Scientific Collaboration through the lens of Social Network Analysis

A case study using the 2015 Canadian Aquaculture R&D Review

C. Mimeault and M. Williamson
Ottawa Graph Meetup

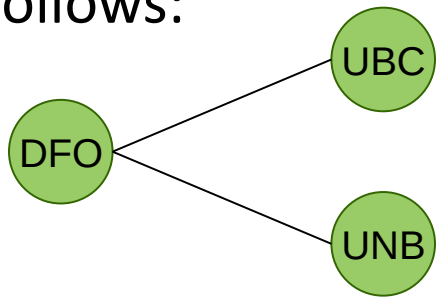
July 2017
(tentative)

Scientific collaboration

- Science advice relies on knowledge generated by a large community of researchers
- New investments are being used to increase scientific collaboration to foster and leverage important partnerships
- Increasing collaboration represents an intervention in an existing community which can be data driven and informed by science

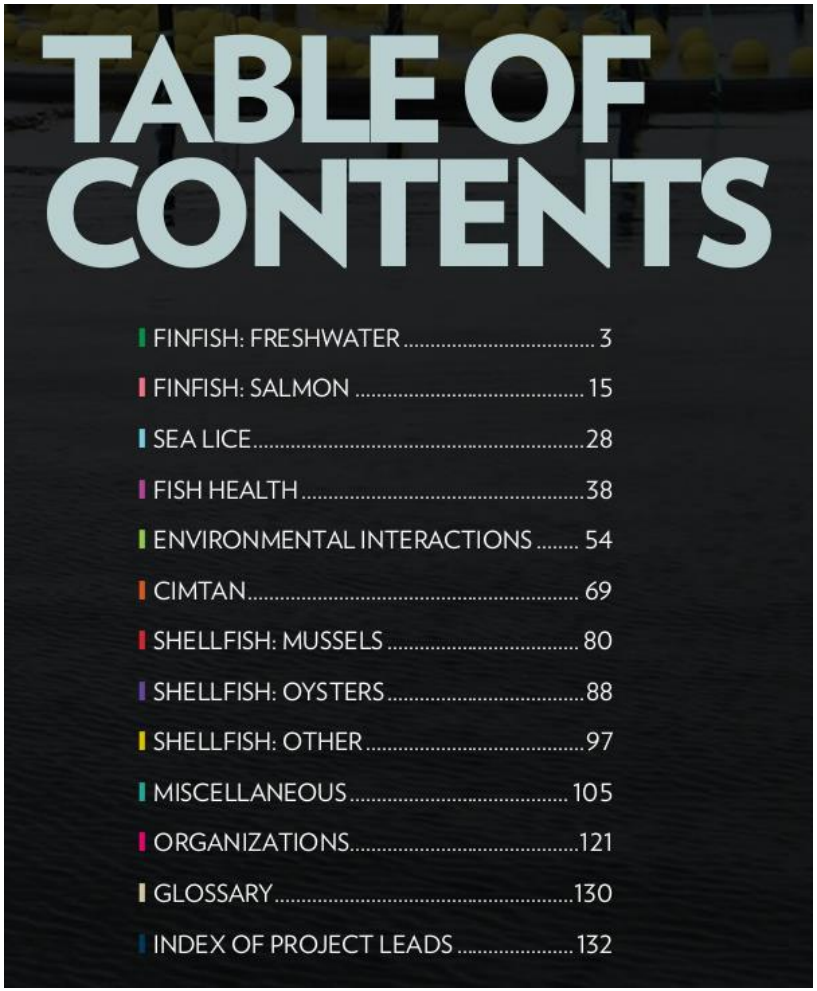
Social Network Analysis

- Science that studies the structure of communities providing tools to describe and measure collaboration
- A network is a collection of points (nodes) joined together in pairs by lines (edges) (Newman, 2010)
- For example, applying those concepts, collaboration between DFO, UBC and UNB could be modelled as follows:



- Conceptualizing available data in the form of a network allows to apply multiple social network analytical tools

Case Study



http://www.dfo-mpo.gc.ca/aquaculture/sci-res/rd-eng.htm

Aquaculture R&D Reviews

- Every two years since 2007, the Aquaculture Association of Canada publishes, in partnership with Fisheries and Oceans Canada, a compendium of the on-going aquaculture research and development projects in Canada
- Projects include governmental, academic, industry and non-governmental organizations
- The Aquaculture R&D Reviews provide an overview of the Canadian aquaculture research community over the last decade
- Extracting project details from those documents creates a dataset amenable to a variety of analytical techniques including descriptive statistics and social network analysis

