A Neural Network Based Content Analysis of Antarctic Science Research: Studies into 25 years of research publications (Part I)

Prabir G. Dastidar* and Deepak Kumar Jha§

ABSTRACT

In this study content analysis of Antarctic Science research using a network analysis based algorithm was done. Titles of the research articles were used as proxy to do the content analysis; 10942 of research articles published during 1980 to 2004 (25 years) in Science Citation Indexed (SCI) journals were used for the analysis. A neural network based algorithm—CATPAC was used to generate normalized co-word matrix from significant most-used words. The matrix was used to study the thematic structure of research at cognitive level. Parameters like significant words, window size, slide size, cycles, clamping, threshold and decay were defined for the analysis. Network analysis tools like structural equivalence, centrality were used to study cognitive blocks and characterize cognitive associations of the significant words embedded in titles of the publications.

Keywords: Antarctica, Policy studies, Content analysis, Cognitive analysis, thematic analysis, Contextual analysis, Scientometrics, Neural network, Co-occurrence matrix, Co-word , UCINET, CATPAC, Social Network Analysis (SNA)

---

* Ministry of Earth Science, Block No. 9& 12, CGO Complex, New Delhi - 110003, India, Email — Prabirgd11@gmail.com.

§ NISCAIR, Education and Training Division, 14, S.V Marg , New Delhi - 110067 , India, Email —

deepaksnd@gmail.com

1.0 INTRODUCTION

The research journals provide a macro level view of the main research themes, subfields and their linkages in a research domain. Therefore, a thematic analysis of a research specialty constitutes an important study. However, from this Various research fields within and between the subfields, among the concepts, etc. The journal network helps in delineating a knowledge domain, but a particular knowledge domain is further organized in a network of knowledge sub-domains. To understand these knowledge sub-domains, analysis of the indicators is required at the micro levels, like cognitive units in the form of representative words and their association patterns, leading to concept formations, activity structures in subject specialties.

This contextual analysis can be done through the study of words in titles or in abstracts of published articles. This can also be done through analyzing the keywords and indexing terms which can be seen as representing concepts described as ‘poles of research interest’, ‘research themes’, ‘problems domain’, etc. Therefore, this study is termed as content analysis and cognitive analysis. The assignment of an appropriate set of codes can be viewed as manifestation of an expert assessment of the scientific publication’s cognitive structure. The network of co-occurrences between different codes, collected on a specific set of publications, allows a quantitative study of the structure of publication contents, in terms of the nature and strength of linkages.

A scientific field is characterized by a terminology of ‘words’, which signify concepts, operations, processes or methodologies. These important ‘words’ are reflected in the titles or abstracts, as a research worker attempts to convey or highlight the important and salient points of his/her paper. Co-occurrences of conceptual words in a large number of documents, in titles or abstracts of papers, signify the important relationships among these words. Thus, the structure depicted by the frequency of co-occurrence of conceptual words reveals important and interesting linkages among them and provides a further insight into the framework of a research field. These contextual analyses of co-occurrence of codes and of conceptual words enable the investigator to grasp the static and dynamic aspects of the manner in which scientists relate and place their work in a hierarchy of scientific research concepts. In addition, this method provides a direct quantitative way of linking the conceptual contents of publications. Hence, such a co-occurrence structure can represent research activities within a scientific area via depiction of concepts and topics, which are active, and the relations among them.

The network of co-occurrences between different words, collected on a specific set of publications, allows a quantitative study of the structure of publication contents in terms of the nature and strength of linkages between the pairs of words.

Word usage is more codified, and it seems always possible to distinguish between words with a major theoretical, methodological, or observational meaning within the context of a given specialty. It provides an analytical framework for carrying out dynamic analysis of the contents of articles (Leydesdorff 2001). The keywords are often used to identify sub-domains of research specialties. For this study, the sub-domains were identified using structural equivalence techniques, by grouping keywords at different levels (Borgatti S.P., Everett M.G. and Freeman L.C. 2002)

This method labeled as ‘co-word analysis’, provides a direct quantitative way of linking the conceptual contents of publications, by comparing and classifying publications with respect to the occurrence of similar word-pairs. Hence, such a co-word structure can represent research activities within a scientific area. It does so through the depiction of the state-of-art research in that scientific area by delineating and underscoring the relations between various research themes. The co-word analysis was applied in this study to identify the emerging research areas in Antarctic Science. As a result, co-word approach was applied to uncover the topics/areas which were active. The picture that emerged depicted the micro level description of the specialty of a field. This is a valuable supplement in understanding the intellectual structure of the field (Figure 1).

