Scientific Co-authorship Network at IITM

Rahul Ramesh(CS14B061)*, Ashutosh Kumar Jha(ME14B148)†, Parikshit S Hegde(EE14B123)‡, Aravind Shankar(EE14B012)‡, Sarthak Peshwe(EE14B126)‡ and Prasanth BJ(EE14B131)‡
* *Computer Science and Engineering,IIT Madras*
† *Mechanical Engineering, IIT Madras*
‡ *Electrical Engineering, IIT Madras*

***Abstract*—The analysis of scientific co-authorship networks provides valuable insights into the flow of information among various publishing individuals, and also sheds light on inter-departmental research and knowledge transfer. In this document, we analyze the collaboration scenarios in various departments of IIT-Madras, along with their visualization in the form of networks individually and collectively; the analysis is based on network metrics that are widely used. Co-authorship being a direct measure of collaboration between the participating entities gives a deeper understanding of the social interactions and the exchanges between them.**

1. Introduction

The co-authorship network is a network with the authors at the nodes(or vertices) and links(edges) connecting them if they have co-authored a publication. These form the network representation of the co-authorships in an institute. Visualisation of this network gives an insight into the nature of collaborations within the departments and even between them. Also, the network metrics carry useful information on the same.

Adjacent nodes (co-authors) have one edge connecting them in case of collaborations between the authors. If the number of documents involving co-authorship of adjacent nodes is more than one, the colour of the edge is different, and the link or edge is undirected as it involves collaborations from both parties. This makes it an undirected weighted simple network.The colour coding for the links is given as : Yellow, for the low end of the spectrum and Red, for the higher end of the spectrum in terms of collaborations.

We have modelled the Coauthorship Network as a Social Network, such as a network of friends. Here, collaboration is analogous to friendship in the Friends Network. Evidence suggests that in most real-world networks, and in particular social networks, nodes tend to create tightly knit groups characterised by a relatively high density of ties; this likelihood tends to be greater than the average probability of a tie randomly established between two nodes [1] [2]. We will examine if the coauthorship network too follows such trends.A clustered group would denote a well connected department whereas sparse connections would imply lesser inter-departmental collaboration.

Tasleem A. *et al* [3] have previously studied Scientific Coauthorship Network for the CS departments in some of the top IITs in India. They analysed the intra- and inter-institute networks and drew some conclusions from the networks.For example, they came to the conclusion that this network exhibited a lot of the properties seen in Social Network. In our work we add upon this previous work by extending it to all the departments in IIT Madras. We would like to see if the IITM network also exhibits similar properties.

This document is organized as follows : We first talk about the procedure we underwent for data acquisition, followed by the extraction of the co-authorship collaboration network. On obtaining the network representation and graph, we put forward the metrics used for analysing the collaboration network. Based on these values, we then mention the observations and note down the results and inferences. Finally, we conclude the data and its analysis and provide the references.

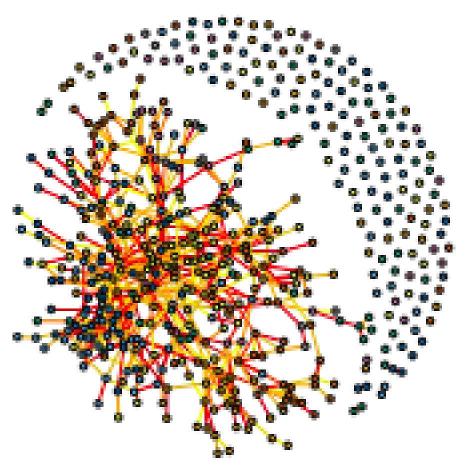

Figure 1. Co-authorship network of IITM Professors

2. Data Acquisition

In this project, the first step we took was to collect the data for the co-authorship network, which was the list of number of coauthored papers between pairs of professors of

