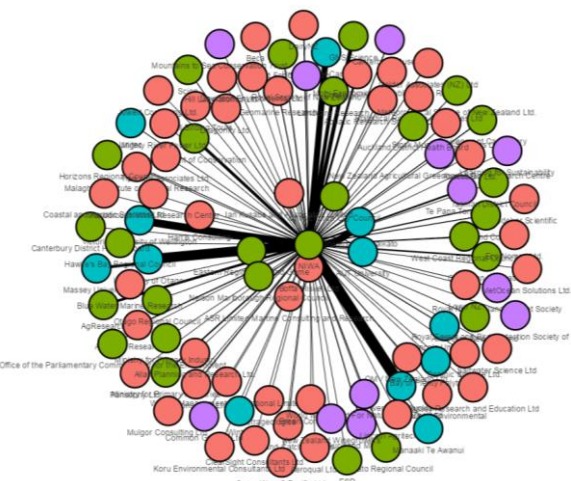


Figure 9: NIWA Ego Network

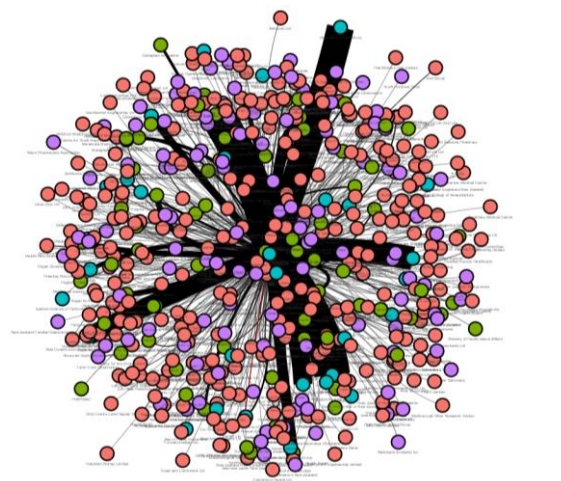


Figure 12: University of Auckland Ego Network

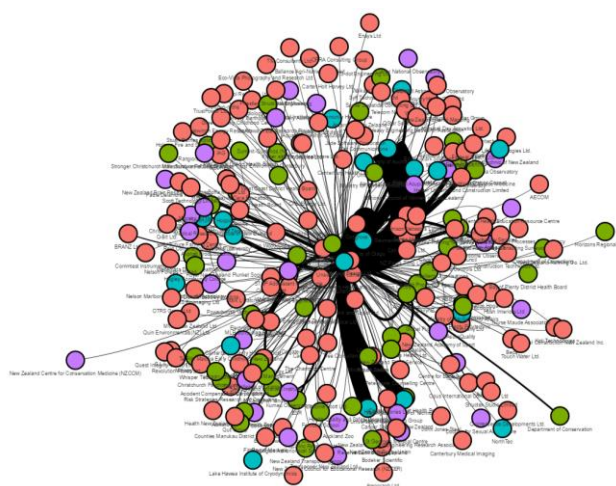


Figure 13: University of Canterbury Ego Network

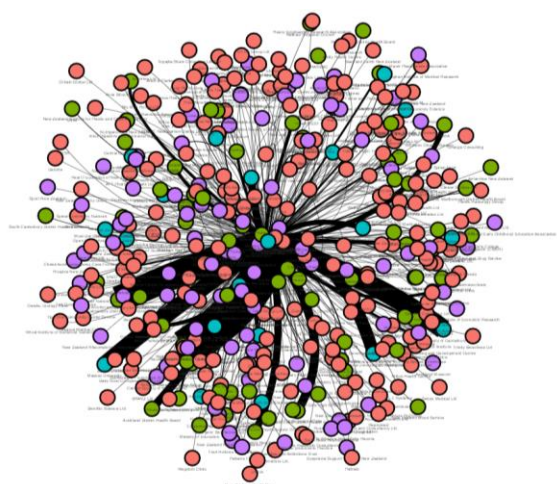


Figure 14: University of Otago Ego Network

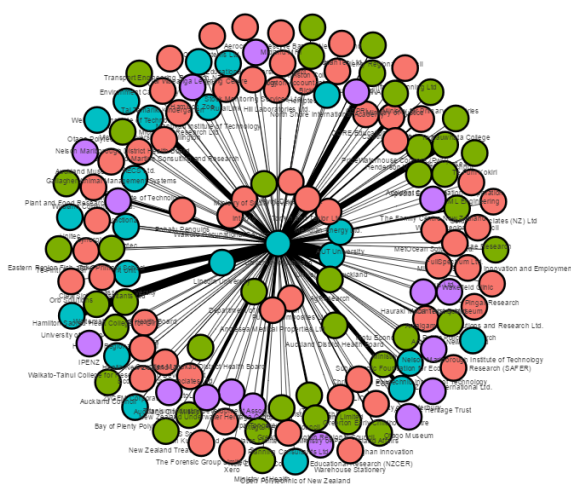


Figure 15: University of Waikato Ego Network

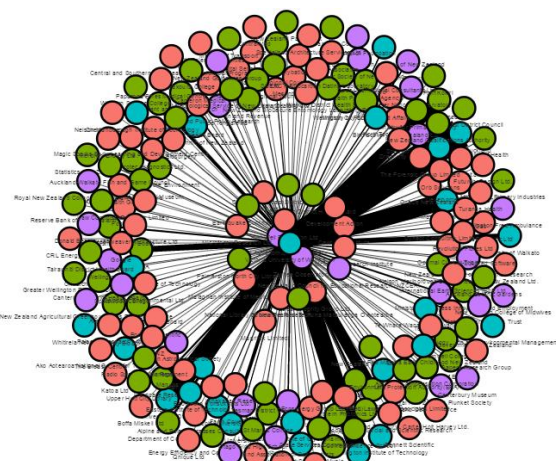


Figure 16: Victoria University of Wellington Ego Network

From the ego network in Fig. 2 we observe that AgResearch has strong collaboration ties to higher education institutions. There are many business enterprises collaborating with AgResearch as shown in Fig. 2. All four types of collaborators can be seen for AUT University featuring strong ties to other higher education institutes in Fig. 3. The ESR ego network shows strong collaboration ties to higher education institutions. There are only a few collaborating PNPs in the ego network of ESR in Fig. 4, while government category seems to account for many ESR collaborators. Like networks of AgResearch and ESR, GNS Science ego network and that of Landcare Research show strong collaboration ties to higher education institutions. Fig. 5 shows that a few PNPs collaborate with GNS Science, while many PNPs have collaboration with Landcare Research as illustrated in Fig. 6. From Fig. 7 we can see that Lincoln University collaborating institutions are less than most other NZ universities.