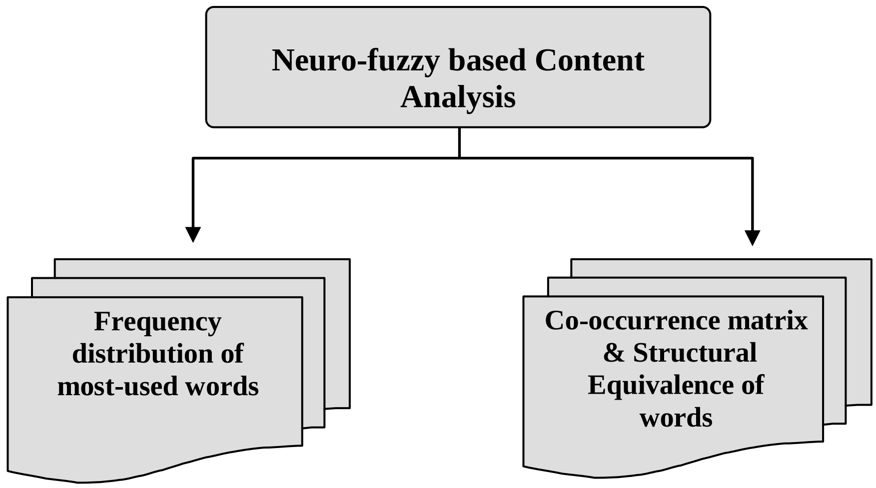

Figure 1 A schematic presentation of content analysis conducted on data

The first well-documented case of quantitative analysis of printed material goes back to the eighteenth-century in Sweden involving a collection of 90 hymns of unknown authorship entitled 'Songs of Zion'. Content analysis or thematic analysis has numerous applications, spanning from marketing research, propaganda analysis and lately computer text analysis. In psychology, this technique has found three primary applications. The first is the analysis of verbal records to discover motivational, psychological or personality characteristics. The second is the use of qualitative data gathered in the form of answers to open-ended questions, verbal response to tests, thematic testing, and the third is concerned with the processes of communication in which content is an integral part (Krippendorff 1980). In previous works authors have studied the global intellectual structure of Antarctic Science research and identified active players—countries, organization and authors in Antarctic Science Research.

(Dastidar and Persson 2005 and Dastidar and Ramachandran 2005). To visualize a complete picture of Antarctic Science research authors felt necessary to study the content / thematic structure of research as well.

In this work attempt has been made to analyze the content of Antarctic Science research during 1980 to 2004—which is denoted as part I of the study. In part II and part III, findings of Ocean Science & Ocean Engineering dataset will be presented.

2.0 MATERIALS AND METHODS

2.1 Neural Network-based Content Analysis

A real biological brain consists of a set of neurons, which are essentially biological “switches.” In the simplest form, these switches can be either “on” or “off”, but in more complicated models, the neurons can take on several “levels” of activation. When sufficiently stimulated, a neuron becomes active or “fires.” Many of these neurons are connected to other neurons by neural pathways which can conduct stimulation from one neuron to another. Some of these pathways are in place at birth while others are formed during life as a result of experience. Because of these connections, activating some neurons in the brain generally results in activating others as well.

Artificial Neural Network (ANN) is an information-processing paradigm which mimics the parallel structure of the neuromorphic system of mammalian brain. Artificial neural networks are collections of mathematical models that emulate some of the observed properties of biological nervous systems and draw on the adaptive biological learning. It is composed of a large number of highly interconnected processing elements that are analogous to neurons and are tied together with weighted connections that are analogous to synapses.

Learning of biological systems involves adjustments to the synaptic connections that exist between the neurons. This is true of ANNs as well, where the training algorithm iteratively adjusts the connection weights (synapses). These connection weights store the knowledge necessary to solve specific problems. They are good pattern recognition engines and robust classifiers with the ability to generalize in making decisions about imprecise input data. They offer ideal solution to a variety of classification problems, such as speech, character and signal recognition. The advantage of ANNs lies in their resilience against distortions in the input data and their capability of learning.

To carry out the analysis in this research, a self-organizing neural network based algorithm (software)**—CATPAC**, was used to derive the normalized matrix of word associations (Woelfel 1998).