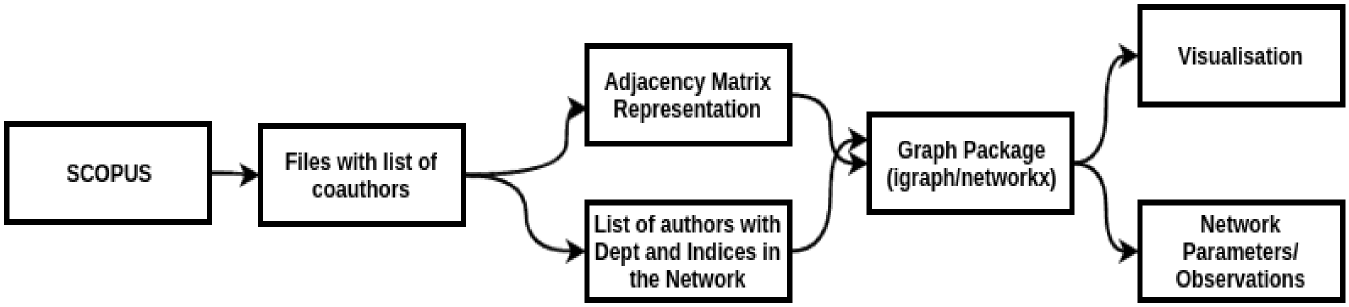

Figure 2. Program Workflow

IIT Madras. We collected the details of the professors from all the Science-Stream departments(i.e., all departments except Humanities and Management). Publication lists are available at several databases like Google Scholar,DBLP, homepages of authors etc. Since we were interested in the number of co-authored papers only, and not the details of the publication, the data in the above mentioned databases was not in a user-friendly format for our purposes.

The database on Scopus [4] contained the data in the format we required under the co-authors tab of each authors profile. We made attempts to use a web-crawler using a Python API [5] to automatically obtain data from the website. However, the authors on Scopus did not have a label to identify them as IIT Madras professors. We then used the faculty list on the institute website [6] to look up the names of the professors and manually extracted the details of the co-authors from their Scopus profile using the API. We then saved the list of coauthors for each author in a separate file. These files were used in the next step where we parsed the data and formed the networks using R [7].

In the data collection process, we have only looked at number of co-authorships between pairs of professors. We haven't looked at publications in which a set of authors have collaborated.

All the code files for the project and all the plots (some of which haven't been shown in the paper) have been put up in the GitHub Repository of the project.

3. Network Formation

Figure 2 gives a pictorial representation of the methods we used to process the data and work with the network. After extracting the data about the co-authors of each faculty member, we went about building an adjacency matrix to represent the network. This was done by using our own scripts to parse the csv (comma-separated value) files obtained from Scopus and form an ordered list of all the faculty members of IIT Madras; then, the adjacency matrix was populated using values from the same files that specify which author collaborated with which other author and for how many publications. This adjacency matrix represents an undirected weighted simple network with a node assigned to each individual faculty member and edges with weights corresponding to the number of publications which the two linked faculty members have co-authored. We then used the *igraph* [8] library in R to visualize the graph and obtain the images shown in this report. We used the packages *igraph* [8] and *networkx* [10] for obtaining the performance metrics.

4. Metrics used for Network Analysis

We used the following network metrics to analyse our network. This analysis was done across the whole network as well as department-wise.

Average Path Length(APL) : It is equal to the average of the shortest distance between every connected pair of nodes. It is a measure of the mean separation of the nodes in the network. In the presence of disconnected clusters we evaluated the the APL individually for the clusters and then used a weighted average(weighted by the size of the cluster). Mathematically, we have:

$$a = \sum_{s,t \in V} \frac{d(s,t)}{n(n-1)}$$

where V is the set of nodes in the network G. $d(s,t)$ is the shortest path from s to t and n is the number of nodes in G.

Diameter(DI) : The longest distance between all pairs of nodes. It is a measure of how far apart is the most distant pair. Here, in the presence of disconnected clusters, we considered the largest diameter amongst all the circles.

Collaborators (CL): Average number of collaborators an author has. We find an approximate average clustering coefficient for G by repeating n times (or trials) the following experiment: Choose a node at random, choose two of its neighbors at random, and check if they are connected. The approximate coefficient is the fraction of triangles found over the number of trials.

Clustering Coefficient (CC): Clustering co-efficient of the entire network.

Largest component (LC): The number of the nodes that are present in the largest connected component of the network.

5. Observations

The following observations were made by using visualisations of the networks

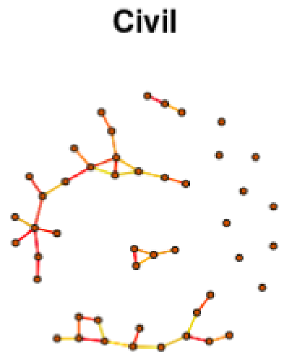

Figure 3. Co-authorship network in Dept. of Civil Engineering

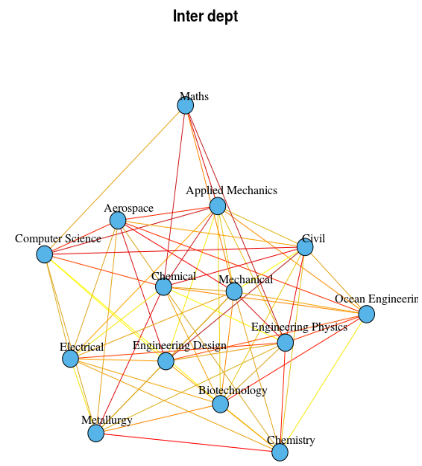

Figure 4. Co-authorship between departments

Intra-Departmental Connections. EE(Figure 5) and EP(Figure 7) department networks appear to be the most well connected networks.These networks have a single well branched chain. CE has two chains of appreciable lengths which are not connected(as seen in Figure 3). CH(Figure 8) graph indicates that in the connected chain many pairs of nodes are connected by multiple edges indicating that the links are strong (weights are represented by number of edges). Chemistry graph features small well connected groups.

