Network Analysis of the Intergovernmental Panel on Climate Change

**An examination of team structure, network
architecture, and other major influences
of the
IPCC's Third Assessment Report**

MIT esd

Team members:
Travis Franck
Robert Nicol
Jaemin Song

Table of Contents

Introduction

In 1988 the World Meteorological Organization and the United Nations formed a joint organization: the Intergovernmental Panel on Climate Change. This new organization, the IPCC for short, was charged to fairly and openly assess the science and socio-economic challenges regarding climate change. (IPCC, 2006) To date, the IPCC has published three full reports: 1990, 1995, and 2001. This research project analyzes the latest report to understand the team collaboration structure of climate change researchers.

The IPCC and Its Report Process

The IPCC report is prepared through complex procedures to successfully synthesize all relevant issues related to climate change with agreement of stakeholders. Since climate change is a sensitive and controversial issue in policy and science, achieving credibility is an important and necessary goal for the IPCC report.

Figure 1 briefly shows how the IPCC report is prepared. To develop scope and outline of the report, frequent experts meeting are held. In addition, policy makers are also encouraged to address currently important policy issues related to climate change. Once the scope and outline are defined, governments are invited to nominate experts and each working group selects authors/editors from the nominated experts for the relevant volume. Authors are in charge of the contents and work based on peer-reviewed and internationally available literatures. The draft written by the authors are sent to experts for review. After authors prepare final draft based on the experts review, governments review the draft and working groups of IPCC finally approves the report and published it.

Figure 1 - The flow chart of actors for an IPCC report.

Motivation to Study the IPCC

The IPCC was chosen because it is often regarded as the definitive collection of research. UN members refer to the reports when negotiating treaties and discussing other possible courses of action. Since the IPCC report plays a vital role on the world stage, we focused our efforts to see how well the IPCC lives up to its mission of openness and fairness by:

- Exploring how team collaborations differ in the many fields of expertise (e.g., economics and atmospheric chemistry).
- Evaluating core network metrics, such as clustering coefficient and the degree distribution, for their relevance two to identifying IPCC characteristics.
- Identifying the nationality of key authors to see if they represent the goal of the IPCC to included developing and developed researchers.
- Understanding who are the most connected and least connected authors. More importantly, to understand characteristics of the least connected authors, including discipline, research labs, and nationality.

By studying the IPCC, we choose a seminal work in the climate change field, which many researchers defer to on a variety of issues.

Preparing the Data for Analysis

This project required a significant effort to obtain, extract, and clean the IPCC data. Because even minor typographic errors could skew the network analysis, careful extraction of the data was essential. Scripts were written to extract the authors from the IPCC raw text citations. Upon examination the data contained many inconsistencies, typos, and other errors. The first was to

remove all non-author text imported into author fields, such as citation journal information. Next was the separation of authors from their initials that were incorrectly joined in the source data (A.C.Fisher into A.C. Fisher). Next was the most time consuming step of placing all the data and sorting by last name. Excel formulas were written to perform string comparison and identify similar strings in order to find additional errors such as: A.C. Fisher is the same as AC Fisher; D. TanrŽ is the same as D. Tanrz; or R.S.J Tol is the same as R.S. Tol. Some disambiguation required going back to the papers cited to verify the identity of each author. Systematic rules were applied where possible such as replacing all Ž characters with z. Certain phrases still remained even after this, such as "http" or "10others" which also required significant manual editing. A final cleanup was performed after UCINET crashed with certain non-standard characters (such as ë, ö, č).

At this point a UCINET compatible file was produced in "DL" format file listing the papers and authors as a bipartite network using the form: paper12, author1, author2, etc. This file contained 14,337 lines each corresponding to a paper in the IPCC and 17,901 unique authors. UCINET was unable to correctly deal with this large data set, producing out of memory errors[1] when the entire data set was loaded, and skipping authors when attempts were made to reduce the initial memory allocated by modifying the DL input file parameters. These errors and the general extremely slow performance of UCINET with large data sets forced a change to Pajek and a reformatting of all the data into a Pajek format. To reduce the data set to its absolute minimum, repeated authors or author pairs were removed (i.e., of 10 individual papers from R.S.J. Tol only 1 was kept since the others would have been disconnected in any case). Removing these repeats preserves all connections, but enables the data set to be smaller. Pajek was able to deal with this data set without any problems.

After the network was analyzed, authors were ranked based on different network metrics. These rankings are explored in further detail in the following sections. For the Top 24 authors and the Longest Path, hours were spent researching author's institutional affiliations, educational background, and nationality. We felt this 'micro' relationship information was key to understanding the networks beyond the Pajek metrics.

The IPCC Report Network Structure

The IPCC network was analyzed similar to other collaboration network research. Similar Newman (2001a; 2001b; 2003; 2004), Li and Guanrong (2003), and Moody (2004), authors were

---

[1] This was performed on a PC with the maximum windows addressable memory installed

designated as the vertices (or nodes) and papers were the edges. If three authors published a paper together, the network graph would be a triangle, since there are three vertices each connected by a common paper.

The IPCC Third Assessment Report co-authorship network contains 17,901 unique authors[2] connected by over 71,341 edges. The average degree for a node is 7.96. As describe above, this network proved a challenge to analyze and pushed the limits of several of the software packages. Pajek proved best in the end, so many of the metrics, such as clustering coefficients and centrality betweenness, used Pajek's algorithms. Some other metrics were calculated without a network analysis package.

Figure 2 and Figure 3 are visualizations of the entire network. The sheer number of nodes makes distinguishing patterns difficult, but these drawings demonstrated that the nodes are vastly interconnected with some authors remaining on the periphery. Figure 2 is the giant connected component of the network and Figure 3 contains the remaining nodes in their significantly smaller clusters and isolates.

