

COLLABORATION NETWORK IN DEPT. OF COMPUTER
SCIENCE & ENGINEERING
IIT KHARAGPUR



*Term Project on
Complex Network Theory*

BY

**Arnab Sinha
Mithun Dhali
Lalit Narayan Paswan**

Department of Computer Science and Engineering,
Indian Institute of Technology,
Kharagpur

Done under the guidance of

Prof. Niloy Ganguly
Department of Computer Science & Engineering,
Indian Institute of Technology,
Kharagpur

➤ Introduction:

Collaboration networks of research communities reveal a lot of insight into the community structure and the *scientific productivity* of individual as well as group. With an aim to study the various parameters (e.g. *degree distribution*, *clustering index* etc.) of the collaboration situation in the department of Computer Science and Engineering, Indian Institute of Technology, Kharagpur we present our findings.

➤ Various Definitions:

We have observed the degree distribution and the clustering index of each person who have so far contributed in the scientific productivity of the department through their published works in various conferences and journals.

By degree distribution, we want to measure the “collaborative” attitude of research community in CSE, IIT Kharagpur. Thereby we want to validate the *law of scientific productivity*. We plot the degrees in increasing order in the X-axis and the number of corresponding persons who co-authored with that many people.

Through computation of clustering index of each member of the network, we want to guess the different groups who have co-authored in several papers. Formally, we define clustering index (C_w) as follows.

$$C_w = \frac{\sum_{u,v \in N(w)} e(u,v)}{n(n+1)/2}$$

where $e(u,v) = 0$, if no edge between u and v
1, if edge exists between u and v,
 n is the number of neighbors of node w,
 $N(w)$ is the set of neighboring nodes of w.

➤ Observations:

We have collected most of the data from faculty homepage (<http://www.facweb.iitkgp.ernet.in>), and DBLP bibliography (<http://www.informatik.uni-trier.de/~ley/db/>). We have about 358 members in all. We have accounted for all the faculties, students who contributed over a decade. We show our observations in the following charts. Chart 1 shows the log-log plot.

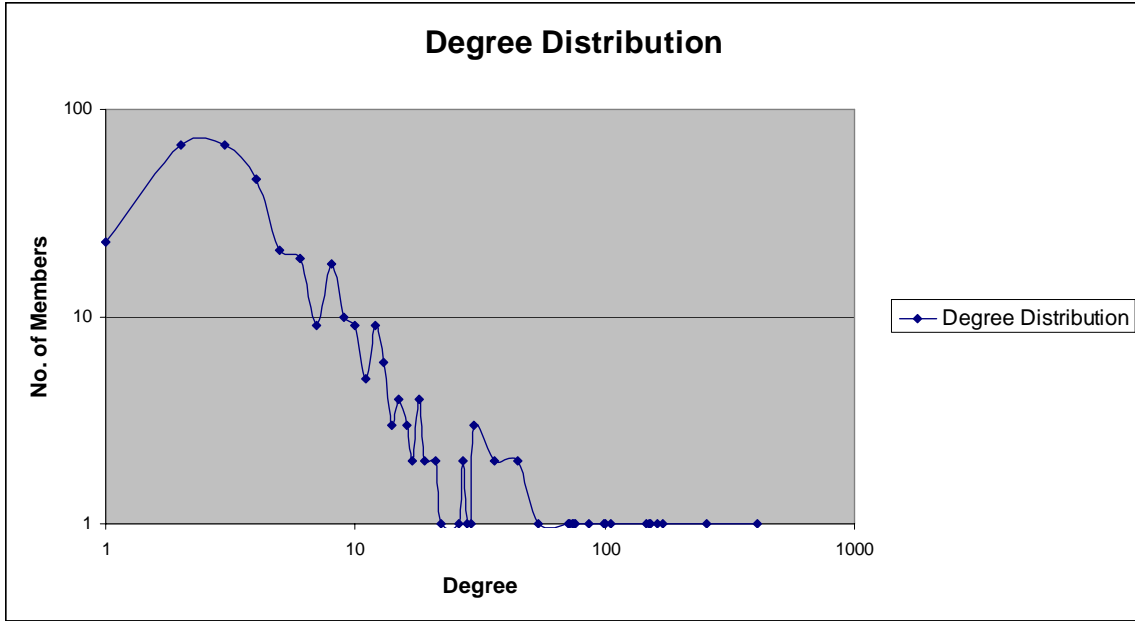


Chart 1: The Degree Distribution of the Collaboration Network.