A general theme among all the department graphs is that there are tightly knit groups, or often cliques of size 3. Also, a large number of isolated vertices exist in each departments network. If we analyze any clique, we note that the collaborating professors often work in the same group within the department, for example Dr. Sarit K. Das and Dr.Sundararajan,T. have a lot of collaborations and also happen to be working in the same research group in the Thermal Power Engineering Lab in the Mechanical Engineering Department. This close knit structure is in accordance with the trend in other social networks where people who have a lot in common often form cliques in such networks.

Inter-Departmental Connections. The inter department connections are uniformly present among the departments(as seen in Figure 4,also see Table 2). Each department has a degree of around 9-11 barring the Mathematics department. It has to be however noted that the number of co-authored documents in inter-disciplinary research is significantly lower as compared to the total number of co-authored documents published within a department.Also, we can see that there are pairs of departments which have a relatively high number of collaborations, for example ME and CH, and MM and PH, and EE and CS. This is as expected because these pairs of departments have a significant overlap in the fields which they study. Hence, this is not complete evidence that there is interdisciplinary research being conducted.

Network Attributes. Network attributes were calculated department-wise using igraph and networkx packages. The attribute values have been recorded in Table 1.

TABLE 1. DEPARTMENT-WISE NETWORK PARAMETERS

	APL	**DI**	**CC**	**LC**
CH	3.142	7	0.120	21
CE	3.616	9	0.079	18
ED	1.285	2	0.0	4
EE	3.979	10	0.142	45
AE	2.531	5	0.0	10
AM	1.25	2	0.0	3
BT	2.890	6	0.170	13
CS	2.361	4	0.0	12
ME	3.662	8	0.076	34
MM	2.388	5	0.096	12
OE	2.839	6	0.269	17
CY	2.425	5	0.057	9
PH	3.197	8	0.058	34
MA	1.461	2	0.0	5
All	**5.505**	**14**	**0.011**	**193**

Department Leaders. The following professors were found to have the most number of collaborations:

- Sayan Gupta(AM)
- Vignesh Muthuvijayan(BT)
- Manikandan Mathur(AE)
- Raghunathan Rengasamy(CH)
- Ashok Kumar Mishra(CY)
- Ramaswamy Sivanandan(CE)
- S.Ganesh Sundara Raman(MM)
- Veezhinathan Kamakoti(CS)
- Venkatesh Balasubramanian(ED)
- Shanthi Pavan(EE)
- Ashwin Joy(EP)
- Raghavan Rama(MA)
- Ravikiran Sangras(ME)
- Govindarajan SureshKumar(OE)

Similarity to a Social Network:. In trying to analyse the Co-authorship network as a Social Network, we can draw an analogy to a Network of Friends. Authors could be analogous to people,collaborations to friends(and hence the strength of the friendship is signified by the number of collaborations),departments to cities, cliques to a friend-circle(each person knows the other) and a chain of collaborations to a chain of friends. Using this analogy we can see the remarkable resemblance of our network to a Social Network.

We see that most collaborations happen within the department(friendships within the city). Even in those inter-departmental collaborations that do exist, the number of collaborations seems to be low(just like friendships across city). There are many more such comparisons that can be done.

6. Conclusion

In this paper we analysed the co-authorship network in IIT Madras. We provided visualisations of the entire network as well as the intra- and inter-department networks. Also, we also measured a few vital network parameters in the department networks. As expected, the intra-departmental collaborations were more than that of inter-departmental collaborations. Again, as expected but less obvious, even within departments we saw formation of clusters of authors which in many cases corresponded to fields of study within the department.

This study has given us considerable insights into the nature of collaborations among the professors in the institute. The network throws light on the nature of intra- and inter-departmental research(or the lack of?) happening in the institute. This could help the institute plan its future research endeavours in the direction of inter-disciplinary research.