---

[2] The network contains 17,901 vertices, but as we have done analysis, we have noticed some data inconsistencies. While we attempted to clean the data to control for different spellings and initials, we did miss a few. (For example "John Ron Smith" maybe be "J.Smith" and "J.R.Smith".) The size of the network prevented us from catching all incidences. Alternatively, the size of the network means that these small errors most likely shouldn't change our results dramatically.
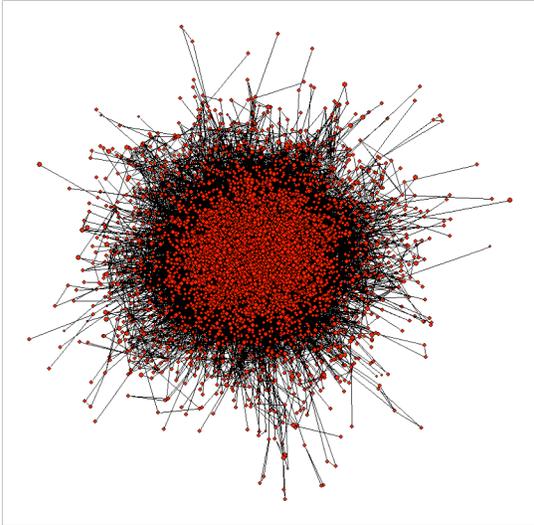
Figure 2 – The Giant Component of the IPCC author collaboration network. (k=11264)
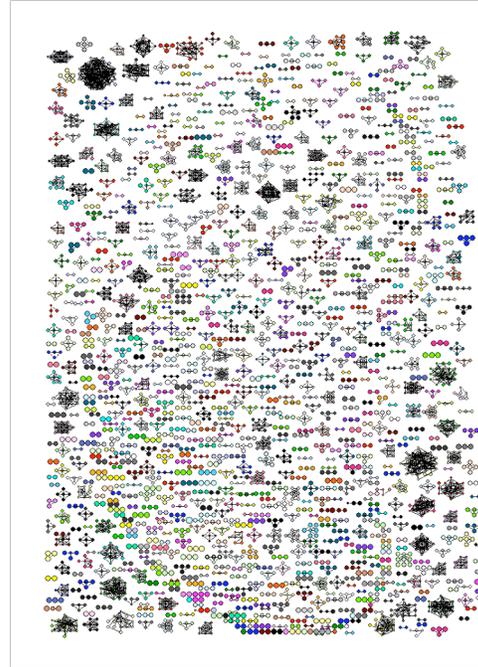

Figure 3 – Remaining Components of the IPCC author collaboration network. (k=6637)

Clustering coefficient

The clustering coefficient metric measures the interconnectedness of a network. The values range from zero to one, with one being a completely connected graph. Table 1 compares the clustering coefficient for the IPCC network with two other fields of study taken from Newman (2003) but using the local clustering coefficient (Watts – Strogatz). The clustering coefficients vary widely between global (Newman) and local (Watts-Strogatz) so it is difficult to compare the two, but in general most co-authorship networks show much higher values than the IPCC data. The IPCC data is significantly below the ranges in the literature for co-authorship networks with a value of 0.022 (local). Authors in the IPCC report are even less connect than Newman's least connected research field.[3]

This conclusion can be rationalized by the nature of the IPCC report. The entire network contains many different authors writing on varied subjects such as insurance, upper atmospheric chemical reactions, and deep ocean currents. With such a broad set of topics, a completely connected graph is unrealistic. The clustering coefficient value highlights, though, that even though climate change

---

[3] Two principal formulas are used in the literature to calculate clustering coefficient, Newman's triangle counting method and the Watts – Strogatz average of local values for clustering coefficients. The Pajek CC1 value corresponds to the Watts-Strogatz method and is the value reported here. It also corresponds to C(2) in Newman's table.

requires many different types of expertise to analyze the problem, researchers are working in a fragmented manner and not collaborating with one another.

	IPCC Cluster Coeff. (Pajek's CC1)	IPCC Principal Component (CC1)	Biology (Newman, 2003)	Physics (Newman, 2003)
Result	0.022	0.031	0.60	0.56

Table 1 - Comparison of the clustering coefficient with two other fields.

Power Law check

A network follows a power law, with respect to node degree, if a small number of nodes have a very high degree and a high number of nodes have a low degree, with an exponential decay in between. In the real world, this may occur if a prominent researcher is invited to collaborate on an increasing number of papers as they become increasing prominent, and hence are preferentially chosen over other researchers. Power law trends have been found in social networks. (Newman, 2003)

To check if the IPCC report has a power law trend, we plotted the degree distribution on a log-log scales. (Figure 4) The distribution does not show a power trend. Papers selected for the IPCC report don't show the power law nature, but this does not disprove any prediction that real-world collaboration networks are scale-free. The IPCC process should select a representative sampling of papers and views, which might give more weight to less-frequently published authors. Thus even if the author network grows scale-free in journals, the might not be reflected in the IPCC co-authorship network.
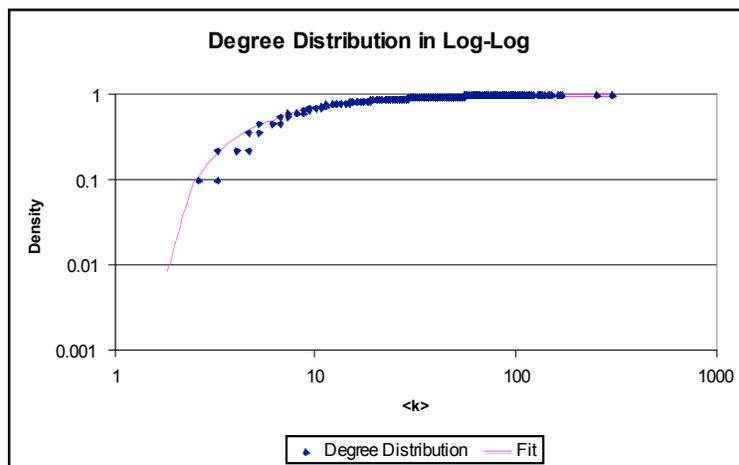

Figure 4 - Log-log plot off the degree distribution to verify that the network is not scale-free.