We can find certain peaks in the plot but it has a heavy trail in the right. These peaks stand for the cliques who probably share the same research interest and have co-authored in many papers, and hence they have similar degree distribution. This actually validates the law of scientific productivity.

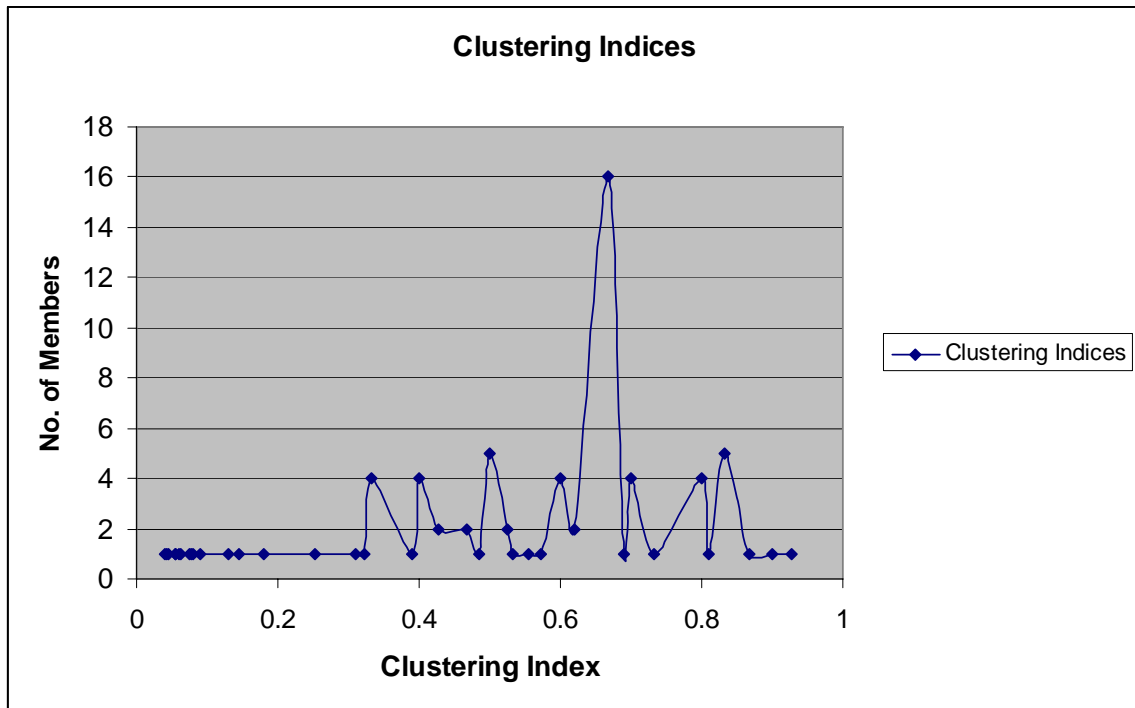


Chart 2: Distribution of Clustering Indices

As we can see, there are some peaks in the chart 2. Since the peaks are caused by no. of members having same clustering indices, this fact directly implies the existence of various groups in the

department that existed at some point of time and shared their research findings among themselves.

Now we present the typical findings regarding the current faculty members of the department. Due to some practical problems, we are yet to receive some more data, which might be throwing significant light on the collaboration network analysis.

Faculty Name	Degree	Clustering Index
A. K. Majumdar	149	0.055904962
A. Basu	163	0.0589372
A. Gupta	86	0.055172414
A. Pal	100	0.08
Indranil Sengupta	105	0.045404207
Jayanta Mukherjee	150	0.07439613
Pallab Dasgupta	255	0.06169772
P. P. Chakrabarti	407	0.038961038
Rajeev Kumar	152	0.04244898
Rajib Mall	76	0.12987013
S. C. De Sarkar	72	0.52380955
S. Ghosh	170	0.090756305
S. P. Pal	99	0.07862903

Table 1: The Faculty Findings

For further illustration we plot the above data in the Chart 3. Here we plotted the normalized degree and the clustering indices for a comparative study.

