MAPPING COLLABORATION NETWORKS IN EUROPEAN AQUACULTURE AND FISHERIES RESEARCH

Linda Seidel-Lass
University of Kiel, Dept. of Agricultural Economics, lseidel@agric-econ.uni-kiel.de

ABSTRACT

The security of world fish supply from wild fisheries and from aquaculture is an increasing concern due to dwindling yields from wild catch and undesirable ecological impacts of aquaculture. The extent to which aquaculture can substitute for fish from wild catch will depend on the productivity of aquaculture, which in turn depends on the level and productivity of aquaculture and fisheries research. Scientific research in aquaculture and fisheries, as in other fields, has become a highly specialised enterprise that could not function without the collaboration of the geographically dispersed specialists.

In this study we investigate one aspect of the extent and pattern of research collaboration among aquaculture and fisheries researchers and scientists. In particular, we measure the extent of co-authorships between researchers from the EU-15 and from Norway using ISI's Science Citation Index Expanded data from 1990 and 2000. Over the observed ten years, the number of publications related to fish and aquaculture has increased nearly six fold. But this considerable growth does not provide evidence, whether the publication pattern itself has changed. We use techniques and measures from social network analysis to identify patterns in aquaculture research collaboration.

Keywords: social network analysis, collaboration, co-authorships, aquaculture, fisheries

INTRODUCTION

Accompanying the enormous rise in the production of fish and other aquatic species from aquaculture, research on this topic has emancipated itself from fisheries research as such and is increasingly assisting the development towards a production similar to intensive agriculture and food production.

First of all, according to the number, the publication scheme of articles in peer-reviewed journals does reflect the positive development international aquaculture has been following in the last decades. In the time frame from 1990 to 2000, the number of publications has increased nearly six fold, according to data of the ISI Science Citation Index Expanded.

In this paper we will go beyond counting numbers and furthermore investigate the structural identities of researchers publishing with one or more co-authors. We will identify the publication pattern by indicators of graph theory and statistical measures of Social Network Analysis (SNA). A social network is defined as a collection of people, each of them being connected to one or more others. In this particular case, the nodes (actors) are the scientists, each of them being linked if they have a joint publication, that is, they have a relationship in which they share information, e.g. on research questions and results. The co-authorship network is a very strong expression of social linkage in science as the decision to collaborate on a paper is made entirely by the authors [1].
The history of the science of social networks goes back to the beginning of the 20th century with roots in graph theory and sociometrics, being united in the Social Network Analysis and becoming popular by the experiment of Stanley Milgram [2]. The aspects of collaboration in scientific publications as a social network have first been analysed in relation to the Hungarian mathematician Erdos, who published at least 1500 articles during his life. The so called Erdos number measures the distance of an author to Erdos: an Erdos number of 1 marks direct co-authors of Erdos, whereas an Erdos number of 2 marks co-authors of co-authors and so forth. Barabási and Albert [3] identified the phenomenon of authors seeking to collaborate with famous colleagues as “preferential attachment”.

Modelling the interactions of scientists as a social network implies the assumption of a self-organisational structure in science. It is widely accepted that the spontaneous organisation of research is the most efficient way to produce high-class research outputs [4]. But why do scientists want to publish their results and why do they seek collaboration on papers? In other words: which economic considerations is the organisation of science based upon? We will introduce the concept of researchers transforming knowledge into information that can be exchanged within the community. We will discuss in detail the economic properties of information as a public good and their influence on the modalities of research outputs being exchanged. In particular, the market organisation will prove to be no optimal allocation mechanism. Instead, it can be observed a self-organisational structure with actors seeking the validation of their results and hence reputation.

The interactions between the members of the social community of researchers can be depicted as a social network model. The preliminary results of this study confirm the assumption that the network in general has significantly enlarged. The individual researchers tend to collaborate more intensely. However, there are limitations in the informational value of the results deriving from the static character of the survey. In an outlook on further research needs, we propose to apply a longitudinal analysis to the aquaculture and fisheries network with periods of at least 5 years to gain further insights into the evolution of the network.