Discussion of Top 20 Authors

Even though the IPCC collaboration network doesn't show a scale-free property, some authors are more connected and important than others. The 17,901 authors were ranked using "centrality betweenness", a measure of their importance in the network for connecting other authors to one another. Table 2 show the 24 highest-ranking authors from the IPCC report.

Some interesting trends can be seen from the data. First, the majority of the authors are from natural science fields. Environmental and atmospheric scientists dominate the list. There are only two economists in the list: Richard Tol and Joel Smith. One explanation could be that climate change started as a scientific issue and economists are relatively new to the issue, and hence are less senior and connected in the field. Another explanation is that economists often publish on many topics, not just climate change. Atmospheric researchers might instead focus on greenhouse gas properties and publish multiple papers, increasing the chance of being cited by the IPCC. Even environmental economists often focus on many environmental issues as opposed to only focusing on climate change costs or policy implications. (Jacoby, 2006)

The authors with highlighted names[4] (Table 2) ranked in the Top 24 when sorting the author list by node degree. It shows a significant overlap between the metrics of centrality betweenness and node degree, though they do have differing results. Centrality betweenness is most often used in social networking analysis. (Magee, 2006) Scientists dominated the node degree rankings again. This could be because of the nature of the IPCC report. The report is divided into three volumes, roughly speaking: climate science; climate science, ecology, and economics; and modeling and economics. The science-based chapters are longer than the economics and policy chapters. This means that more scientists are cited and their node degree scores ought to be higher.

Most of the researchers in Table 2 work in institutions that have large budgets or can tap a large number of researchers. For instance, the Tyndall Centre in the UK is a research lab associated with several universities. They employ researchers directly but also work with professors at the host universities, which are likely salaried by the university, not the Centre. This allows them to collaborate with one another and creates a presence that is persistent, attracting experts from all around the world. When analyzing the various IPCC Report volumes below, we see that the developing countries aren't represented as well as the developed countries. These well funded research laboratories maybe one of the reasons.

---

[4] Note: These highlighted names may not print in black and white. Please refer to PDF file. The authors highlighted are the first eight, the tenth, the twelfth, and the twenty-second.

There is one further explanation how these Top 24 authors became important: they are modelers and synthesizers. Some of the high scoring authors (e.g., R.Leemans, T.M.L.Wigley and M.Hulme) work on models that are used by other researchers in the field, even researchers in other labs. These authors often get invited to as co-authors on papers that use their models. Also, since their models normally synthesize economics and climate systems into one model, they publish over a broad range of climate change issues.

Author	Centrality Betweeness	PhD	Year	Discipline	Affiliation 1	Affiliation 2
T.M.L.Wigley	0.02930	U of Adelaide	1967	Physics	U of Waterloo	Tyndall Centre
R.Leemans	0.01589	Uppsala University		Ecology	Wageningen U	UK AEA
C.Rosenzweig	0.01510	U of Massachusetts	1991	Environmental Science	NASA	RIVM
J.F.B.Mitchell	0.01465	Belfast	1973	Physics	Hadley Centre	
J.E.Penner	0.01402	Harvard	1977	Mathematics	LLNL	UCAR
I.C.Prentice	0.01373	Cambridge	1977	Environmental Science	U of Bristol	U of Michigan
P.D.Jones	0.01132	U of Newcastle	1977	Environmental Science	U of East Anglia	Tyndall Centre
M.E.Schlesinger	0.01089	UCLA	1976	Atmospheric Science	U of Illinois U-C	
J.A.Patz	0.01081	Case Western	1987	Molecular Biology	Johns Hopkins	IPCC
M.Heimann	0.01068	U of Bern	1982	Biogeochemistry	Max Planck Institute	NCAR
F.S.Chapin	0.01035	Stanford	1973	Biology	U of Alaska	IPCC
A.HendersonSellers	0.00989			Atmospheric Science	NSF	NAS
R.S.J.Tol	0.00882	Vrije	1997	Economics	Energy Economics (J)	ANSTO
M.Hulme	0.00855	U of Wales	1985		Tyndall Centre	Carnegie Mellon
S.H.Schneider	0.00787	Columbia	1971	Physics	NCAR	Tyndall Centre
K.P.Shine	0.00770			Meteorology	U of Reading	MacArthur
Y.Zhang	0.00769	U of Washington	1996	Atmospheric Science	IPCC	
G.Marland	0.00755	U of Minnesota	1972	Geology	Indiana State U	IPCC
F.Giorgi	0.00749	Georgia Tech	1986	Physics	NCAR	Oak Ridge Natl. Lab
W.H.Schlesinger	0.00706	Cornell	1976	Biology	Duke	ENER (Italy)
S.Brown	0.00700			Biogeochemisty	Hadley Centre	
M.Weber	0.00691				Technical U Munich	UK AEA
D.Rind	0.00638	Columbia	1976	GCM	Earth Institute	Columbia
J.B.Smith	0.00626	M.S. Public Policy	1982	Economics	EPA	Stratus Consulting

Table 2 - Top 24 IPCC Third Assessment Report authors ranked by centrality betweenness.