The network demonstrated for Massey University in Fig. 8 shows various types of collaborators. Fig. 9 indicates that NIWA collaborates with all four types of institutions and most actively with higher education institutions. There are many PNPs collaborating with NIWA compared to other CRIs. Fig. 10 demonstrates relatively many business enterprises collaborating with Plant and Food Research. Strong ties to higher education institutions are visible for Plant and Food Research ego network. The number of Scion collaborators is less than most other CRIs as evident in Fig. 11 which only shows one PNP collaborator.

From Figs. 12 - 14, we observe various types of institutions collaborating with University of Auckland, University of Canterbury, and University of Otago respectively. For University of Waikato, the number of collaborators is less than other universities as demonstrated in Fig. 15. All four types of collaborators can be seen in Fig. 16 that represents Victoria University of Wellington ego network.

4.3 Centrality Analysis of the Network

The rest of this Section provides quantitative results on two centrality analyses that determine the most central institutions of the network.

We use betweenness [22] and eigenvector centrality [23] to find the most influential nodes of the network. Betweenness centrality captures the importance of a node in a network based on its role of connecting other nodes. It measures how often a specific node appears on a path between two other nodes [22]. Eigenvector centrality contains an aggregate of a node's degree and its neighbours' degrees summed up based on a decreasing weight of distance to the neighbour [23].

Figs. 17-18 show ten most central institutions based on betweenness and eigenvector centrality measures.