AN ECONOMIC PERSPECTIVE ON RESEARCH ORGANISATION

The interaction of scientists in terms of collaboration on publications can be modelled as the exchange of information. From the economics perspective, information as a good has characteristics that are significantly deviating from private goods, and are more of a public good character [5]. Moreover, exchange of information in science is more than a simple transaction. As Ziman argues “the nature of the knowledge produced by this system is closely bound up with its social structure” [6]. This idea can be transformed into a social network model, if we consider the knowledge to be produced by individual actors rather than by “a system”, but where the actors are in turn embedded in a specific social structure. The organisation of research, the process of generating knowledge and the process of the diffusion of knowledge in scientific communities have a long history in the social and behavioural sciences.

However, the transactions of the social entities remain an economic activity. To draw a picture of the production process in science, it is initially important to gather insights into the inputs of research, what do researchers exchange, thus what is their output, and why do they exchange their “raw materials” and the output. In the production process of research, actors (scientists)
transform inputs into outputs. The input in science can be understood as information, that is, knowledge translated into communicable messages between a set of actors. This means, that a) knowledge can be materialised in oral or written expressions as well as information stored in a researcher, and b) knowledge needs to be transformed to information. Otherwise, a recipient would not be able to make use of it or more precise: it would not be tradable [7].

To deepen the economic perspective on the production of scientific output, we need a detailed discussion of the good “information”. Information is characterised by distinct attributes. There exists non-rivalry in consumption, i.e. it can be consumed by more than one person without loosing its valuable content and at the same time it is possible to exclude other people from using it. Another specific characteristic are the high economies of scale, that is, the first output unit produced is by far the most expensive, and the production costs of the second can be assumed to be negligible (equate the material costs). The edition of a book might cost the remuneration of the author, the editorial office, the print and the distribution of the product. Once the author has written his book and the editorial office has finalised its work in form and content, the costs for a second print are barely noteworthy. From the economics perspective, a second interesting part begins, as information as a good implies a paradox of value. The bookseller can not accept, that a potential buyer of a new book reads it before he buys it. Once he has read it, the value of the print edition is gone (to the consumer). From the consumer’s perspective this means, that he cannot convince himself of the quality he can expect from this good before buying it. This will emerge as a very important characteristic of information exchanged within the social community of science. Scientists have to rely on something that emerges within the social network: reputation. A distinct discussion of the social concept of reputation has to be excluded from the considerations here, as it would exceed the scope of the study. A final important characteristic is the positive network effect, i.e. the value of information increases with the number of its consumers. Thus, the more a scientific publication is read (citations are widely used as an indicator for this), the higher is the value of its content.

Considering these formal characteristics as the determinants of the extent and manner to which information is exchanged in the scientific community, we can gather insight into the incentives scientific collaboration is based upon. Again, from the economics perspective the efficient allocation of scientific effort is not possible in a market organisation due to the public good character of information. At the same time, these characteristics also inhibit the organisation to follow a hierarchic model. If the organisation were hierarchic, who was to value the knowledge produced, if it was worth to be transformed to information and exchanged on the market or not? Research has been forced into a hierarchic organisation model during the Cold War. Specific targets were set, and science was expected to meet these political targets. However, no one would regard this allocation mechanism as efficient. Another extreme would be to consider optimal allocation to derive from perfect laissez-faire [4]. The scientific studies on the sociology of science have proved that it is much more than an unstructured chaos that represents the organisation of research. Based upon distinct mathematical and statistical analysis, the methodology of social network analysis validates the assumption of a self-organisational structure in research [8]. The sociological concept of codified communication in science remains unclear about the nature of this code that is said to determine the differentiation of disciplines. From the social networks perspective we know, that scientific collaboration can be modelled as a complex system, i.a. implying a self-organisational structure, the emergence of structures which can be explained by the interdependence of different social entities within the system, and path
dependence, thus connecting incidences of longer periods. This complex evolving system will be formulated as a social network model of research organisation.