Discussion of Longest Path

Using Pajek, we found that the "longest shortest" path in the network was nineteen edges. We researched the affiliations and research areas for each of these twenty authors, as summarized by Table 3. The chain of connections has some interesting characteristics: First, the nationalities of the countries flows from a small, developed country (Sweden) through some US researchers, and then briefly to some US-trained Brazilians. The Brazilians collaborated with more Americans, who worked with a Canadian. Finally, Russian researchers worked together with the Canadian professor. Though this is only one chain, one conclusion from the flow is that the least connected countries in the world are smaller or slightly off the world stage (Sweden and Russia). Also, the researchers of these countries are connected through their ties to developed countries, primarily the United States.

Another characteristic of interest is the connection of disciplines. The Swedish researchers were working on negotiations strategies. They in turn worked with individuals from economics and biodiversity, some of who work at the United Nations (i.e., the UN Millennium Ecosystem

Assessment). From biodiversity, the expertise chain followed energy, atmospheric science, plant science (impacts of atmosphere change on crops), to soil chemistry, to soil core sampling techniques (to study historical climate change records), to artic ice core sampling, and then finally the Russian researchers who were interested in Arctic fish species and were part of an international arctic research program.

This domain or expertise is consistent with the centrality measurements above. Researchers working in the sciences, as opposed to economics and social aspects like negotiations, are more important in connecting researchers in the the IPPC network. Figure 5 shows that the researchers in the middle of the chain are more central to the network as a whole. W.H.Schlesinger and G.Marland are two researchers from the Top 24 list above.

	Author	Full Name	Nationality	PhD	Field/Year	Discipline	Affiliation 1
1	B.Aniansson	Britt Aniansson	Swedish*			Negotiations	Planning and Coordination of Research
2	U.Svedin	Uno Svedin	Swedish*			Negotiations	Planning and Coordination of Research
3	C.Folke	Carl Folke	Swedish*			Economics	Royal Swedish Acader
4	M.Gadgil	Madhav Gadgil	Indian	Harvard	Biology*/1969	Ecology	Indian Institute of Scier
5	J.A.McNeely	Jeffery McNeely	US	UCLA	Anthropology/196X	Biodiversity	IUCN
6	W.V.Reid	Walter Reid	US	U of Washinton	Zoology/1987	Biodiversity	Director of the Millenni
7	J.Goldemberg	Jose Goldemberg	Brazilian	U of Sao Paulo	Physics	Energy	Instituto de Eletrotecni
8	J.R.Moreira	Jose R. Moreira	Brazilian	U of Sao Paulo	Physics		Instituto de Eletrotecni
9	G.Marland			U of Minnesota	Geology/1972	Geology	Indiana State Universit
10	I.Fung	Inez Fung		MIT	Meteorology/1977	Atmospheric Science	UC Berkeley
11	P.A.Matson	Pamela A. Matson	US*	Oregon State	Ecology/1983	Earth Science	UC Berkeley
12	W.H.Schlesinger			Cornell		Biology	Duke
13	S.A.Prior	Stephen A. Prior	US*	Auburn	Agronomy/1993	Earth Science	National Soil Dynamics
14	B.A.Kimball	Bruce A. Kimball	US*	Colorado State	Ecology/1997		USDA
15	M.Johnson	Mark Johnson	US*	Cornell	Soil Chemistry/1986	Earth Science	EPA
16	J.King	John W. King	US*	U of Minnesota	Geology/1983	Oceanography	University of Rhode Isl
17	D.J.Noakes	David L. G. Noakes	Canadian*	UC Berkeley	Zoology/1971	Zoology	University of Guelph
18	V.V.Ivanov	Vladimir Ivanov	Russian*	St. Petersberg	Oceanography/1992	Oceanography	University of Plymouth
19	A.P.Nagurnyi	Andrei P. Nagurnyi	Russian*			(arctic fish/ocean)	AARI
20	P.A.Abaza					(arctic fish/ocean)	

Table 3 - The details of the 20 authors that make up one of the longest paths.
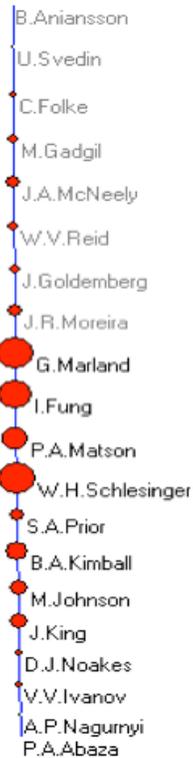
Figure 5 - The longest path drawn such that node size corresponds to the centrality betweenness metric.

Discussion of Field of Expertise Differences

In this section, we investigate the characteristics of collaboration networks in an individual field of climate change research by looking at each volume of the IPCC report. As previously mentioned, the IPCC report consists of three volumes and each volume describes different topics. In general, Volume 1 is pure science, Volume 2 is a mixture of science and policy and economics, and Volume 3 is primarily policy and economics of climate change, according to the contents of each volume. Therefore, the analysis of each volume will enable us to understand the different behaviors of research collaboration across science and social science as a whole as well as each field of science and social science related to climate change.

Basic statistics on collaboration behavior

Moody (2004) shows that social scientists tend to collaborate less than scientists due to the nature of the field, even though collaborations in social science are increasing over time. This is because social scientists are rarely dependent on large-scale lab or resources compared to pure science research. We also found similar results in the IPCC network.

As shown in Table 4, scientists have the highest average number of collaborators per author, while, on average, social scientists have the smallest number of collaborators. The average number of collaborators refers to the scope of their close researchers whom they indeed work with. When we consider collaborators as core team members, we can see that scientists have three times many members per team than social scientists do.

Not only do social scientists work with small numbers of people, but also have a propensity to work alone. The result is that almost 20% of papers in Volume 3 is written by a single author while only 3% is by a single author in Volume 1. In terms of the number of authors per paper, Volume 1 still has the largest average number of authors per paper while smaller numbers of authors work together per paper in Volume 3. Therefore, we can conclude that scientists collaborate more often than social scientists do and, furthermore, scientists also work with a more diverse range of researchers.