Methodology

- Collaboration was defined as two or more organizations either :
  - working on a given project, i.e. listed as project lead, project team or collaborators; or
  - funding a given project, i.e., listed as funders or co-funders
- For each project included in the 2015 Aquaculture R&D Review, specific variables (category, title, funded by, co-funded by, project lead, project team, collaborators) were extracted into a structured, consistent and parsable data format (yaml)
- A series of Python scripts were created to generate csv files from the yaml files for all collaborative projects
- csv files were imported in Gephi, a Social Network Analysis software, for visualisations and analyses

Example of a yaml file


```
- title: "OPTIMIZATION OF CULTURED WALLEYE (SANDER VITREUS) EGG QUALITY"
  year: "2015"
  category: ["Freshwater finfish"]
  species: ["Walleye"]
  period: "SEP. 2011–APR. 2013"
  funded_by: [{"org":"Société de Recherche et de Développement en Aquaculture Continentale Inc.",
      "program":"undefined"}, {"org":"Ressources Aquatiques Québec", "program":"undefined"}]
  cofunded_by: [{"org":"Fonds de Recherche Nature et Technologies", "program":"undefined"},
      {"org":"BMP Innovation scholarship", "program":"undefined"}]
  project_lead: [{"name":"Réjean Tremblay", "org":"Université du Québec à Rimouski"}]
  project_team: [{"name":"Céline Audet", "org":"Université du Québec à Rimouski"}, {"name":"Grant
      Vandenberg", "org":"Université Laval"},{"name":"Marco Blanchet", "org":"Station piscicole Trois-
      Lacs"}]
  collaborators: [{"name":"Ines Ben Khemis", "org":"National Institute of Marine Sciences and
      Technologies"},{"name":"Mari Moren", "org":"Nofima"}]
```


Blue: parameters related to funding
Green : parameters related to collaboration

Example of Python script

```
10 from docopt import docopt
11 import yaml
12 import csv
13 import itertools
14 from collections import Counter
15
16 if __name__ == '__main__':
17     arguments = docopt(__doc__, version='0.1')
18
19 filename = arguments["<file>"]
20 file = open(filename, 'r')
21 dataset = yaml.load(file)
22
23 filename_without_extension = filename.split(".")[0]
24 csvfile = open("{}_org_edges.csv".format(filename_without_extension), "w")
25 csv_output = csv.writer(csvfile, delimiter = ",")
26 csv_output.writerow(["SOURCE", "TARGET", "TYPE", "WEIGHT"])
27
28 def clean_up():
29     global file
30     file.close()
31     global csvfile
32     csvfile.close()
33
34 def extract_people(project):
35     return project["project_lead"] + project["project_team"] + project["collaborators"]
36
37 def extract_organizations(people):
38     # we are removing duplicate org names
39     # by converting the list to a set
40     return set([person["org"] for person in people])
```