SOCIAL NETWORK MODEL OF RESEARCH ORGANISATION

A social network model measures the relationship between social entities like researchers exchanging information and publishing their scientific output as papers in journals, working papers or conference proceedings. The social network analysis enables research to identify regular patterns within the social group under observation which are referred to as structural identities of this relation. Following Wasserman and Faust [9] the distinctions that characterise social network analysis in comparison to other relational concepts are the interdependence rather than the independence of actors, the linkages that are a channel for transactions between the actors (e.g. information or reputation), the embedding of actors into a social network of constraints and opportunities for the individual and the conceptualisation of pertinent social structures (economic, political and social relations).

From the network perspective, the linkages between actors can be characterised by any relationship like group membership, kinship or flow of resources. However, it is very important to note, that relations are always measured on a pair of actors, connected by a linkage. The pair is the basic unit under study. The analysis of relationships between actors can include as well as exclude attributes of the individual actor. The relational tie of a pair of actors always comes first in the collection of the structural identities of the network. The term “social network” refers to a collection of people, each of them being connected to one or more others. Using measurements of social network analysis (descriptive and statistical), one is enabled to identify the strength of influence an individual actor has in the particular social structure.

In the past, many networks have been analysed on the basis of data collected in interviews, e.g. as so called ego-centric networks with actors being asked to name other people they feel acquainted with. Thus, most data sets did not contain more than a hundred actors. With the development of data processing facilities, analytic tools in social network analysis and the World Wide Web as an extensive source of information, social network studies are now workable for large data sets and complex analytic tasks. However, one needs to distinguish between data sets that describe loose social relationships like the interactions in the World Wide Web itself [10], and more “real” social relations like actors being associated to each other by the common membership in clubs and associations. In the so called “affiliation networks” the actors have a strong social relationship, and the data set is reliable by the assignment of individual actors as members of such organisations in cold print on membership lists [11]. An even more precise network is the network of scientific collaboration in co-authorships on published papers. In most cases, it can be assumed, that the common output of scientific interaction in the form of a joint publication is the expression of researchers working closely together and, although being increasingly geographically separated, sharing the attribute of knowing each other on a face-to-face basis, at least they will have met each other in person at the initial contact.

Recently, there has been a growing interest in the analysis of scientific collaboration. First approaches have mapped the flow of information through science within and between disciplines by analysing the citation pattern in publications. This has a long tradition in Bibliometrics (a sub discipline of information science dealing with publication and citation patterns in science). As
mentioned above, the intensity of relations in co-authorships has attracted more attention to these approaches and several studies have been conducted on scientific collaboration in various disciplines. So what can be expected when inquiring the research network of aquaculture and fisheries publications? Applying the concept of social networks on the organisation of aquaculture and fisheries research is a challenging task. Firstly, methodologically. The available data sets offer an excellent basis for descriptive and high-numerical statistical analysis, and they do not extend a reasonable size. Thus, a detailed analysis of the attributes of individual researchers is still possible. Secondly, in comparison to agricultural research, fisheries production and fisheries research undergo an immense development process, shifting the long-term perspective of world fish supply more and more towards sustainable aquaculture production. It is a blue revolution after the green revolution thousands of years ago. The intensity of research on this topic has been volatile within the last decades. The organisation behind the research and development of aquaculture and sustainable fisheries production systems can be expected to have developed accordingly. Like other studies on scientific collaboration show, collaboration has increasingly become international with some clusters of highly productive researchers (by number of co-authored publications) and a periphery of those still seeking to be “absorbed” into the centre of scientific excellence. On the other hand, in the aquaculture and fisheries case, we are talking about an applied and interdisciplinary science unlike physics, biomedicine or computational science [11], just to name some of the other studies published on scientific collaborations. Can we expect a similar topology of the aquaculture and fisheries network at all? That means, can we expect well fitted power law distributions of the number of papers and the number of collaborators of scientists? Which role does the interaction between different disciplines play for the intensity of collaboration? We will propose a social network model of research organisation that can be applied to the European network of aquaculture and fisheries publications.