The reason why environmental scientists have different collaboration patterns from social scientists and vice versa is mainly due to the nature of the field as mentioned above. Most science research related to climate change requires large-scale experiments or data collecting procedures, which inevitably leads to collaborations among researchers. On the other hand, social science studies tend to be theoretical or qualitative, which does not necessarily require large-scale collaboration. Also, as mentioned before, economists or policy makers rarely work exclusively on climate change issues. Thus, few common collaboration networks have been established in this particular field. Contrarily, we can find many large scientific research groups and scientists who primarily study climate change.

	Number of collaborators per author	Number of papers with a single author	Number of authors per paper
Volume 1	11.52	209 (2.97%)	3.78
Volume 2	6.06	1069 (10.61%)	2.83
Volume 3	4.14	669 (19.22%)	2.23

Table 4 – Basic statistics on collaboration in each volume.

Distance and clustering

Short distance path and high clustering are main features of the small world network, which is commonly observed in many social networks including collaboration networks such as the IPCC network. We studied the distance and clustering patterns that appeared in the IPCC network.

Table 5 shows that the collaboration network of scientists has the shortest average path length. The difference in the average path length is not significant over the three volumes. On average,

any two researchers in a particular volume can be reached (if they are reachable each other) by tracing across 5 or 6 researchers on the path connecting the two. Meanwhile, the percentage of unreachable pairs appears to be very different over the three volumes. In Volume 3, only 23% of pairs are reachable, with 77% of pairs remaining unreachable. The larger percentage of unreachable pairs implies that fast and/or smooth knowledge dissemination might not be easy in the social science research network, with many isolated groups and people.

	Average path length	Unreachable pairs (%)	Longest path length
Volume 1	4.85	34.50	13
Volume 2	6.09	74.45	24
Volume 3	5.98	86.96	15

Table 5 – Path length statistics.

Our result on the shortest path length distributions shown in Figure 6 describes an interesting feature of collaboration over different fields related to climate change. For Volume 1 and Volume 3, which are pure science and social science related to climate change respectively, all reachable pairs can be reached within 13 or 15 steps of researchers to link each other. However, Volume 2 has a long tail, with some pairs having very long path lengths. This indicates that a longer path length is needed to link two researchers from different fields, one from science and the other from social science. Considering the path length as time needed to transfer knowledge in this research collaboration network, the graphs clearly show that it takes almost twice as long as time to disseminate knowledge or information cross the different fields than in one individual field. In climate change research, the importance of collaborations over different fields cannot be overestimated considering its complex nature and significant impact on policy and economics of the entire world. In this sense, it is desirable and necessary to have a shorter longest path length in Volume 2.
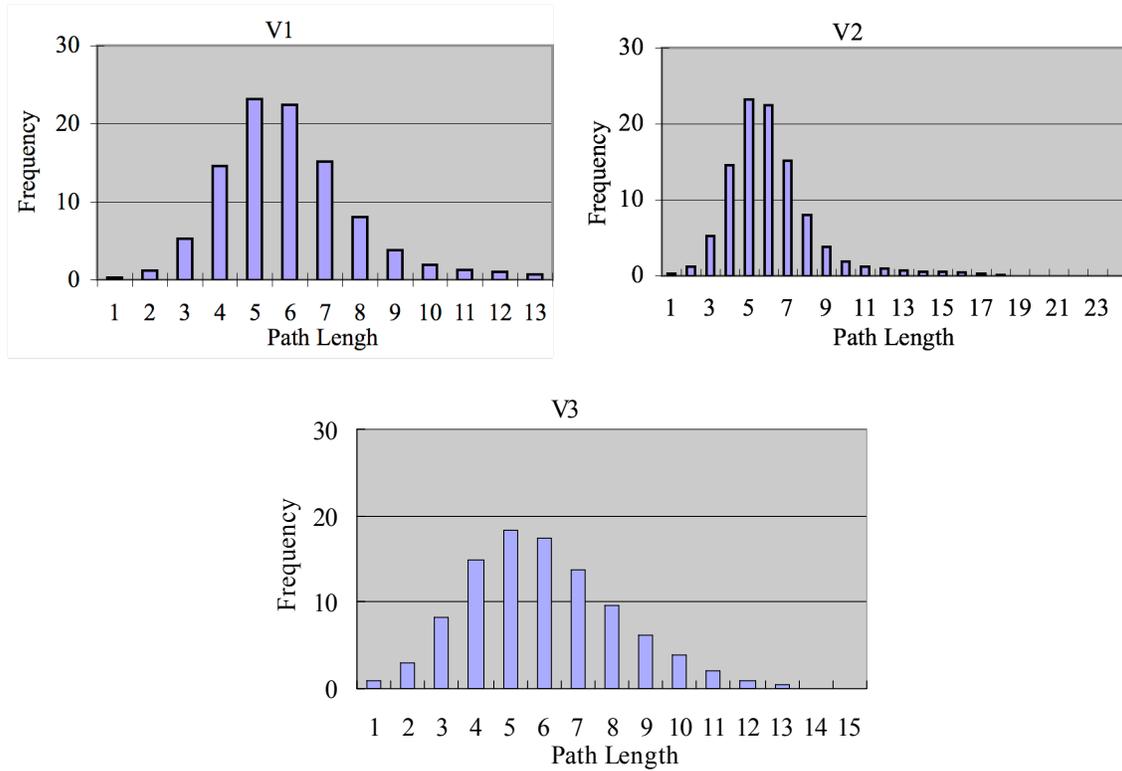
Figure 6 – Path length distributions.