Each word that **CATPAC** sees is associated with an artificial "neuron" in **CATPAC's** simulated brain. As a result of the learning and forgetting rules, **CATPAC** will produce a 'brain" consisting of a network of interconnected neurons, each of which represents a word in the text. Some of these neurons will be tightly and positively connected, indicating that they are closely associated. Whenever one of them is activated, likelihood is great that the other will also be called to mind. Other neurons will be strongly negatively connected, indicated that one is very unlikely to be active when the other is active. Such neurons actually inhibit each other, so that activating a node will tend to de-activate other nodes to which it is negatively connected.

The neural network based algorithm makes it possible to retrieve episodic memories of the text document. Episodic memories differ from semantic memories by containing circumstantial info (who, what, when, where, etc.). Remembering episodic memories is generally more complex than recalling semantic memories, involving the evaluation of cued memories based upon the current goal (Raye et al 2000).

The algorithm works by passing a moving window of size n (in the present analysis 3-word window was used) through the text. In our study the text was a collection of all the titles of the papers. Each title was separated by delimiter '-1' to single out contributions from individual publications. Any time the window encounters a word, the neuron representing the word becomes active, connections among active neurons are strengthened, so the words that occur close to each other in the text tend to have higher level of connections. In subsequent scanning, if a word is encountered again, its value will go up, while in the absence of it, the activation level of words (neurons) goes down.

Because of its self-organizing characteristics, the algorithm can learn from the patterns of associations and generate normalized matrix. This matrix can be used to generate non-hierarchical clusters and to perform other network analysis.

A word association matrix was constructed by taking into account the connection strengths among the neurons that represent the words. It is not a simple co-occurrence matrix. This matrix not only represents the direct co-occurrences among the words, but also their indirect connections. For example if word 1 and word 2 co-occur, and word 2 and word 3 co-occur, but word 1 and word 3 never co-occur, nevertheless, algorithm links the words 1 and 3 because of their indirect connection through word 2. The resultant matrix is a generalized scalar product matrix normalized to approximately plus or minus 1.1. This may be treated as a generic similarities matrix. Cluster analysis of the resultant matrix gives a better expression about its purpose than the results obtained from simple co-occurrence of words. Like 'Pacific' and 'Ocean' do not convey much meaning independently but if the word 'wave' comes with this group, it conveys that 'wave research on Pacific Ocean'.

2.2 Title Words as Indicator of Research Activity

Titles constitute an important indicator of the content of a research article, and provided clue to the importance of the work. Numerous surveys have shown that bibliographies appearing in papers are one of the most valuable sources of information in literature searching (Garfield 1968). Words and citations are important indicators of research activity. Title words provide a special perspective on science and scholarly activity and for identifying research fronts (Garfield 1986). Search terms extracted from titles of articles are useful search terms for retrieval of information from databases and augmenting retrieval efficiency (Garfield 1990)

2.3 Generation of Matrix

Following neural network parameters were selected to generate the normalized matrix:

2.3.1 Significant Words

Zipf (1972) had described the frequency with which words occur in a given piece of literature. It was found that multiplying rank (r) of the words by its corresponding frequency of occurrence (f) gives a constant, C, i.e.

$$C= rf$$

Power Law behaviour provides a concise mathematical description of sheer dominance of few members over the total population (Luscombe 2002). The power law behaviour has been observed in different population distributions, these included income levels, relative sizes of cities (Zipf 1972), and connectivity of nodes in large networks (Barabasi and Albert 1999).

In a typical distribution profile of words in a text document (in the present study it was a list of titles of papers published in peer-reviewed journals), the dominant word groups reflect the main theme of the document. Synonyms were clubbed together to derive a consolidated picture on the technical words. The words with high frequency of occurrence signify that the concepts which they depict are important. Among the highly ranked words, the cut-off values were determined to generate the matrix. The most frequent words occurring in the top layer were chosen to generate a matrix, which was used to carry out network analysis.

2.3.2 Window Size

The software algorithm works by passing a moving window of size *n* through a file. For example, for a window size of 4, and a slide size of 1, **CATPAC** would read words 1 through 4, then words 2 through 5, then words 3 through 6, and so on.

Any time a word is in the window, the neuron representing this word becomes active. Connections among active neurons are strengthened, so words that occur close to each other in the text tend to become associated in **CATPAC's** memory.