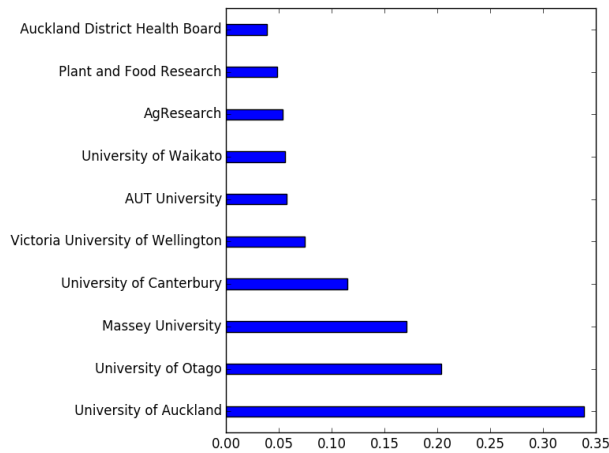


Figure 17: Top 10 central institutions among all 1500 institutions based on betweenness centrality

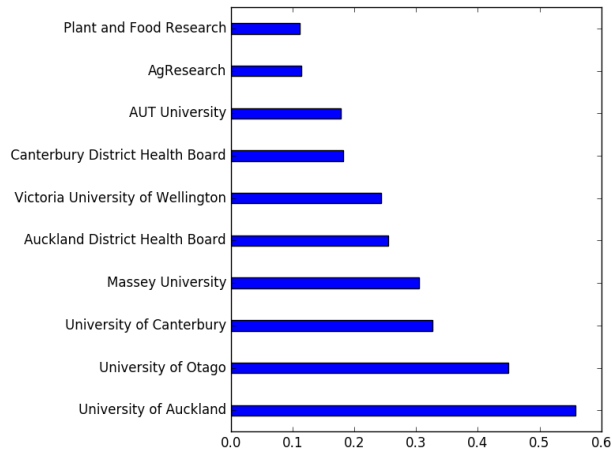


Figure 18: Top 10 central institutions among all 1500 institutions based on eigenvector centrality

The most striking observation to emerge from Figs. 17-18 is that not all NZ universities are among the 10 most central institutions in terms of collaboration. Instead some district health boards and CRIs can be seen in Figs. 17-18. Recall that in Fig. 1, the eight universities were not the largest nodes of the network.

5 COLLABORATION RATIOS

In this Section, we analyse the ratios of collaboration for each of the 15 universities and CRIs. This section can be considered as an analysis of weighted degrees of the nodes.

The collaboration ratios are provided as proportions in Fig. 19 and total counts in Fig. 20. Numbers on the vertical axis refer to the 15 universities and CRIs as listed in Section 3. Purple, blue, green, and red colours in Figs. 19-20 represent collaboration with PNPs, higher education providers, government institutions, and business enterprises respectively.

Three faceted plots based on All Science Journal Classification (ASJC) are provided in the appendix. They can be used for comparing scientific collaborations in different fields of research as well as comparing universities and CRIs based on their collaboration records in each ASJC subject.