The characteristics of research collaboration can be also found in different structures of clustering in the three volumes. The largest cluster of the science network in Volume 1 accounts for 80% of the network, while the largest cluster in Volume 3 covers only 36%, which is consistent with our findings on the different portions of unreachable pairs in the three volumes we observed above. Both the small coverage of the largest cluster and the low cluster coefficient explain why we have such a high percentage of unreachable pairs in Volume 3. Newman (2001) suggests that the largest cluster usually accounts for 80 ~ 90% of the network in the scientific research collaboration. The largest cluster of the scientific collaboration network in the IPCC report also falls in this range. On the other hand, social science network is relatively sparse and has a number of small size clusters, which confirms the independent collaboration patterns in social science again.

	Cluster coefficient (Pajek's CC1)	Largest Cluster (%)
Volume 1	0.0468	80.92
Volume 2	0.0377	50.49
Volume 3	0.0392	36.02

Table 6 – Clutering coefficeints.

15

Centrality

We measured centrality of the IPCC network using key centrality metrics. The three centrality metrics are betweenness, degree, and closeness centrality. All three metrics are used to measure the level of centrality of the network with slightly different implications: betweeness centrality is related to power by being an intermediary in information flow; degree centrality is related to power by having the most connections; and closeness centrality is related to power by having short distance to other nodes (Magee, 2006). As we expected, the science network is the most centralized network according to all three metrics and the other two volumes have rather decentralized networks. This result is not surprising, recalling our findings on different collaboration patterns between scientists and social scientists where social scientists tend to work independently and, furthermore, there are not many interactions among research groups, while scientists do indeed collaborate with each other. The centrality measures are similar for Volume 2 and Volume 3, while the centrality measures for Volume 1 is twice as much, as shown in Table 7.

	Betweenness	Degree	Closeness
Volume 1	0.045	0.023	0.237
Volume 2	0.024	0.009	0.167
Volume 3	0.028	0.014	0.143

Table 7 – Centrality metrics.

Collaboration network in other studies

Many studies have been done to investigate research collaboration or co-authorship patterns in various research fields. Among them, we will compare the IPCC network with co-authorship networks studied by Newman (2004) and Moody (2004). Newman studied several collaboration networks in science while Moody focused on the analysis of social science networks. Table 8 is the summary of their analysis on the different networks.

Generally, their results are similar to our findings that scientists are more collaborative than social scientists. In particular, the Biology network seems to be very similar to the IPCC science network except that the Biology network has a slightly bigger largest cluster and more collaborators per author. The average number of authors per paper and the average path length in Biology are almost the same as we observed in the IPCC network.

Meanwhile, Moody's Social Science network is slightly different from the IPCC social science network. Much higher percentage of papers, almost three times as much as in the IPCC social science network, was written by a single author in Social Science, which accounts for 67% of the papers studied. In addition, the Social Science has an average path length two times longer than

the IPCC network. All of these results suggest that social scientists tend to conduct research more independently and collaborate much less than we observed in the IPCC social science network. There are two possible explanations. The first explanation is that Moody's study on Social Science network includes all sub-fields in social science, which results in a sparse and fragmented network. Another explanation can be attributed to the nature of climate change research, which requires relatively large-scale analysis compared to other social science areas, so that this particular IPCC social science network is more collaborative than the usual social science network.

Interestingly, none of the papers calculated any centrality metrics, which we found very important and useful components in the analysis of social networks, such as co-authorship networks.

	Newman (2004)		Moody (2004)
	Biology[5]	Physics[6]	Social Science[7]
Number of collaborators per author	18.1	9.7	-
Percentage of papers with a single author	-	-	66.8%
Number of authors per paper	3.75	2.53	1.64
Average path length	4.6	5.9	11.53
Largest cluster	92%	85%	42%

Table 8 – Statistics of collaboration networks from other literatures.

Examination of Political Influence

Since climate change is a very hot topic with many controversial issues from science to policy, we can expect to have high likelihood of the IPCC report having political influence. In this section, we investigate two factors, nationality representation and influence of editors, which presumably reveal political influence or biases embedded in the IPCC report, if there are any.

Nationality representation

Due to the time constraint, we focused on 20 authors from each volume who are ranked as the top 20 according to the betweenness centrality metrics. We explored web pages to determine their nationality. Figure 7 shows their nationality distribution. Out of sixty authors, there is only one researcher from a developing country. In contrast, approximately 77% of researchers are from either the US or UK. In general, we can see large numbers of researchers from developed countries in any academic research field. However, this result, one out of sixty, is still more than

---

[5] Co-authorship network from the Medline from 1995 to 1999.
[6] Co-authorship network from "Physics E-print Archive" from 1995 to 1999.
[7] Co-authorship network from "Sociological Abstracts" from 1989 to 1999.

what we expected, though we acknowledge the trend. This uneven nationality distribution of top 20 authors indicates a high possibility of underestimation of the needs of developing countries and overrepresentation of specific developed countries in the IPCC report.
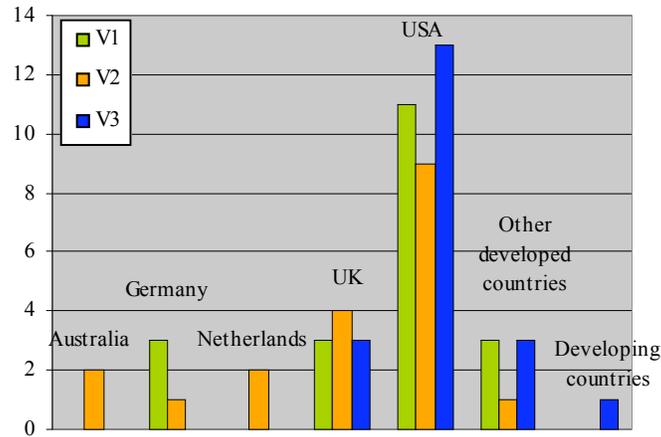

Figure 7 – Nationality distribution of top 20 .