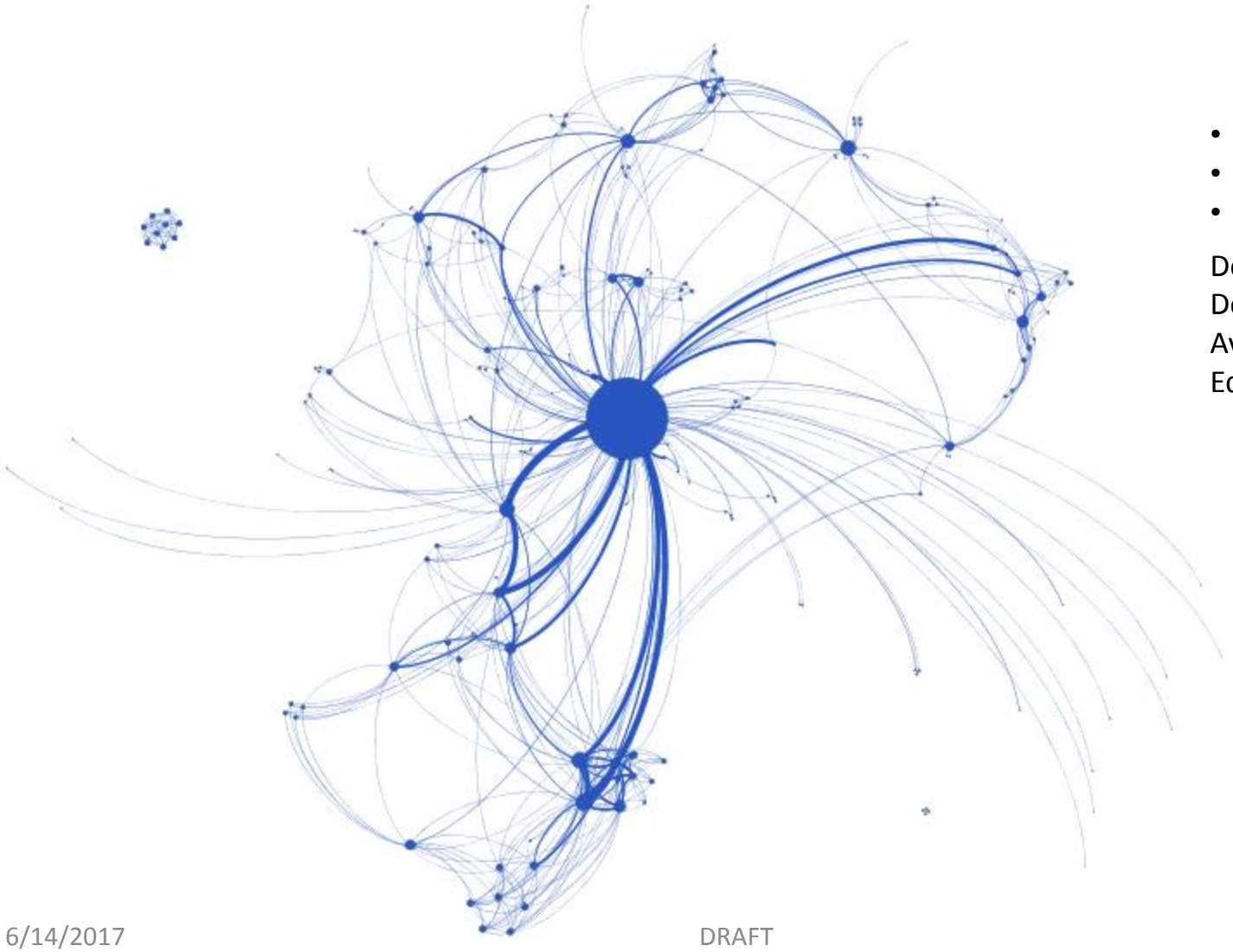
Example of a csv file

SOURCE	TARGET	TYPE	WEIGHT
Atlantic Canada Fish Farmers Association	New Brunswick Department of Agriculture, Aquaculture and Fisheries	undirected	1
Atlantic Canada Fish Farmers Association	Cooke Aquaculture Inc.	undirected	1
Genome Atlantic	Genome Canada	undirected	2
Atlantic Canada Fish Farmers Association	Huntsman Marine Science Centre	undirected	1
EWOS Innovation	Genome Atlantic	undirected	1
Genome Atlantic	National Research Council	undirected	1
Cold Ocean Salmon Inc.	Fisheries and Oceans Canada	undirected	2
Société de recherche et de développement en aquaculture continentale Inc.	Université Laval	undirected	1
Cooke Aquaculture Inc.	Fisheries and Oceans Canada	undirected	1
Cermaq Canada	Grieg Seafood	undirected	1
Ressources Aquatiques Québec	Université Laval	undirected	1
Huntsman Marine Science Centre	National Conservation Plan funding	undirected	1
Atlantic Canada Opportunities Agency	Huntsman Marine Science Centre	undirected	2
Grieg Seafood	Marine Harvest Canada	undirected	1
Huntsman Marine Science Centre	New Brunswick Department of Agriculture, Aquaculture and Fisheries	undirected	1
Fisheries and Oceans Canada	Marine Harvest Canada	undirected	1
Fish, Food and Allied Workers	Fisheries and Oceans Canada	undirected	1
Huntsman Marine Science Centre	Northern Harvest Sea Farms	undirected	2
Cooke Aquaculture Inc.	Huntsman Marine Science Centre	undirected	1
Cermaq Canada	Fisheries and Oceans Canada	undirected	1
Cold Ocean Salmon Inc.	Fish, Food and Allied Workers	undirected	1
Natural Sciences and Engineering Research Council	PEI Aquaculture and Fisheries Research Initiative Inc.	undirected	1
Cooke Aquaculture Inc.	Natural Sciences and Engineering Research Council	undirected	1
Fisheries and Oceans Canada	Ressources Aquatiques Québec	undirected	1
Ressources Aquatiques Québec	Société de recherche et de développement en aquaculture continentale	undirected	1
Genome Canada	National Research Council	undirected	1
Cooke Aquaculture Inc.	National Conservation Plan funding	undirected	1
Cooke Aquaculture Inc.	Fort Folly First Nations	undirected	1
Fort Folly First Nations	New Brunswick Department of Agriculture, Aquaculture and Fisheries	undirected	1

Results

- In 2015, a total of 215 projects were included in the Canadian Aquaculture R&D Review of which 168 (78%) were collaborative, i.e. involved two or more organisations
- In 2015, Fisheries and Oceans Canada was involved in 138 (64%) of the 215 projects of which 108 (78%) were collaborative