2.3.3 Slide Size

This prompts to ask how you would like the moving window to "slide" through the text. The number defines how many words the window will skip prior to reading the text. It may select any increment one may specify. For example, in case of a window of 5, and a slide size of 1, **CATPAC** would read words 1 through 5, 2 through 6, etc. In case a window size of 5 and a slide of 2, **CATPAC** would read words 1 through 5, then 3 through 7, etc.

2.3.4 Cycles

CATPAC's network analysis procedure works in the following manner: When words are present in the scanning window, the neurons assigned to those words are active, and the connection among all active neurons is strengthened. In addition, the activation of any neuron travels along the pathways or connections among neurons, and can in turn activate still other neurons whose associated words may not be in the window. These neurons can, in turn, activate still other neurons, and so on.

In an actual (biological) neural network, these processes go on in parallel and in real time, so that the signal coming into the network is spreading at different rates of speed throughout the network, and neurons are becoming active and inactive at different times. This process of delay is called *hysteresis*.

Very little cycling (or none at all as in the simple co-occurrence model) tends to find only highly superficial associations. Too much thinking (cycling), however, is not always a good thing, since **Catpac** can tend to see things as all pretty much alike if it is allowed to cycle too many times. In the analysis the default 'cycle 1' was used.

2.3.5 Clamping

When a word is found in the window, its neuron is activated. However, it can become de-activated again as the network goes through its normal processes, just as we see things, become aware of them, and then forget them. Clamping the nodes (another word for neuron) would prevent them from turning off again.

Chip-Head network options: The most generally useful neuron and some reasonable values for the three generally useful neurons (functional forms), and some reasonable values for the three general parameters have been chosen as defaults in the analysis.

2.3.6 Function Form

Out of four available function forms: a logistic varying between 0 and +1, a logistic varying between –1 and +1, a hyperbolic tangent function varying between –1 and +1, and a linear function varying between –1 and +1, the default one, i.e. logistic varying between 0 and +1, were used for the analysis.

2.3.7 Threshold

Each neuron in **Catpac** is either turned on by being in the moving window, or else receives inputs from other neurons to which it is connected. These inputs are transformed by a *transfer function*. After the inputs to any neuron have been transformed by the transfer function, they are summed, and, if they exceed a given threshold, that neuron is activated; otherwise it remains inactive. By lowering the threshold, you make it more likely for neurons to become activated; by raising the threshold, you make it less likely for neurons to become activated. Default threshold zero was used for the analysis.

2.3.8 Decay Rate

The decay rate specifies how quickly the neurons return to their rest condition (0.0), after being activated. The default rate of 0.9, means that each neuron, if not reactivated, will lose 90% of its activation in each cycle. Raising the rate makes them turn-off faster; lowering the rate means they are likely to stay on for a longer period.

2.3.9 Learning Rate

When neurons behave similarly, the strength of the connection between them is strengthened. The learning rate is how much they are strengthened in each cycle. The default 0.01 was used in the analysis.

2.4 Structural Equivalence Blocks as Specialty Areas

Lorrain and White (1971) proposed that if nodes are people, then social positions may be conceived as equivalence classes or 'blocks' of people who relate in a similar way to other such blocks. A concrete network can be transformed into a simplified model of itself where the nodes are combined into blocks and the relation(s) between nodes are transformed into relations between blocks.

Ideally, if two individuals (nodes) have exactly the same pattern of giving and receiving ties, they are structurally equivalent to each other. A set of such nodes jointly occupy a common position in the network. In principle, a set of positions, each occupied by nodes structurally equivalent to each other, can be determined. These positions are structurally non-equivalent. These are the blocks.

The relations between nodes, both within and between blocks, can be used to construct the relations between the blocks. It is important to note that the reduction operates simultaneously on nodes and relations yielding a structural image that is simpler and amendable to more abstract analysis. In a network of individuals the members of a jointly occupied position (block) may not even know each other, just as two judges in different cities may not be acquainted or otherwise related-but share a common set of relational patterns, to prosecutors/defendants, jury members, and the like (Doreian and Fararo 1985).

Structural equivalence within a block implies that a block is formed with members that have a similar pattern of association, i.e., a similar pattern of giving and receiving ties. This is not the same for the groups that are formed through cluster analysis. In cluster grouping, only strong cohesive linkages among members result in their being in a particular group. In structural equivalence, the main criterion of a member being present in a block is that it has a strong association with another node. Thus, members are expected to be connected in a relationship among themselves through this external tie. Similar to cluster grouping, they are also expected to have linkages among themselves. But this is not a necessity to form a block/group, as it is in the case of cluster approach. Thus, this provides us a new method of looking at the relationships. (Hannema Robert 2006).