A social network can be represented in different ways. In the graph theoretic notation the network is represented by a graph, which consists of nodes and lines. Let \( N = \{n_1, n_2, \ldots , n_g\} \) denote a set of \( g \) actors. They represent the individual researchers. The set of actors \( N \) is also named “mode”, that is a distinct set of entities on which structural variables are measured. They are called one-mode networks. All actors belong to one set. Very generally, we can further assume that each actor in \( N \) relates to every other actor therein and that the relation is binary-coded: \( n_i \) relates to \( n_j \) or does not. If a link exists, thus, if \( n_i = 1 \), the ordered pair is an element of a collection of pairs, denoted as \( L \). If an ordered pair is an element of \( L \), then the first actor relates to the second in this relation [9]. A graph consists of a set of nodes \( N \) that are connected by a set of lines \( L \) and can be described by the two sets \( (N, L) \). The symbol \( G \) denotes the graph. The simplest possible network can be represented by the algebraic structure \( S =< N, L > \).

As the co-authorship relation is an affiliation network, a modification has to be made to general social networks. In this particular case, a second mode exists. A second set of actors in a two-mode network can represent e.g. an event, the actors participate in or an organisation, they are members of. In our case, it represents the publication, the authors have published jointly. We denote the second mode by \( M = \{m_1, m_2, \ldots , m_h\} \). The corresponding graph to a two-mode network is called a bipartite graph, as lines connect only nodes from one set to another, but not within the sets. Another distinction to general social networks is that affiliation networks contain information on collections of actors larger than pairs, assuming that social events usually are
attended by more than two people. And finally, the relationship between the first and the second mode can be considered as complementary. That is, the actors can be connected to one another by their affiliation with organisations and at the same time the organisations can be connected to one another by the actors who are members of them. It is thus called a co-occurance relation [9]. For analytical purposes, two-mode networks can be transformed into one-mode networks, with the nodes of the second set (articles) becoming elements of the first set (authors). Whenever two papers share an author in the two-mode network, there is a link between them in the one-mode network. In a two-mode network the interconnectivity of an actor with one or more organisations is represented by loops. A loop is a line that connects an actor to itself, e.g. it represents the number of co-authors on a paper. Whenever an author has participated in writing more than one paper, he is connected to the other papers by multiple lines, which will be labelled by his name in a two-mode network. These multiple lines can be replaced by a valued single line, indicating by the value the number of former single lines of an author. To this transformed and simplified two-mode network, the standard techniques of one-mode networks can be applied, although there remain some differences one has to consider when analysing results from such networks [12].

THE EUROPEAN NETWORK

The data for this study were derived from the bibliographic online database of the Institute for Scientific Information (ISI) between April and July 2006. The Science Citation Index (SCI) Expanded that is produced by the ISI provides an extensive database on journal articles as well as on proceedings, reviews and letters for every year since 1990. The SCI Expanded is a multidisciplinary index, including some 5900 journals of the sciences. We collected data on the world total publications on aquaculture and fisheries for the years 1990 and 2000, and extracted from that the publications of authors coming from the EU-15 and Norway (in the following referred to as European papers) for a detailed analysis in a social network model. The EU-15 countries were chosen because they jointly account for more than one third of aquaculture and fisheries publications worldwide. Norway was included because of the country's considerable involvement in aquaculture production and research. A distinct European analysis is possible as authors are indexed with their institution and country address.

In a first step, the two data sets have been analysed by absolute measures to gain an encompassing view on the data involved. Thus, counting the number of international papers and extracting from that the European papers, validates the assumption that the European aquaculture and fisheries research sector is vitally important when it comes to the dissemination of research results. Europe accounts for one third of all published articles in the discipline under study, and is able to increase that share even more in the time frame until 2000 (Tab. I). Another impressive figure characterises the development of the publications as such: the number of papers has increased nearly six fold between 1990 and 2000. This interval has been chosen as it begins with the first available data set in the ISI, but it excludes data beyond 2000 which has proved not to be stable enough yet, due to the ongoing indexing process. However, the interval covers a ten year period of scientific work which has been accompanied or driven by a restructuring process in the fisheries sector and a strong growth in the aquaculture sector [13, 14]. Apart from the developments in the sector and in the related scientific disciplines, the general conditions for science have changed in the sense that information and communication technologies have improved, the scientific collaboration within Europe has been emphasised by the research policy of the European Union and cross-country cooperation has been eased after the collapse of the
Eastern Bloc [15]. The comparison of the two data sets shows a slight increase in the mean number of authors per paper as well as in the mean number of papers per author. More significant is the decrease of single-authored papers to a negligible level. Thus, only a few authors have no collaborator when publishing a paper.