Influence of editors

Since editors of the IPCC report are those who select relevant journal articles and write its contents, we cannot exclude the possibility of their significant influence on the report. To investigate their influence on the report, we counted the number of editors among the top 20 and top 50 of each volume. We listed top 20 and 50 authors according to betweenness centrality. Indeed, many of the editors are in top 20 or top 50, as shown in Table 9. The editors in Volume 1 and Volume 2 also appear to have great power on the network by having approximately 10% of editors in the top 50. In particular, Volume 3 turns out to be the most influenced by editors, which has 20% of editors, twice as many as Volume 1 and Volume 2, in top 50. This is consistent with comments made by Jacoby (2006) and Reilly (2006) concerning the IPCC report. Jacoby and Reilly were involved in the IPCC preparation process as editor and reviewer, respectively. They pointed out that Volume 3 was the most politically controversial volume and there were even many debates and disagreements among the editors. Due to the nature of the contents in Volume 3, it has many sensitive issues regarding international policies, equity, and economics of countries related to climate change. Thus, many governments tried to address their own interests in the report, which might appear to have as many editors as key players in the network.

	Number of editors	Number of editors in top 20	Number of editors in top 50
Volume 1	139	9 (6.5%)	18 (12.9%)
Volume 2	199	9 (4.5%)	19 (9.5%)
Volume 3	116	11 (9.5%)	23 (19.8%)

Table 9 – Number of editors in the IPCC report.

Following the nationality analysis of the top 20 authors in the previous part, we also studied the nationality distribution of editors. Figure 8 shows relatively balanced nationality distribution with around 30% of editors from developing countries and the rest from developed countries. We might think not all the editors have the same level of control over the report. Therefore, it is meaningful to focus on the key editors who have more influence on the network than others. We identified editors who are listed among top 50 authors according to betweenness centrality. Figure 9 provides the nationality distribution of those key editors in each volume, which tells a somewhat different story from what we see in Figure 8. We see very few editors from developing countries in the top 50. This means that even though we could find a number of editors from developing countries on the editor list, many of them are not influential key players in the network. Rather, many of them might have been selected owing to political reasons just because they are from developing countries or they do not have strong power at the international level in spite of their substantial expertise in the relevant field. This again questions a possibility of under-representation of the needs from developing countries.
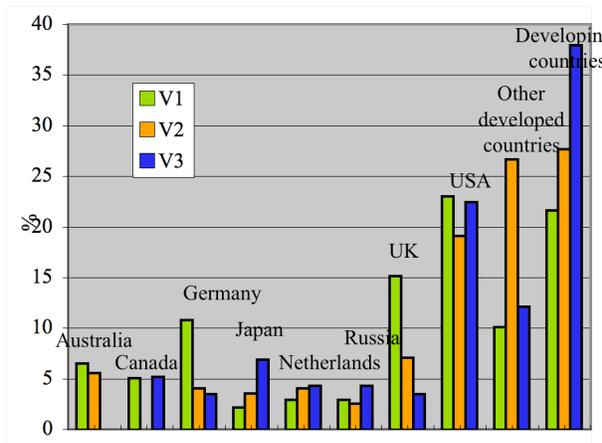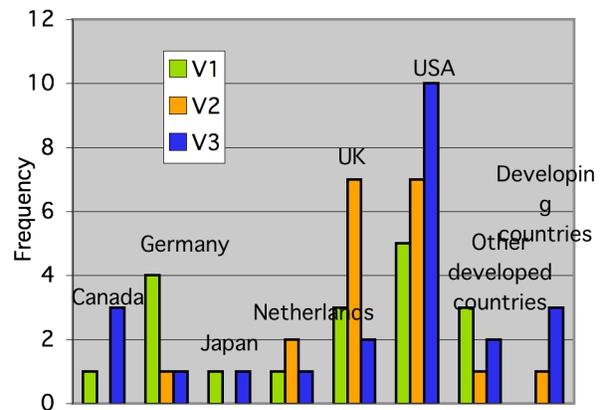

Figure 8 – Nationality distribution of editors.

Figure 9 – Nationaltiy distribution of editors in Top 50.

Potential Architectural Improvements and Lessons Learned

Helping the IPCC reach its mission

The IPCC, as part of the UN, is charged to represent the interests of the developing countries. Our analysis shows that researchers from developing countries might not have the influence that their counterparts in the developed countries have. While they have a similar number of total editors, they have an under-representation of editors in the top 50. Developing country authors and editors are not 'central' to the network. One key reason for this is the lack of capacity in developing countries. These countries don't have many dedicated research laboratories. If they do have a lab, it is often focused on the local conditions and not of general interest to the global community of researchers. (Jacoby, 2006)

One solution to this is to increase the capacity can capabilities of developing countries to conduct climate change research. This is important in the long-run, but there might be a short-term solution using our network analysis above. If developing country researchers worked closer with the key authors above, those most central to the international community, then their views and research might have more impact throughout the world. The IPCC would better achieve it mission of including all views from various parts of the world.

MIT's rankings

MIT's Joint Program on the Science and Policy of Global Change is one of the leading climate change labs. The co-directors of the lab, Professor Prinn and Professor Jacoby, however, aren't among the Top 24 authors. In fact, they rank 875 and 1196, respectively. (Table 10) While not ranking high, their relative rankings with one another match the fundamental trend we found with other authors: the scientist (Prinn) ranked higher than the economist (Jacoby).