7. Future Work

We see two ways of expanding on the project in the future:

- We could expand the analysis to multiple institutes and even look at inter-institute collaborations. In this paper we have only considered the professors' collaboration in our analysis. We could also extend this to the students and other researchers in the institute.
- In this paper we have constructed the network using only the number of co-authorships as the data. In the future, we could also look at papers in which a set of professors have collaborated. We could then construct hyper-graphs which would contain much more information.

Acknowledgments

The authors would like to thank Dr. Venkatesh Ramaiyan for having given them an opportunity to work on this project, and also for providing references and guidance during the project.

References

[1] P. W. Holland and S. Leinhardt (1971). "Transitivity in structural models of small groups". Comparative Group Studies. 2: 107124.

[2] D. J. Watts and Steven Strogatz (June 1998). "Collective dynamics of 'small-world' networks". Nature. 393 (6684): 440442.

[3] Tasleem Arif, Rashid Ali, M.Asger (August 2012). "Scientific Co-authorship Social Networks: A Case Study of Computer Science Scenario in India".International Journal of Computer Applications (0975 8887) Volume 52 No.12

[4] Scopus — The largest database of peer-reviewed literature — Elsevier

[5] John R. Kitchin: Python API for Accessing Scopus databases with their REST API

[6] Indian Institute of Technology Madras: Official Website

[7] The Comprehensive R Archive Network

[8] python-igraph

[9] NetworkX-Python

[10] GitHub Repository of the Project

Appendix-A

Inter-Department Collabration Matrix:. In Table 2, we have recorded the number of inter-department collaborations among different departments. We have set the intra-department count to 0 as that has been studied extensively in the rest of the paper.

TABLE 2. NUMBER OF INTERDEPARTMENT COLLABORATIONS

	AM	BT	AE	CH	CY	CE	CSE	ED	EE	PH	MA	ME	MM	OE
AM	0	4	6	2	0	10	1	9	1	3	2	14	0	3
BT	4	0	0	6	11	8	6	0	4	6	0	4	17	9
AE	6	0	0	5	0	1	1	3	11	0	0	26	7	6
CH	2	6	5	0	5	6	1	0	4	9	7	33	13	1
CY	0	11	0	5	0	5	0	1	6	23	0	23	4	16
CE	10	8	1	6	5	0	4	3	0	9	0	9	0	8
CSE	1	6	1	1	0	4	0	2	24	0	1	0	1	0
ED	9	0	3	0	1	3	2	0	17	2	0	8	2	3
EE	1	4	11	4	6	0	24	17	0	18	0	12	12	0
PH	3	6	0	9	23	9	0	2	18	0	3	28	51	2
MA	2	0	0	7	0	0	1	0	0	3	0	2	0	0
ME	14	4	26	33	23	9	0	8	12	28	2	0	23	2
MM	0	17	7	13	4	0	1	2	12	51	0	23	0	0
OE	3	9	6	1	16	8	0	3	0	2	0	2	0	0

Intra-Department Plots:. Following are the plots of a selected number of departments. For a collection of all the plots please refer to the GitHub Repository of the project.

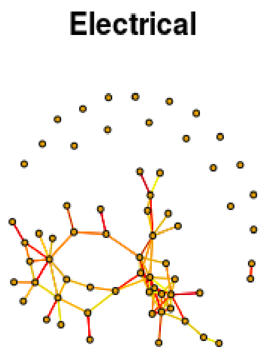

Figure 5. Co-authorship network in Dept. of Electrical Engineering

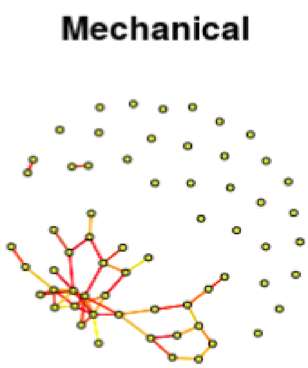

Figure 6. Co-authorship network in Dept. of Mechanical Engineering

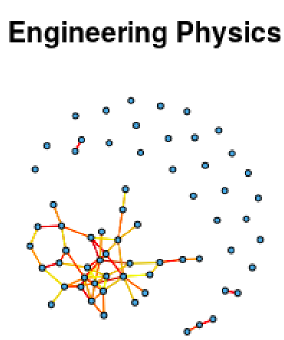

Figure 7. Co-authorship network in Dept. of Physics

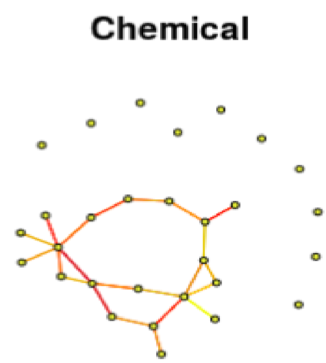

Figure 8. Co-authorship network in Dept. of Chemical Engineering