2015 Canadian Aquaculture R&D Community



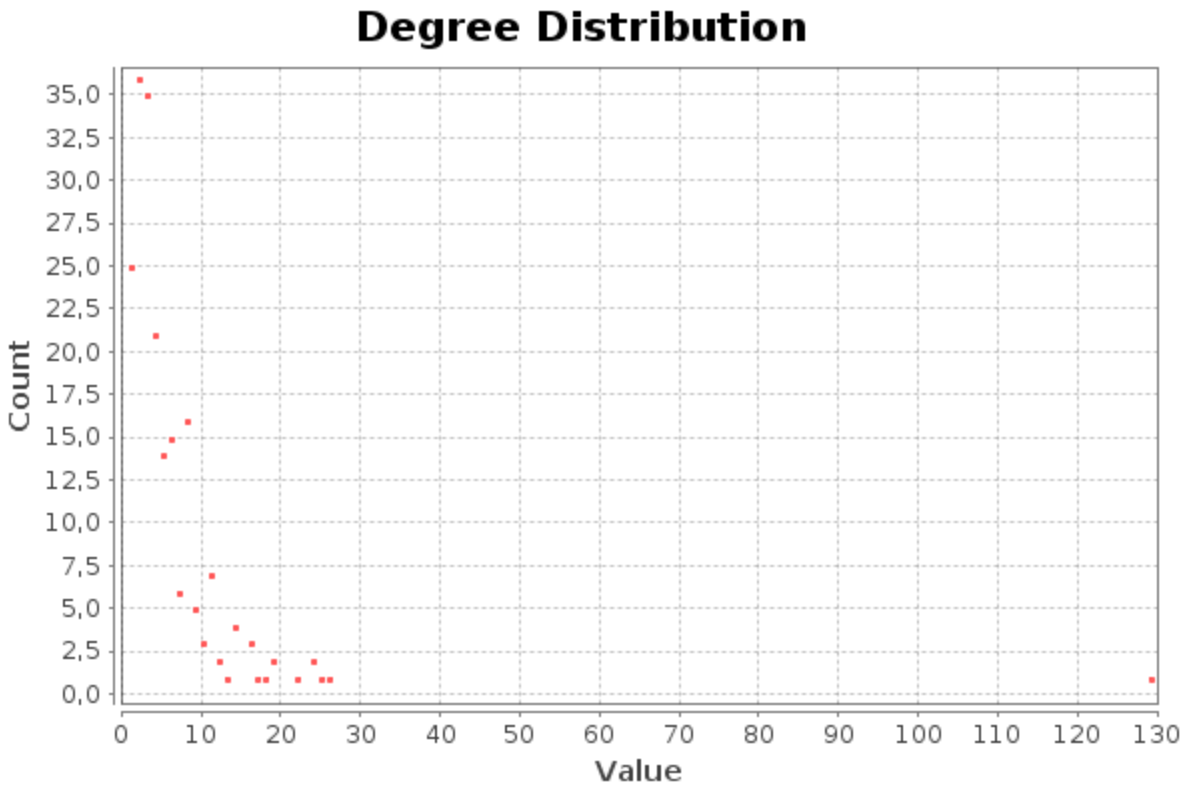
- 168 projects
- 203 nodes
- 616 edges

Density: 0.03