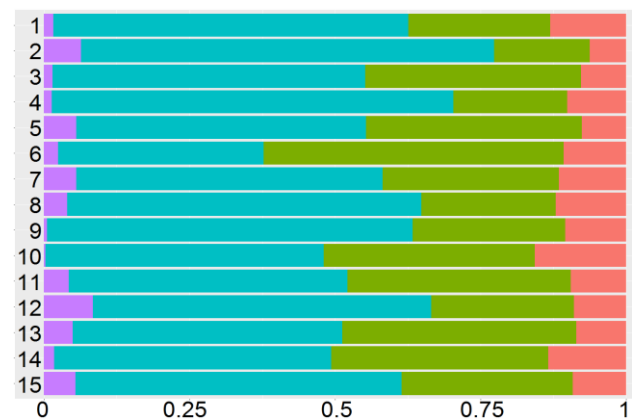


Figure 19: Collaboration records proportions for 15 universities and CRIs

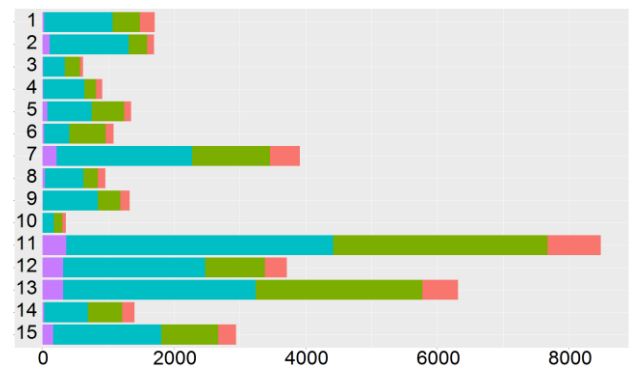


Figure 20: Collaboration records counts for 15 universities and CRIs

A comparison of relative collaborations with the four institution categories can be made based on Fig. 19. Lincoln university (6) has the highest proportion of collaboration with government. Regarding collaboration with business enterprises, Scion (10) has the highest proportion. University of Canterbury (12) has the highest proportion of collaboration with PNPs.

Fig. 20 shows that University of Auckland has the highest collaboration count followed by University of Otago, Massey University, University of Canterbury, and Victoria University of Wellington. This order corresponds to the five largest nodes in

Fig. 1 (recall that, the collaboration record counts equal weighted degrees of the nodes).

6 CONCLUSIONS

This study investigated research collaborations among New Zealand institutions based on all Scopus publications from 2010 to 2015. Research collaborations were quantified by numerical measures based on joint publications of New Zealand institutions.

We have considered four classifications for the institutions, namely higher education, government, business enterprise, and private not for profit. The raw data containing thousands of Scopus affiliation IDs was categorised into standard institution names and classes defining nodes of a scientific collaboration network where collaborations are represented by weighted edges. The network is unique in its representative capability for research collaborations at a national level.

The centrality analysis, demonstrated in Figs. 17-18, indicated the most central institutions in terms of scientific collaboration. The quantitative results on collaboration records, illustrated in Figs. 19-20, shed light on collaborations between New Zealand universities/CRI and different types of institutions at a national level. We have also used ASJC subjects to analyse research collaborations in 27 different fields of research whose comparative results can be found in the Appendix.

While we focused on 15 universities and CRIs in New Zealand, the analysis was performed for over 1500 New Zealand institutions comprising of business enterprises, charitable trusts, union trusts, incorporated societies, and limited partnerships registered in New Zealand companies office register as well as central and local government institutions, schools, district health boards, private training establishments, polytechnics, institutes of technology, and independent research organisations.

This research has opened many avenues to be explored by more in-depth analysis on New Zealand bibliometric data. For one research direction, the analysis can be extended allowing for measures of research quality to play a role in evaluating collaborations. Field-normalised citation based measures [24] might be suitable candidates to be used as measures of research quality. Evaluating the potential improvement capacities across different disciplines would be another direction that can be taken from a policy development perspective. Observing a few institutions accounting for a large proportion of collaboration and most of the results confirming intuitive expectations, a third recommendation for future work is incorporating a measure of institution size to get the relevant measures per capita and use them for a better comparison of research collaboration performance of the institutions.