Words with strong structural connections were observed to be coming in a structurally equivalent block. Mainly the connections are associated with prosperities, types, effects or methods used for investigations. The blocks are categorized into possible research areas. This assigning is done based on observing the strength of linkages among the words inside the blocks. Further the context of these words is seen from the titles, i.e., words which are embedded in the titles. This contextual understanding is a prerequisite exercise visualizes the research area. (Bhattacharya and Basu 1998)

The empirical, or operational, methods of reducing a concrete social network to a simpler image of itself are referred to as “block modeling.” The model proposed by Breiger et al (1975) relies on iterated correlations, while Burt and Schott (1990) proposed technique uses Euclidean distance. The method of Structural Equivalence which looks at the relationships among words as well as structural equivalent blocks is more appropriate for mapping research specialties at the micro levels, as it considers indirect linkages also. As proposed by Doreian and Thomas, 1985, the mean densities of the matrix were used as cut-off points to generate image matrices from the density of the blocks. These structures were viewed as reduced images of initial cognitive networks. These image matrices were used to draw network maps.

Ucinet software (Borgatti et al 2002) was used to study the structural equivalent blocks and calculating Freeman’s centrality values of the most-frequently used words.

3.0 Data Cleaning

SCI Database search with ‘Antarc*’ in title, from the year 1980 through 2004 (25 years), retrieved 10,942 records. The titles of all the articles were used for thematic analysis, following synonyms and word variants were clubbed to bring similar words together. It ensured that the words with similar meaning were placed together and were not listed under variant entries.

- All ‘Antarctica’ words replaced by 'Antarctic'
- All ‘Island’ replaced by the word 'Islands'
- All 'Waters' replaced by the word 'Water'
- The Words- 'Art', 'Sp', 'Superba', 'Land', 'Late', 'Polar', 'Sub', 'Study' etc. were kept excluded from the analysis.
- No additional words were included in the top layer.

4.0 RESULTS AND DISCUSSION

The rank-ordered list of most-frequently used words is given in Table 1. 'Ice', 'Sea', 'Islands' are the most-frequently used words. Though, the word 'composition' is at the bottom of the list (Table 1), it is the most-connected word in Antarctic science (Table 1.1) with Freeman's degree centrality value of 10.56. Top 35 words were selected to generate the matrix of word-associations, which was subsequently used for network analysis.

Table 1. Most-frequently used words in Antarctic Science Subject Specialty

Total Words	13672
Total Unique Words	35
Total Episodes	13669
Total Lines	27324

Threshold	0.000
Restoring Force	0.100
Cycles	1
Function	Sigmoid (-1 - +1)
Clamping	Yes

Descending Frequency List				
			CASE	
WORD	**FREQ**	**PCNT**	**FREQ**	**PCNT**
Ice	1681	12.3	5318	38.9
Sea	1040	7.6	3683	26.9
Islands	921	6.7	3052	22.3
Water	628	4.6	2292	16.8
East	621	4.5	22.6	16.1
Peninsula	463	3.4	1698	12.4
Southern	444	2.9	1679	12.3
Species	396	2.9	1439	10.5
Krill	393	2.9	1358	9.9
Distribution	376	2.8	14446	10.6
Ocean	360	2.6	1312	9.6
Ross	359	2.6	1363	10.0
Fish	353	2.6	1246	9.1
Marine	320	2.3	1178	8.6
Ozone	294	2.2	9.3	6.6
West	291	2.1	1066	7.8
Evidence	284	2.1	1072	7.8
Surface	282	2.1	1058	7.7

Alphabetically Sorted List				
			CASE	
WORD	**FREQ**	**PCNT**	**FREQ**	**PCNT**
Bay	263	1.9	1011	7.4
Changes	226	1.7	881	6.4
Composition	228	1.7	886	6.5
Distribution	376	2.8	1446	10.6
East	621	4.5	2206	16.1
Euphausia	251	1.8	929	6.8
Evidence	284	2.1	1072	7.8
Fish	353	2.6	1246	9.1
Ice	1681	12.3	5318	38.9
Implications	264	1.9	1030	7.5
Islands	921	6.7	3052	22.3
Krill	393	2.9	1358	9.9
Lake	281	2.1	947	6.9
Marine	320	2.3	1178	8.6
Mcmurdo	227	1.7	865	6.3
Measurements	229	1.7	862	6.3
Observations	238	1.7	883	6.5
Ocean	360	2.6	1312	9.6