Tab. I: Data set and statistic analysis results of SNA

<table>
<thead>
<tr>
<th></th>
<th>1990</th>
<th>2000</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total number of papers - world total</td>
<td>1409</td>
<td>8329</td>
</tr>
<tr>
<td>Total number of papers - EU-15 and Norway</td>
<td>465</td>
<td>3316</td>
</tr>
<tr>
<td>Percent European papers of world total</td>
<td>33%</td>
<td>40%</td>
</tr>
<tr>
<td>Total number of authors - EU-15 and Norway</td>
<td>1192</td>
<td>10649</td>
</tr>
<tr>
<td>Mean number of authors per paper</td>
<td>2.6</td>
<td>3.2</td>
</tr>
<tr>
<td>Mean number of papers per author</td>
<td>1.2</td>
<td>1.3</td>
</tr>
<tr>
<td>Max. number of authors per paper</td>
<td>12</td>
<td>32</td>
</tr>
<tr>
<td>Number of single-authored papers</td>
<td>88</td>
<td>15</td>
</tr>
<tr>
<td>Percent co-authored papers</td>
<td>81.08%</td>
<td>99.55%</td>
</tr>
<tr>
<td>Density</td>
<td>0.0027291</td>
<td>0.0006425</td>
</tr>
<tr>
<td>Avg. path length</td>
<td>1.22470</td>
<td>1.01</td>
</tr>
<tr>
<td>Diameter</td>
<td>3.0</td>
<td>2.0</td>
</tr>
</tbody>
</table>

In preparation of the social network analysis, the data have been sorted by the surname and first name initials of the authors. This data set of authors is the first mode of the network. The second mode is represented by the corresponding articles. As the boundary specification of the network included only papers of authors from the EU-15 countries and Norway, we use the online facilities of the ISI to extract those papers from the entire data set of aquaculture and fisheries articles in the respective year. The raw data are then saved in text files from the online data base and imported into a spreadsheet program, where they are rearranged so that every author corresponds to the articles he contributed to. When using surnames and first name initials there is potential for an error in the data set by inconsistent initials. However, the potential error has not been calculated for this data set, but the general problem has been discussed in the literature. Newman calculates the error by surname confusions and incomplete initials to be a few percent [16] and Barabási does not expect a significant impact on the results [1].

Further information relating to the authors and articles, e.g. institutional affiliation, can be derived from the ISI data base, but were not considered for this study. The data have been imported as two files for the particular years into Pajek, a social network software which is available free of charge. Pajek has proved to be a powerful visualisation and analysis program with special strengths in handling large data sets. The data were handled as two-mode networks that were then transformed into one-mode networks of authors. As depicted in Tab. I the networks have been analysed by social network methods to get an impression of the network characteristics.

The density of a network is defined as the number of links in a simple network, expressed in relation to the maximum possible number of links. The results for the two European networks indicate a very scarce density of the networks with a proportion of connected links being 0.02% and 0.006% for 1990 and 2000, respectively. These very low numbers indicate a network that is divided into numerous small subgroups, but with no significant tendency of interconnection between the groups.
Comparing the two networks for the development over the ten year period, we derive with the indicator of average path lengths, a measure on the interconnectedness in the network. A path is defined as a walk, in which no actor in between the first and the last actor occurs more than once [12]. The length of a path is the number of lines in it. The more connected a network is, the shorter is the path length between actors.

The diameter of a network measures the maximum path length between any two actors. As many networks are not connected (in a connected graph there is a path between every pair of nodes), the diameter is reported for the largest component. The component of a graph is defined as the connected sub graph (groups of nodes which are not connected to each other) of the graph. The diameter can range from a minimum of 1 to a maximum of g-1 for all actors. From the two European networks we know that they are not fully connected. Thus the diameter is reported for the largest component. The diameter decreases for the second time period, implying that the distance between two authors within a sub graph is shorter than before which can be understood as an intensification of collaboration, at least within the subgroups.