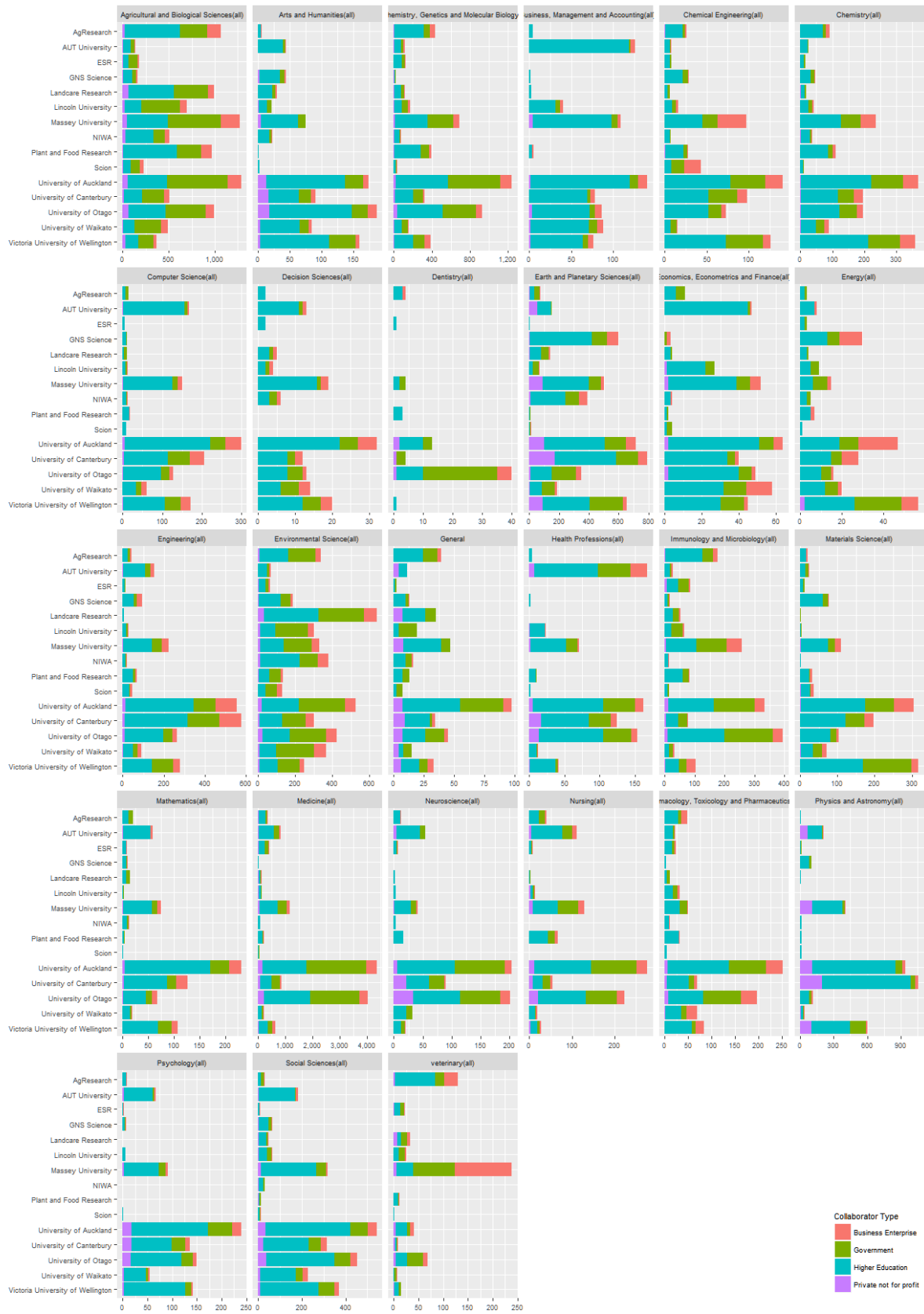
ACKNOWLEDGMENTS

Andrew Marriott and Sam Holmes at Ministry of Business Innovation & Employment performed much of the work classifying New Zealand institutions. This work was partially supported by the Te Pūnaha Matatini 2016-2017 postgraduate internship program at Ministry of Business, Innovation & Employment.