Descending Frequency List				
Lake	281	2.1	947	6.9
Temperature	275	2.0	1041	7.6
Implications	264	1.9	1030	7.5
Bay	263	1.9	1011	7.4
Weddell	256	1.9	981	7.2
Shelf	254	1.9	968	7.1
Euphausia	251	1.8	929	6.8
Snow	246	1.8	896	6.6
Polar	242	1.8	888	6.5
Observations	238	1.7	883	6.5
Station	237	1.7	900	6.6
Measurements	229	1.7	862	6.3
Sheet	229	1.7	879	6.4
Composition	228	1.7	886	6.5
Mcmurdo	227	1.7	865	6.3
Changes	226	1.7	881	6.4
Study	220	1.6	852	6.2

Alphabetically Sorted List				
Ozone	294	2.2	9.3	6.6
Peninsula	463	3.4	1698	12.4
Polar	242	1.8	888	6.5
Ross	359	2.6	1363	10.0
Sea	1040	7.6	3683	26.9
Sheet	229	1.7	879	6.4
Shelf	254	1.9	968	7.1
Snow	246	1.8	896	6.6
Southern	444	3.2	1679	12.3
Species	396	2.9	1439	10.5
Station	237	1.7	900	6.6
Study	220	1.6	852	6.2
Surface	282	2.1	1058	7.7
Temperature	275	2.0	1041	7.6
Water	628	4.6	2292	16.8
Weddell	256	1.9	981	7.2
West	291	2.1	1066	7.8

A four blocks model solution was found to be optimum at R^2=0.998. Table 1.1 depicts the blocks assignments of the words. The density table was dichotomized using the mean density of the table - 0.33 using the following rule (Table 1.2).

Rule: y(i,j) = 1 if x(i,j) > -0.33, and 0 otherwise.

Table 1.1 Block Assignments for Antarctic Science

Block 1	Ice, Island, Sea Water
Block 2	Euphausia (superba), Krill, Measurement
Block 3	Bay, Distribution, East, Implication, Lake, Marine Ocean, Peninsula, Polar, Ross, Sheet, Shelf, Snow, South, Species, Study, Surf, Weddle Sea, West,
Block 4	Changes, Composition, Evidence, Fish, McMurdo, Observation, Ozone, Station, Temperature

Table 1.2 Binary matrix derived from the valued matrix

	Block 1	Block 2	Block 3	Block 4
Block 1	1	0	1	0
Block 2	0	1	0	1
Block 3	1	0	1	0
Block 4	0	1	0	1

The network diagram is given in the Figure 2. Two distinct blocks have come out. The standard deviation of 0.74 indicates a wide range of its variability. The network map has generated two distinct clusters, one between block 1 and block 3, and the other between block 2 and block 4. Block 1 contains words like of 'Ice', 'island' and 'sea' 'water' while block 3 mostly identifies geographical locations, indicating prevalence of research on this subject in the stated locations like Peninsular regions, Ross islands, etc 'Changing' scenarios have been the focus of substantial amount of research.. This may be due to the worldwide concerns about 'global warming' and its relation with Antarctic ice shelf. Substantial research has been done in and around the McMurdo station of the USA, which is Antarctica's largest community[1]. USA sends maximum number of expedition members to Antarctica. They maintain a huge research base in the icy continent, and largest producer of scientific information, as evident through published papers on Antarctica Continent (Dastidar 2007a). Block 2 consists of word like 'Krill' and its scientific name '*Euphausia*' 'measurement' which is linked with the block 4 consisting of words like 'changes', 'composition', 'fish',

[1] It is built on the bare volcanic rock of Hut Point Peninsula on Ross Island, the farthest south solid ground that is accessible by ship. Established in 1956, it has grown from an outpost of a few buildings to a complex logistics staging facility of more than 100 structures including a harbour, an outlying airport (Williams Field) with landing strips on sea ice and shelf ice, and a helicopter pad. There are above-ground water, sewer, telephone, and power lines linking buildings (http://antarcticconnection.com/antarctic/stations/mcmurdo accessed on 06 may 2010).

etc. It is evident from this block modelling that there is prevalence of research on the biological resources like Krill, fish, etc in the Antarctic water.