Apart from the analysis of collaboration in the European network by statistical measures of social network analysis, we can depict the network as a graph of nodes and links. For this purpose, Pajek offers a software tool for visualising networks and partitions. A partition of a network is a classification or clustering of the nodes in the network where each node is assigned to exactly one class or cluster. Pajek discriminates the clusters according to structural identities of the nodes and according to a dissimilarity measure between the nodes. Visualising a network gives some useful insights into the characteristics of the network. However, the picture drawn is an unscaled three dimensional graphic providing no information on the actual distance between the subgroups. What is observable between Figure 1 and Figure 2 for the network in 1990 and 2000, respectively, is that collaboration in aquaculture and fisheries science is concentrated on subgroups that refer to one or more papers being jointly published by a group of actors. Only a small number of authors interact with different collaboration clusters. Some authors have a bridging function between two groups, being a co-author on two papers of different groups. The first visual impression of the network in 1990 is a network of highly separated subgroups with only few interlinkages. There is no indication of very central actors, that is, authors which have a high share of co-authored papers of different subgroups.
Figure 1: The European Aquaculture Network, EU-15 plus Norway in 1990

Figure 2: The European Aquaculture Network, EU-15 plus Norway in 2000
The European network of 2000 has grown immensely. The number of authors has increased nine fold within the ten years, complicating the visual depiction of the network. For simplifying purposes, most of the subgroups are depicted as single spots, identifiable by a representative’s name and the node colour of the corresponding cluster. What can be said from the second network is an obvious intensification of interconnection between subgroups. There are still very strong unconnected subgroups, but there are more bridging actors, thus authors collaborating beyond their subgroup “boundaries”. Again, just by visual impression, the bridging authors seem to relate to smaller subgroups. The last aspect of the individual characteristics of bridging authors can, among other things, be hypothesized and empirically tested with distinct social network analysis which we see as a need for further research.

SUMMARY AND OUTLOOK

This study is a preliminary analysis of the network of scientific collaborations among publications in the EU-15 and Norway. By means of descriptive statistics on a data set of publications in 1990 and 2000 of the ISI Science Citation Index Expanded, this paper is able to detect a large growth in the total number of publications and authors. Applying social network analysis tools to the data sets helps characterise the network as scarcely connected. By means of graphic visualisation, a network of separated subgroups is identified.

For a more detailed analysis of the structural identities of the network, we suggest a longitudinal analysis applied to the network over a time period of at least five consecutive years. The enlarged time frame will provide the opportunity to observe the emergence of structures within the network. On the level of individual actors, a time frame of five years or more enables to observe authors entering and leaving the network, connecting and disconnecting themselves to and from the same or different subgroups.

The concept of preferential attachment provides a promising tool to explain, why central authors, once they are detected in the network, are able to strengthen their position by attracting more co-authors in relation to other scientists. This concept does explain why the successful stay successful or improve their position, but it does not give any explanation of the factors influencing the decision to accept or to deny the co-authorship of a scientist seeking to attach to the famous colleague. What social criteria are the basis for acceptance or deny of collaboration? We suggest extending the affiliation model of scientific publications by further characteristics of the individual actors. Such characteristics will be affiliations to scientific organisations and associations. In this context, we suggest to apply the concept of interlocking networks to scientific organisations and their members. This concept has first been applied to interlocking directorates to measure the importance of simultaneous membership in directorates of large business companies [18]. More on the scientist’s level, the integration of attributes related to the individual career of a scientist will further extend the social network model. In particular, aspects like the origin of the degree (university), the origin of the doctoral thesis (university or tutor) and the current institution.

Extending the social network model on the basis of publication data from a database like the ISI includes the risk to have a) an English language bias and b) a peer-reviewed journal bias. Further studies on the applicability of the data set with special respect to an international analysis will have to be conducted.
ENDNOTES

a Special thanks to Jörg Müller-Scheeßel and Arne Henningsen for their support in programming a shortcut for sorting the data.
b Pajek (V. Batagelj and A. Mrvar, version 1.14, 2005) can be downloaded free of charge at http://vlado.fmf.uni-lj.si/pub/networks/pajek/
c See for example Tullock [4] for an introduction to the incentives of individual researchers and a discussion of the concept of reputation.

REFERENCES