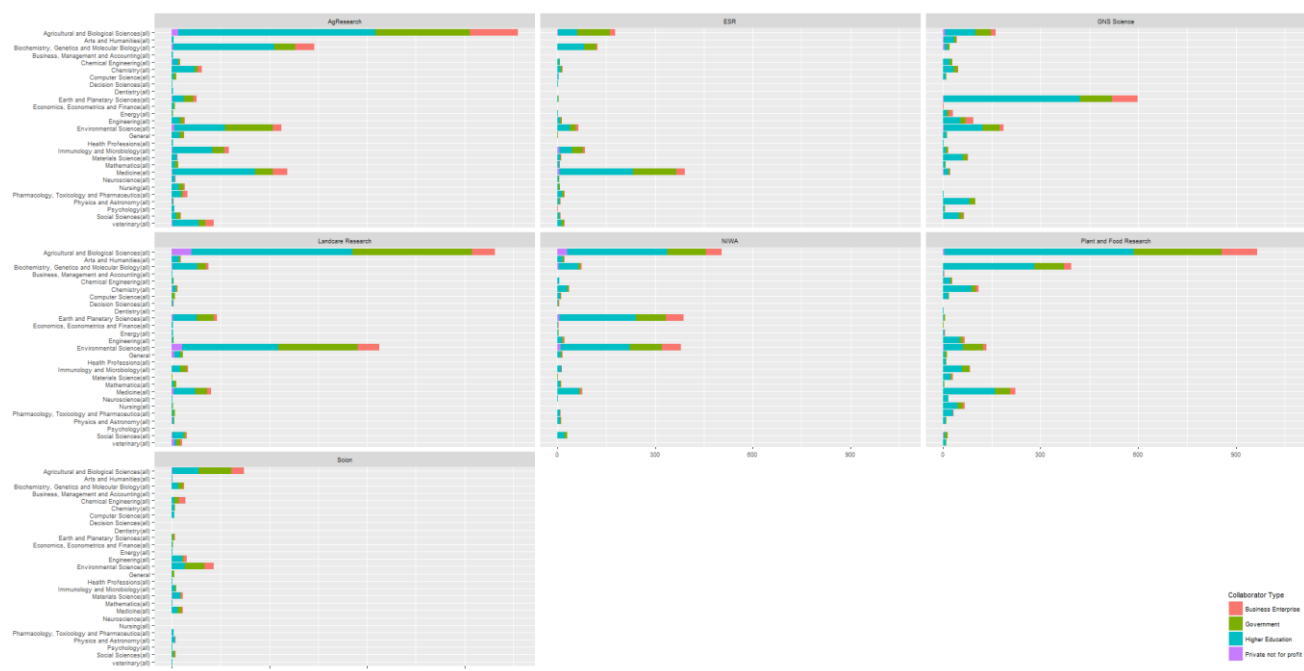
REFERENCES

- [1] Sonnenwald, D. H. Scientific collaboration. *Annual review of information science and technology*, 41, 1 (2007), 643-681.
- [2] Fung, H.-N. and Wong, C.-Y. Scientific collaboration in indigenous knowledge in context: Insights from publication and co-publication network analysis. *Technological Forecasting and Social Change*, 117 (2017), 57-69.
- [3] Newman, M. E. The structure of scientific collaboration networks. *Proceedings of the National Academy of Sciences*, 98, 2 (2001), 404-409.
- [4] Barabási, A. L., Jeong, H., Neda, Z., Ravasz, E., Schubert, A. and Vicsek, T. Evolution of the social network of scientific collaborations. *Physica A: Statistical Mechanics and its Applications*, 311, 3-4 (2002), 590-614.
- [5] Katz, J. Geographical proximity and scientific collaboration. *Scientometrics*, 31, 1 (1994), 31-43.
- [6] Gazni, A., Sugimoto, C. R. and Didegah, F. Mapping world scientific collaboration: Authors, institutions, and countries. *Journal of the American Society for Information Science and Technology*, 63, 2 (2012), 323-335.
- [7] Luukkonen, T., Tijssen, R., Persson, O. and Sivertsen, G. The measurement of international scientific collaboration. *Scientometrics*, 28, 1 (1993), 15-36.
- [8] Abbasi, A. and Altmann, J. *On the Correlation between Research Performance and Social Network Analysis Measures Applied to Research Collaboration Networks*. City, 2011.
- [9] Perc, M. Growth and structure of Slovenia's scientific collaboration network. *Journal of Informetrics*, 4, 4 (2010), 475-482.
- [10] Çavuşoğlu, A. and Türker, İ. Scientific collaboration network of Turkey. *Chaos, Solitons & Fractals*, 57 (2013), 9-18.
- [11] Park, H. W., Yoon, J. and Leydesdorff, L. The normalization of co-authorship networks in the bibliometric evaluation: the government stimulation programs of China and Korea. *Scientometrics*, 109, 2 (2016), 1017-1036.
- [12] Nguyen, T. V., Ho-Le, T. P. and Le, U. V. International collaboration in scientific research in Vietnam: an analysis of patterns and impact. *Scientometrics*, 110, 2 (2017), 1035-1051.
- [13] Abramo, G., D'Angelo, C. A., Di Costa, F. and Solazzi, M. University-industry collaboration in Italy: A bibliometric examination. *Technovation*, 29, 6-7 (2009), 498-507.
- [14] Yoon, J. and Park, H. W. Triple helix dynamics of South Korea's innovation system: a network analysis of inter-regional technological collaborations. *Quality & Quantity*, 51, 3 (2017), 989-1007.
- [15] Etzkowitz, H. and Leydesdorff, L. The Triple Helix--University-industry-government relations: A laboratory for knowledge based economic development (1995).
- [16] Leiden University CWTS *University-Industry Research Connections*, 2014. www.cwts.nl/UIRC2014 (accessed on 17/02/2017)
- [17] Colledge, L. *Snowball metrics recipe book*. Amsterdam: *Snowball Metrics Program Partners* (2014).
- [18] Tijssen, R. J. Co-authored research publications and strategic analysis of public-private collaboration. *Research Evaluation* (2012).
- [19] Elsevier. Scopus Custom Data extracted June 2016.
- [20] NZCO *New Zealand Companies Office Register*. New Zealand, 2017.
- [21] NZCO *New Zealand Companies Office Other Registers*. New Zealand, 2017.
- [22] Freeman, L. C. Centrality in social networks conceptual clarification. *Social networks*, 1, 3 (1978), 215-239.
- [23] Bonacich, P. Factoring and weighting approaches to status scores and clique identification. *Journal of Mathematical Sociology*, 2, 1 (1972), 113-120.
- [24] Waltman, L. A review of the literature on citation impact indicators. *Journal of Informetrics*, 10, 2 (2016), 365-391.