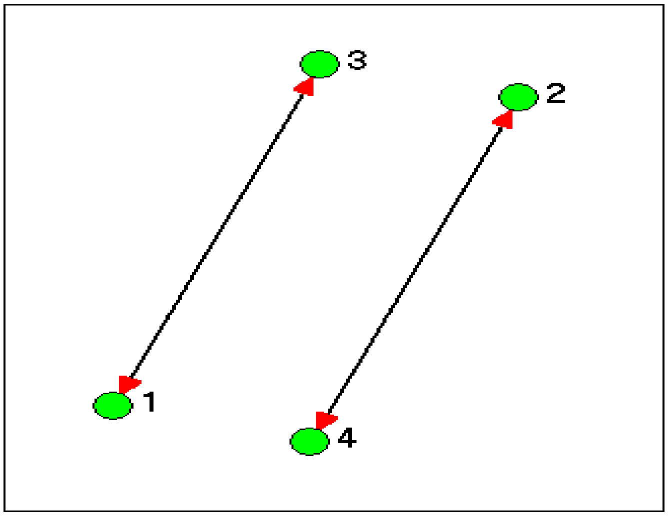

Figure 2 Network map of thematic blocks in Antarctic Science Research

Centrality of the top 35 words is given in Table 1.3. ‘Composition’ is the most-connected word with a centrality value of 10.56, followed by the words ‘Sea’, ‘Ice’, and ‘Water’, signifying its use with many other words. It is evident that considerable amount of research is underway to uncover the ‘composition’ of various attributes. ‘Peninsula’, ‘Weddle’ signifies importance of research on geographical locations.

Table 1.3 Freeman’s degree centrality, normalised degree centrality and share of centrality of words in Antarctic Science subject specialty

Sl. No.	Words	Degree	Normalized Degree	Share
1	Sea	5.586	16.428	-0.216
2	Ice	5.57	16.382	-0.215
3	Water	4.964	14.6	-0.192
4	Island	4.948	14.554	-0.191
5	East	4.935	14.515	-0.191
6	Southern	4.697	13.814	-0.182
7	Ross	4.653	13.686	-0.18
8	Implication	4.613	13.567	-0.178
9	Marine	4.607	13.549	-0.178

Table 1.3 (Continued)

10	Species	4.573	13.449	-0.177
11	Peninsula	4.562	13.418	-0.176
12	Ocean	4.558	13.405	-0.176
13	Distribution	4.545	13.369	-0.176
14	Snow	4.543	13.36	-0.176
15	Sheet	4.479	13.172	-0.173
16	Surface	4.446	13.077	-0.172
17	Lake	4.429	13.028	-0.171
18	Weddle	4.407	12.963	-0.17
19	Shelf	4.354	12.805	-0.168
20	West	4.336	12.752	-0.168
21	Bay	4.224	12.423	-0.163
22	Study	4.148	12.2	-0.16
23	Polar	4.058	11.937	-0.157
24	Composition	10.563	-31.068	0.408
25	Evidence	-10.577	-31.108	0.409
26	Temperature	-10.628	-31.259	0.411
27	Change	-10.826	-31.84	0.419
28	Mcmurdo	-10.858	-31.934	0.42
29	Station	-10.862	-31.947	0.42
30	Observation	-10.966	-32.252	0.424
31	Fish	-11.082	-32.595	0.428
32	Ozone	-11.277	-33.168	0.436
33	Measurement	-11.31	-33.265	0.437
34	Euphausia	-11.463	-33.715	0.443
35	Krill	-11.687	-34.372	0.452

5.0 CONCLUSION

The thematic analysis has been conducted using a dataset of 10,942 titles in Antarctic science. From title the unique words were identified as 13672 for Antarctic science. From the list of unique words, top 35-words were taken for the analysis. Study revealed following words as the frequently used words: 'Ice'

(1681), ‘Sea’ (1040), ‘Islands’ (921), ‘Water’ (628) in Antarctic science. The Freeman’s degree centrality and normalised degree centrality have been calculated for all the selected words in these subject specialties.

Block to block network maps have also been generated to find the linkages of the words in one block with those of the other blocks. The network map in Antarctic science has generated two distinct linkages. One was between Block 1 (having words like Ice, island, sea, water), Block 3 (having location-related words like Peninsula, Polar, Bay). The other was between Block 2 (having words like Krill, measurement) and Block 4 (having words like change, composition, fish). The first depicts the prevalence of research interests in subjects like peninsular regions, Ross islands, etc, while the second has depicted prevalence of research on the biological resources like Krill, fish, etc. in the Antarctic water.