Degree range: 1 to 129
Average degree: 6
Edge weight: 1 to 12

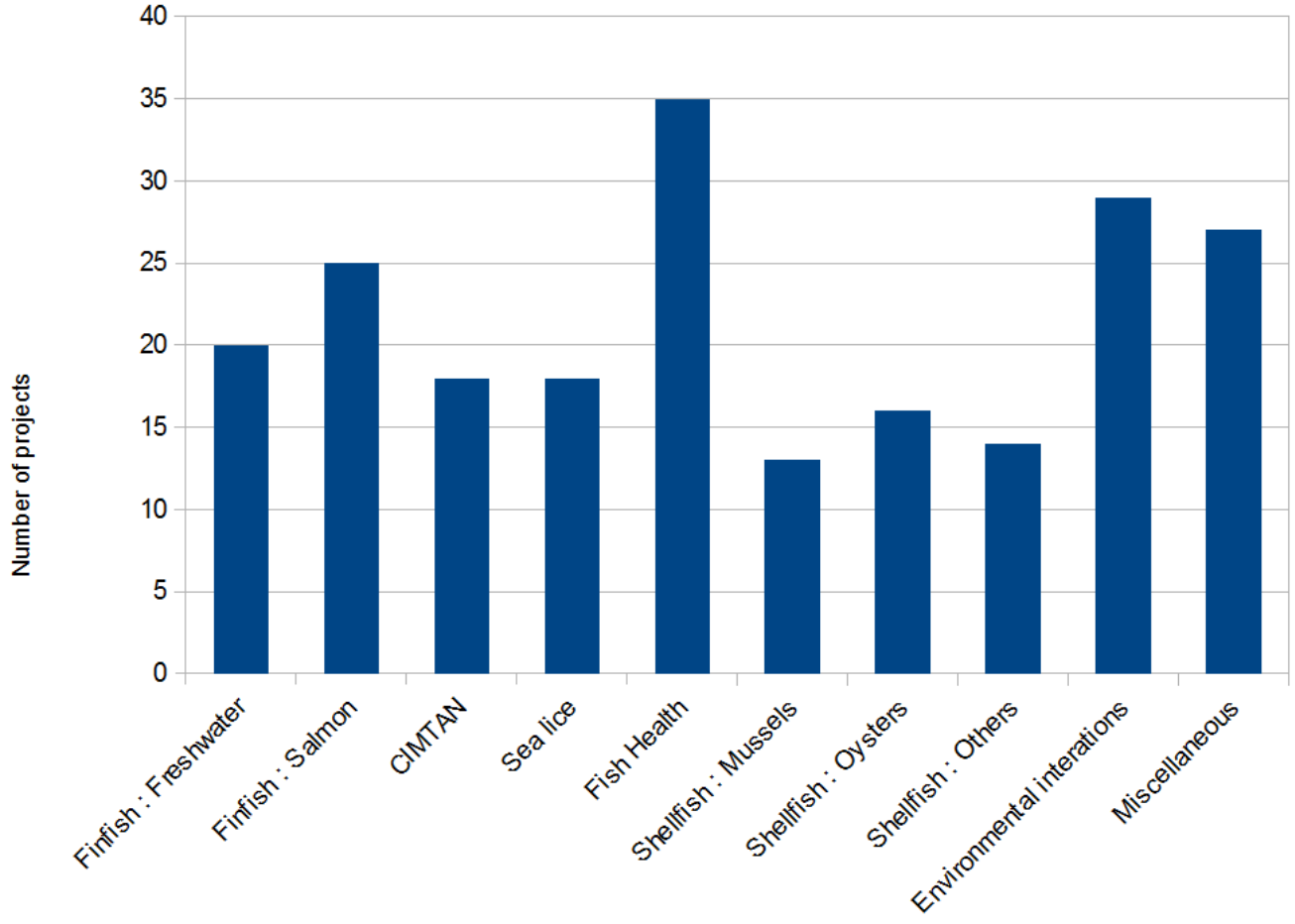
Degree Report

Results:

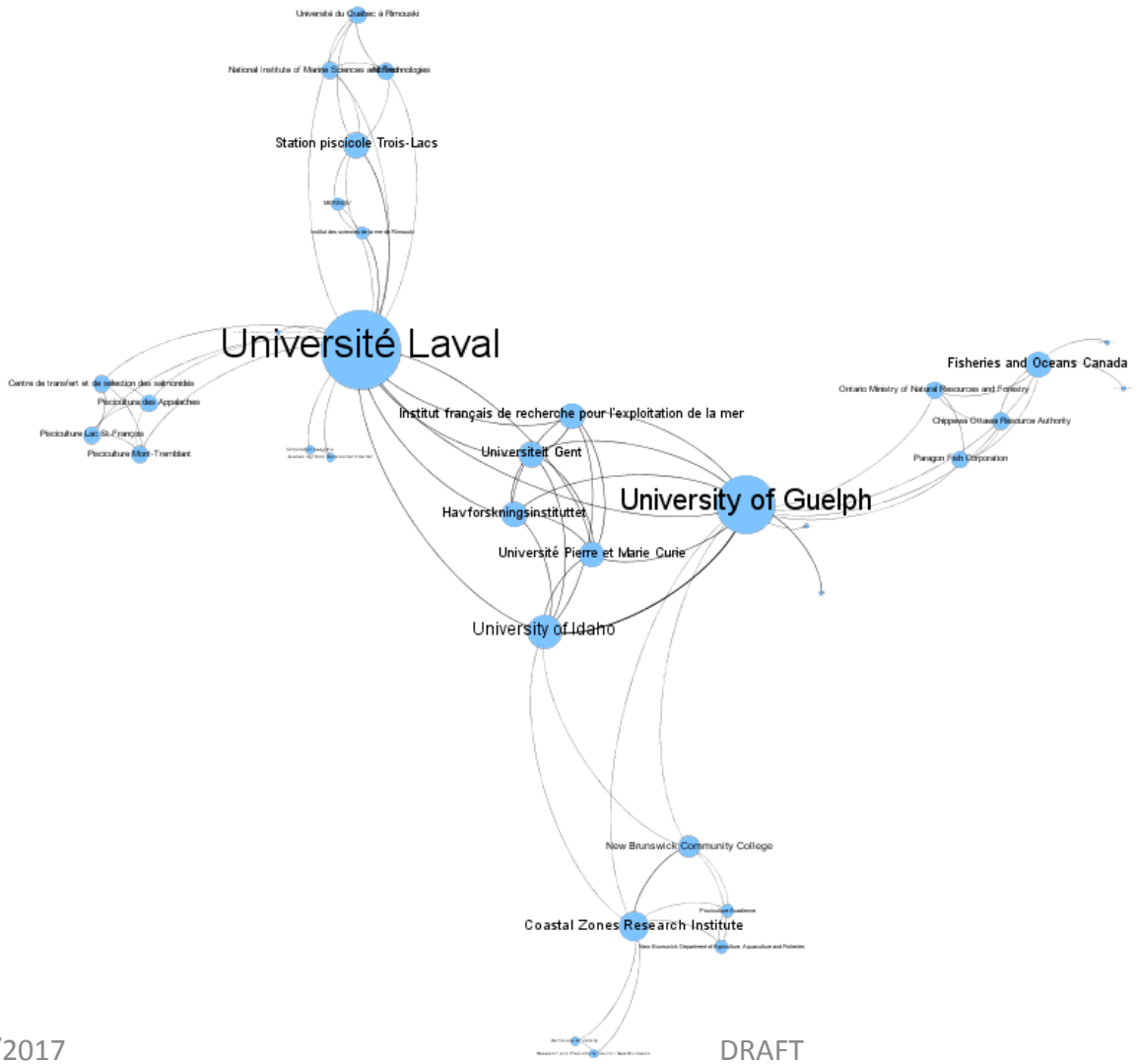
Average Degree: 6,069



2015 Aquaculture R&D Review



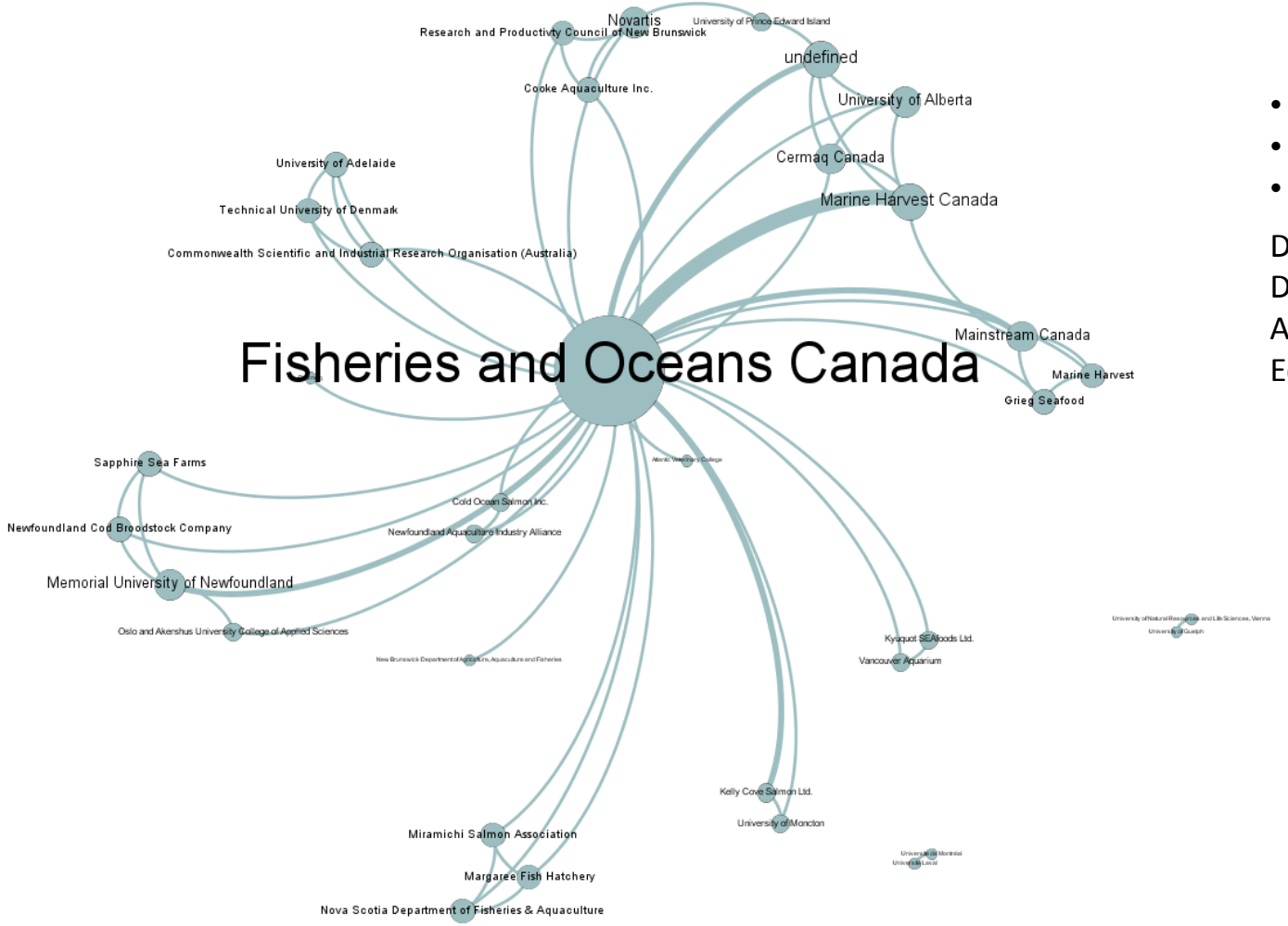
Finfish: Freshwater



- 20 projects
- 36 nodes
- 78 edges

Density: 0.12
Degree range: 1 to 19
Average degree: 4
Edge weight: 1 to 3

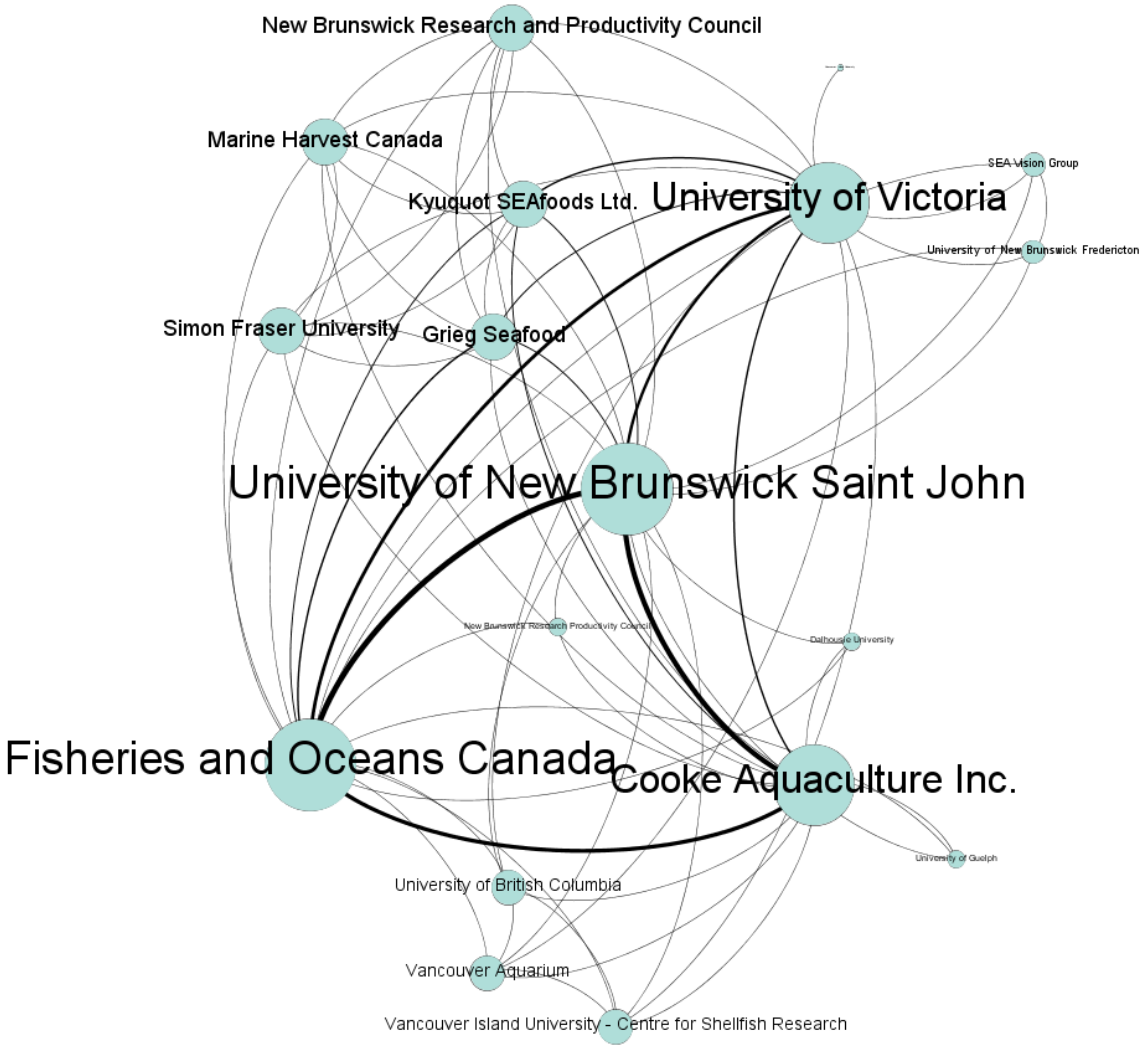
Fish Health



- 35 projects
- 35 nodes
- 59 edges

Density: 0.10
Degree range: 1 to 29
Average degree: 3
Edge weight: 1 to 5

CIMTAN



- 18 projects
- 18 nodes
- 68 edges

Density: 0.44
Degree range: 1 to 16
Average degree: 7
Edge weight: 1 to 10

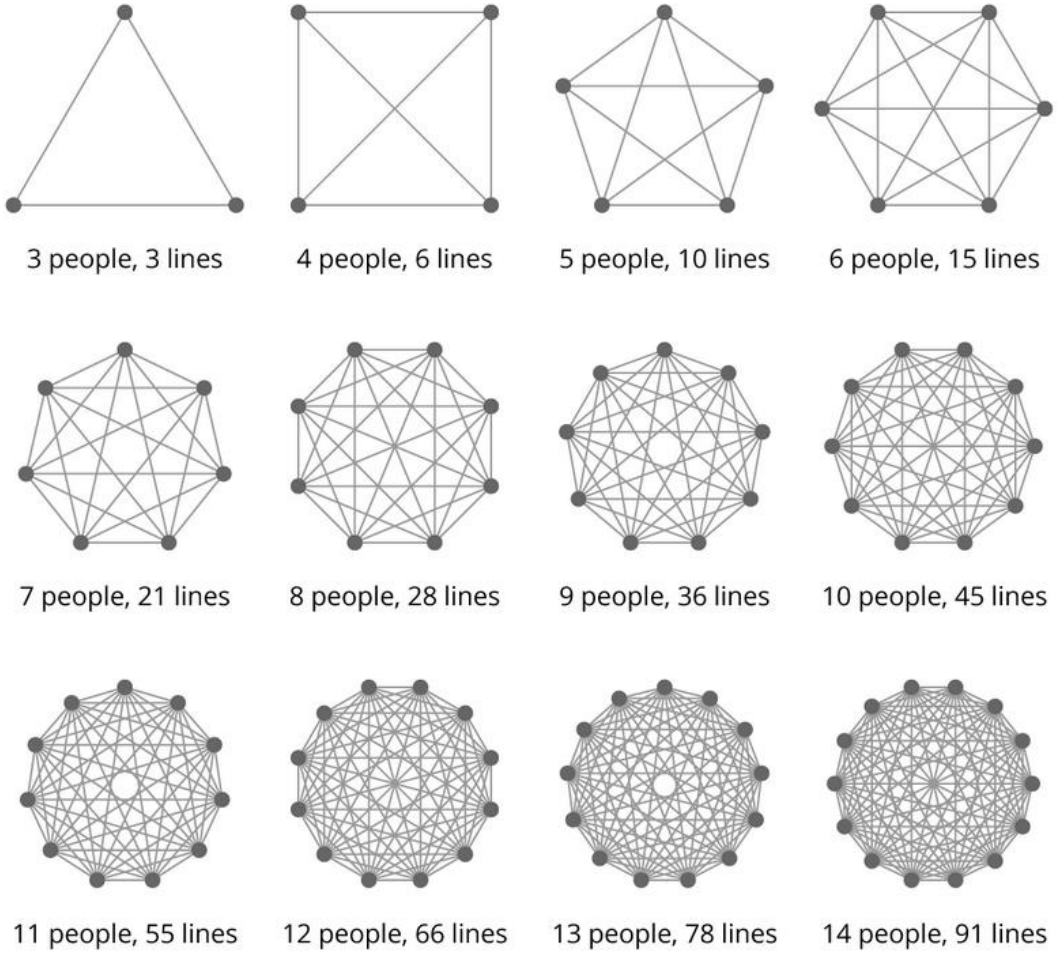
Conclusions

- In 2015, 78% of aquaculture R&D projects in Canada were collaborative
  - The structure and density of the networks varies significantly among different research areas
- In 2015, DFO was an central node in the Canadian aquaculture R&D network being involved in 64% of collaborative aquaculture R&D projects
  - DFO's position varies among networks of different research areas

Discussion

- Increasing collaboration = more edges
  - How many edges is enough?
    - What is the targeted rate of scientific collaboration in aquaculture?
  - Where should new edges be added?
    - With DFO? Should DFO collaborate on more projects? *This would increase DFO's centrality measure*
    - Between partners? Should there be more collaboration between organisations other than DFO? *This would decrease DFO's centrality measure*

Is more always better?



Potential future analyses

- Determine how aquaculture R&D networks change over time
- Determine if increased collaboration correlates with increased valuable outcomes such as publications, science advice, patents, etc.
- Determine the optimal structure of scientific collaboration (cost vs. outcomes)
- Analyse structures of funding networks