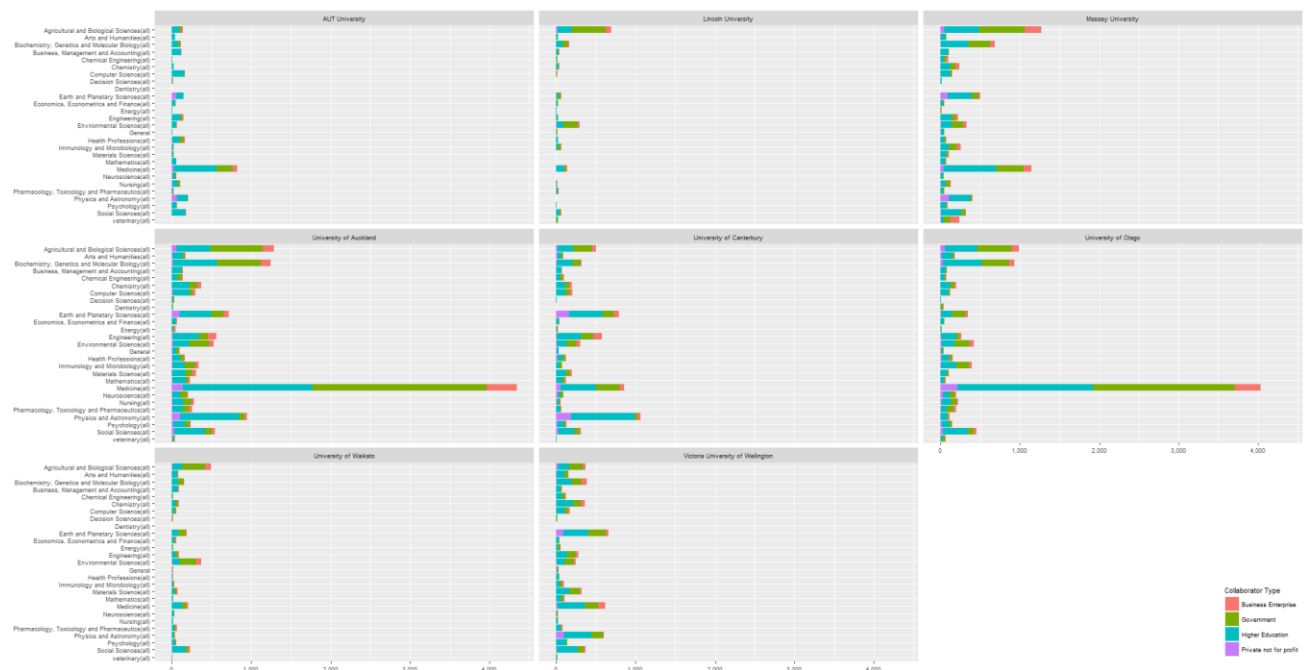
A APPENDIX



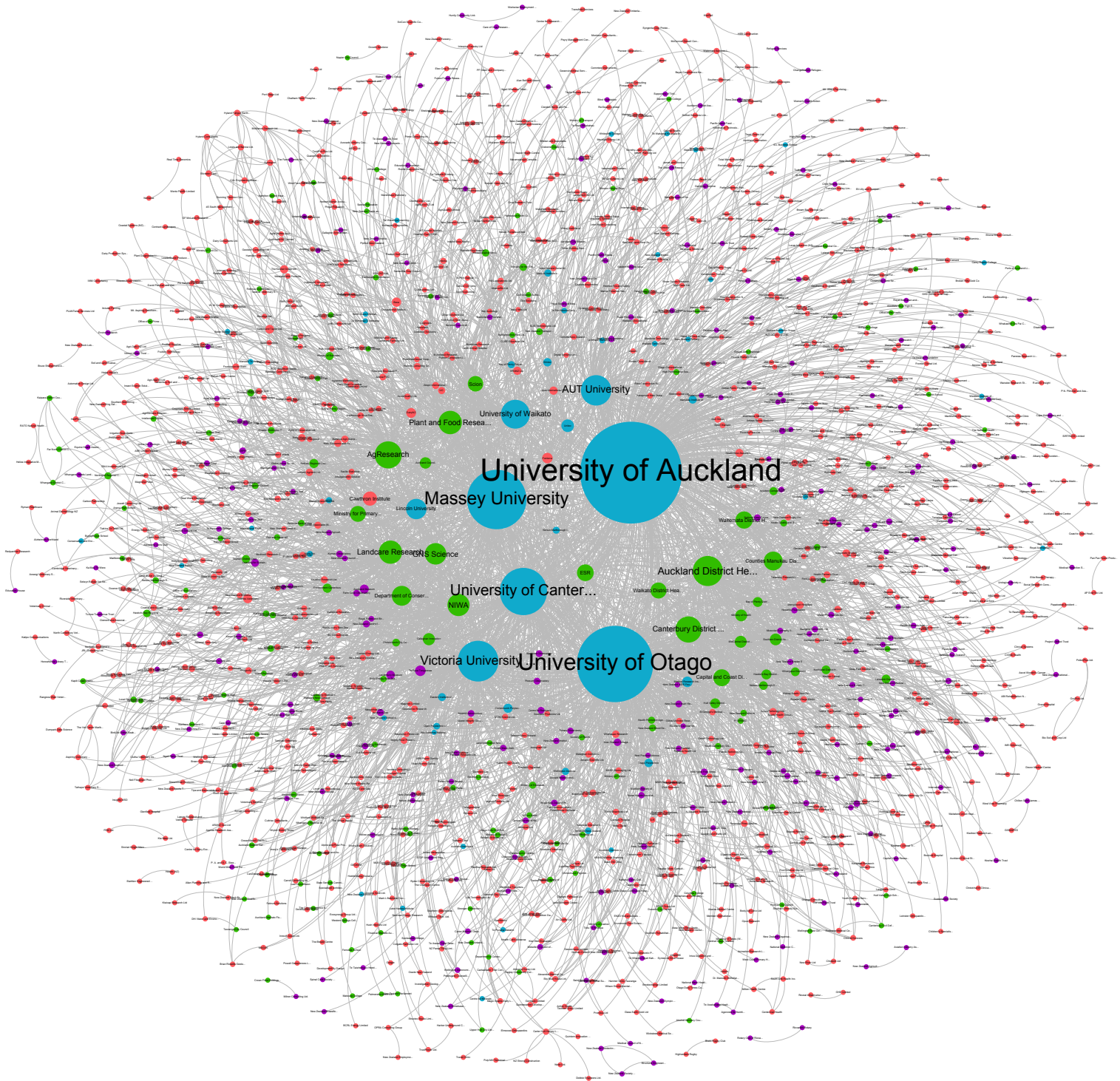
Collaboration records count within New Zealand for each ASJC fields



Comparison of collaboration records of 7 Crown Research Institutes based on ASJC field



Comparison of collaboration records of 8 New Zealand universities based on ASJC field



New Zealand Scientific collaboration network visualised as a weighted network