Jacoby (2006) hypothesized why MIT might not score high in the overall rankings. First, MIT often tries stay out of the IPCC process because it is considered very time intensive. MIT researchers would prefer to continue working on their own projects as opposed to the IPCC report chapters. This means that, if editors have influence on citations, then MIT would be cited less.

Another similar reason MIT might not rank high is that the Joint Program's models aren't reviewed by the IPCC. Several key models were chosen by the IPCC and MIT's were not used. This means that many MIT papers weren't included in the report. Overall, Reilly (2006) thought that Volume 3 was a "civil war" and MIT's economics papers weren't often cited because they just didn't play the political game.

Author	Centrality Betweeness	Rank	PhD	Year	Discipline	Affiliation 1
R.Prinn	0.000470975	875	MIT	1971	Atmospheric Science	MIT
H.D.Jacoby	0.000328577	1196	Harvard	1967	Economics	MIT

Table 10 - The rankings of MIT's Joint Program on the Science and Policy of Global Change co-directors.

Domain knowledge/expertise

An important lesson that we always returned to was the value of domain knowledge. When analyzing a social network, such as a co-authorship network, it is important to have some familiarity with the network. Our initial IPCC knowledge was adequate to start the analysis, but we quickly had to augment this with readings and discussions with IPCC participants (i.e., Professors Jacoby and Reilly).

Domain knowledge was required to make sense of the metrics generated by Pajek and other formulas. For instance, Pajek generated the Top 24 authors list, but domain knowledge was important into understanding who these authors were, what they studied, and why their fields were important to the climate change issue. Their roles in major labs and their roles as synthesizers, etc. required additional information than Pajek could possibly have analyzed. Many of the hypotheses mentioned above in this paper were rooted in domain knowledge.

"Micro" relationship data insights

Co-authorship relationships have been explored in numerous papers, but the connection to the micro level in terms of identifying other layers of connection between the authors appears to be relatively rare. For the IPCC network, we found additional layers such as nationality, field of study, research lab, and chapter editor status. These data are critical since one of the initial motivations for this project was to better understand the team structure and motivation in the IPCC work. Many organizations form temporary collaborations of the sort captured by a co-authorship, but few leave a trace as a paper does. The IPCC data is somewhat skewed in terms of understanding the organizational dimension of system architecture because the intent was to produce as broad a set of papers representing climate change views. But even here significant structure was found. The comparison to other systems is of perhaps best focused on news media and their coverage of issues. In much the same way as the IPCC, the presumed mandate of CNN and others is to provide a broad and balanced coverage of issues facing the world. In this sense the low clustering coefficient of the IPCC data is actually a figure of merit since it can be interpreted to show that a broad and representative set of views was collected. Applying the same technique to news media authors would likely show a much higher clustering and much less

"balanced" reporting. The exact same analysis would yield fascinating insight into who are the most influential editors, reporters, and news media. In addition, the news media stories could be grouped by key sources instead of authors, such as individuals providing "news" to the media. This again would yield a very valuable insight into who the most influential news sources are and how they are connected. This is but one example of the potential uses of the methodology developed here with only slight variations, and why network science is such a powerful tool.

Lessons from the tools

A key learning from this project is that data extraction tools are critical to the advancement of network science. A very significant component of the work in this project was in extracting and preparing the data for use in Pajek. This project emphasized text data, which is probably best handled in database software rather than processing through Excel, but again the gap between data and tools is where most of the effort goes. This project attempted to extract much more nuanced data as to the educational background, institutional affiliation, and funding source for all 17,901 authors, but there were few computer searchable sources of this data. The project certainly demonstrated that even without lots of extra data, but simply the affiliations on papers the key players in a large scientific endeavor can be quickly identified. But other patterns may emerge when other dimensions of the data are added, for example clustering around certain particularly influential educational institutions where key players were trained. One hope is that as network science expands more data sets will become available that can be cross-referenced greatly enhancing the network approaches discussed here.

The scale of the data is also an issue since "mainstream" network tool such as UCINET cannot handle large data sets but is relatively easier to use than Pajek. Many of the text-based data sets of interest to the ESD community are large (i.e., comparing global engineering standards) and analysis tools that can handle the data efficiently will be important (in addition to the pre-processing described earlier). Using well understood "test" sets was found to be very useful in verifying and understanding the large scale results, since large data sets make it impossible to check the validity or correctness of every individual node relationship. Another key learning is that the language of networks is still evolving, and great care must be taken to ensure practitioners are comparing the same metrics as in the case of the local vs. "global" clustering coefficient that are sometimes not explicitly labeled.

References:

IPCC. 2006. "About the IPCC." <http://www.ipcc.ch/about/about.htm> viewed on 10 May 2006.

Jacoby, J. 2006. Personal communications on several occasions in April and May.

Li, C., and C. Guanrong. 2003. "Network connection strengths: Another power-law?" *Paper: City University of Hong Kong*. <http://arxiv.org/abs/cond-mat/0311333> on 24 Apr 06.

Magee, C. 2006. Class 6 from Advanced System Architecture class (ESD 342).

Moody, J. 2004. "The structure of a social science collaboration network: Disciplinary cohesion from 1963 to 1999." *American Sociological Review*. **69**, pp213-238.

Newman, M.E.J. 2001a. "Scientific collaboration networks: I. Network construction and fundamental results." *Phys. Rev. E*, **64**, 016131.

Newman, M.E.J. 2001b. "The structure of scientific collaboration networks." *Proc. Natl. Acad. Sci.,* **98**, pp404–409.

Newman, M.E.J. 2003. "The structure and function of complex networks." *SIREV*, **45**(2), pp167–256.

Newman, M.E.J. 2004. "Co-authorship networks and patterns of scientific collaboration." *Proc. Natl. Acad. Sci.,* **101**(sup1), pp5200 –5205.

Reilly, J. 2006. Personal communications on several occasions in April and May.