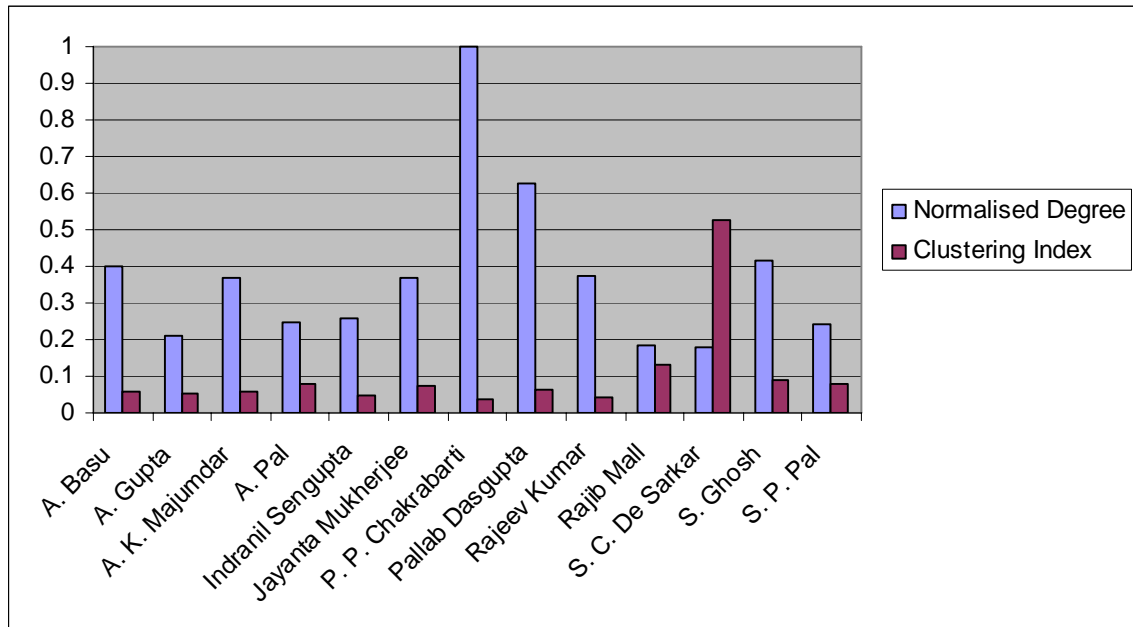


Chart 3: Comparative Study on Degree vs. Clustering indices

It is interesting to note that almost all the faculty members have their clustering indices less than 0.1. Prof. Chakrabarti (degree: 407) and Prof. Rajeev Kumar (degree: 152), as for example have almost the least clustering indices (0.04 each) while Prof. Mall (degree: 76) has a moderate clustering index (approx 0.12). This may be an indication that Prof. Chakrabarti and Prof. Kumar

have interests in various fields and thus collaborated with different set of people while their interests match in certain fields. While on the other hand, the case of Prof. Mall suggests that he restricted his research interest to some specific area and interacted to limited set of people who are known among each other well, compared to that of Prof. Chakrabarti and Prof. Rajeev Kumar.

➤ **Criticism:**

Although the data which we obtained revealed a lot of information on the research communities in department of Computer Science, IIT Kharagpur, the parameters do not reflect the following.

- The productivity (number of papers published from a clique) and its quality (impact of journals, conferences etc.) are not taken into account while estimating the clustering indices. There are about 250 members with clustering indices as 1 but having very low degrees. Some authors after initial productivity have gone into oblivion also share clustering indices as 1, with the co-authors of the papers published.
- Clustering indices are undefined for leaf nodes which are connected to only one node in the network.
- The mentioned parameters fail to identify the leaders in the network.

➤ **Conclusion:**

This kind of analysis of the collaboration network in a research community can be a tool for framing various policies. The selection of groups and leaders for new projects, decisions regarding promotions, the overall and individual scientific productivity etc can be estimated.

➤ **Source of Data:**

1. <http://www.facweb.iitkgp.ernet.in>
2. <http://www.informatik.uni-trier.de/~ley/db/>
3. Personally meeting with various research scholars in different labs.