From the analysis, it was evident that the topology of cognitive networks is subject specific. Three different studies — using Antarctic science, Ocean Science and Ocean Engineering dataset provide three different network structures. In Ocean Science the network is star shaped – each block having connection with other blocks, while one particular block (wave) dominates the network in case of Ocean Engineering (Dastidar 2007b).

*

The authors express deep appreciation and gratitude to Prof. Joseph Woelfel, Department of Communication, State University of New York at Buffalo for his guidance and support for the study.

Note: The opinion expressed in the article is of the authors, not necessarily of the department/ institute where they belong.

REFERENCES

BARABASI, A.L., ALBERT, R. (1999), Emergence of scaling in random networks. Science, 286 : 509-512.

BHATTACHARYA, SUJIT., BASU, PRAJIT. (1998), Mapping a research area at the micro level using co-word analysis. Scientometrics, 43(3) : 359-372.

BORGATTI, S.P., EVERETT, M.G., FREEMAN, L.C. (2002), UCINET for Windows: Software for Social Network Analysis. Harvard: Analytic Technologies.

BREIGER, R.L., BOORMAN, S.A., ARABIE, P. (1975), An algorithm for clustering rational data with applications to social networks analysis and comparison with multidimensional scaling. Journal of Mathematical Psychology, 12 : 328-383.

BURT, R.S. (1987), Social contagion and innovation: Cohesion versus structural equivalence. American Journal of Sociology, 92 : 1287- 1335.

DASTIDAR, P.G. (2007a), National and institutional productivity and collaboration in Antarctic science: An analysis of 25 years of journal publications (1980–2004), Polar Research.

DASTIDAR, P.G. (2007b), Science and Technology Dynamics: Mapping the intellectual structure of Antarctic Science, Ocean Science, and Ocean Engineering , Phd. thesis submitted to Anna University, Chennai, India.

DASTIDAR, P.G., PERSSON, O. (2005), Mapping the global structure of Antarctic research vis-à-vis Antarctic Treaty System, Current Science, 89(2) : 11552-1554.

DASTIDAR P.G., RAMACHANDRAN S. (2005), ‘Engineering research in ocean sector: An international profile’, Scientometrics, 65(2) : 199-213.

DOREIAN, PATRICK., FARARO, T.J. (1985), Structural equivalence in a journal network. Journal of the American Society for Information Science, 36: 28-37.

GARFIELD, E. (1968), The value of articles in bibliographic citations. Current Contents, (5), 25-26.

GARFIELD, E. (1986), ISIs master list of title words provides a special perspective on science and scholarly activity. Part 2. Comparative etymology of neologisms and research fronts. Current Contents, (28) ,214-221.

GARFIELD, E. (1990), Keyword Plus: ISI's breakthrough retrieval method, Part 1. Expanding your searching power on Current Contents on Diskette, Current Contents, (32), 295-299.

HANNEMA, ROBERT. (2006), Introduction to social network methods. http://www.faculty.ucr.edu/~hanneman/, accessed on 22 June 2006.

KRIPPENDORFF, KLAUS. (1980), Content Analysis: An Introduction to its Methodology, The Sage Context series. Sage Publications, London.

LEYDESDORFF, LOET. (2001), Full text analysis of scientific articles in: The Challenges of Scientometrics: The Development, Measurement, and Self-Organization of Scientific communication. Universal Publishers, USA.

LUSMOMBE, NICHOLAS, M., QIAN JIANG ZHANG , ZHAOLEI, JOHNSON, TED., GERSTEIN, MARK. (2002), The dominance of the population by selected few: Power-law behaviour applies to a wide variety of genomic properties. Genome Biology, 3(8).

LORRAIN, F., WHITE, H.D. (1971), Structural equivalence of individuals in networks. Journal of Mathematical Sociology, 1 : 49-80.

RAYE, K.J.H., JOHNSON, C.L., MITCHELL, M.K., NOLDE, K.J.S.F., D'ESPOSITO, M. (2000), FMRI investigations of left and right PFC contributions to episodic remembering. Psychobiology, 28 : 197-206.

WOELFEL, JOSEPH, K. (1998), CATPAC II User's Manual. Rah Press place. New York.

ZIPF, G.K. (1972), Human Behavior and the Principle of Least Effort. New York, Hafner, p.573.