Scientific Co-authorship Social Networks: A Case Study of Computer Science Scenario in India

Tasleem Arif
Department of IT,
BGSB University Rajouri, J&K,
India.

Rashid Ali
College of Computers and IT,
Taif University,
Taif, Saudi Arabia

M. Asger
School of Mathematical Sc. & Engg.
BGSB University Rajouri, J&K,
India.

ABSTRACT
Co-authorship is one of the most tangible and well documented forms of research collaboration. Data mining techniques and social network analysis can be used to extract and study these collaborations. Social network analysis provides an insight into the connections between groups of individuals. It is these connections that channel flow of information and the sharing of knowledge. In order to understand flow of information and interpret collaboration, co-authorship can be used as a measure to study intra and inter organization collaborations. In this paper, we analyze the collaboration scenario in Computer Science in India, and access how researchers in few of the best Indian Institutes of Technology (IITs) collaborate and relate to each other. We construct and visualize scientific co-authorship social network graphs of these institutions. We also compare and contrast network metrics for these institutes and experimentally deduce that these networks like other social networks exhibit "*small world*" properties.

General Terms
Data Mining, Web Mining, Social Network Analysis

Keywords
Co-authorship Networks, Visualization, IIT

1. INTRODUCTION
A social network is a structured representation of the social actors (nodes) and their interconnections (ties). Such a network can be represented as a set of points (or vertices) denoting people, joined in pairs by lines (or edges) denoting acquaintance. Social networks form social groups that share common interests. These groups are steadily emerging on the Web and the demand for forming an on demand social network is immense. Community members profit from being linked to other people sharing common interests, though having geographically dispersed affiliations. One could, in principle, construct the social network for a company or firm, for a school or university, or for any other community up to and including the entire world.

Extraction and visualization of social relations can benefit many applications in areas like crime and terrorism prevention, organizational network analysis, customer interactions, connections and communities. Understanding the graph structure of these networks can benefit many applications in various diversified fields. Adjacent users in a social network tend to trust each other and mostly have common interests. Users normally find the content of their interest in their neighboring regions. It would be useful to have efficient algorithms to infer the actual degree of shared interest between two users and trust or reliability enjoyed by a user among other users in the network.

Sharing of knowledge and interaction between researchers is well known to be the essence of research practice and collaboration. Collaboration is defined as "*working jointly with others or together especially in an intellectual endeavor*" [1]. Researchers interact not only to communicate research activities but also to collaborate with each other to coproduce research and co-author research results. Since collaboration has the potential to promote research activity, productivity and impact, it should be encouraged, supported and monitored. Although it has been argued that co-authorship is no more than a partial indicator of collaboration, studies indicate that it is the highest measure of collaboration [2]. In several studies, for instance in [3], it has been shown that there is a positive correlation between collaboration and co-authorship. In fact, co-authorship is one of the most tangible and documented forms of research collaboration [4]. These collaborations or connections form a social network, and in order to understand their effect, they need to be viewed from a network perspective.

Social network analysis (SNA) focuses on the relationships among social entities, and on the patterns and implications of these relationships, and allows us to examine those patterns in a structural manner [5]. SNA can be used to discover underlying social structure such as: central nodes that act as hubs, leaders or gatekeepers; highly connected groups; and patterns of interactions between groups [5]. SNA has been used to study social interaction in a wide range of domains. Examples include: collaboration networks [6], directors of companies [7], inter-organizational relations [8], and many others. Social Network Analysis examines relationships between social entities i.e. people, groups, teams, tasks, beliefs, knowledge, etc. These entities are modeled with nodes and their relationships are modeled with links between the nodes. Not all nodes in the network are connected and some nodes may have multiple connections. This mathematical model is applicable in many content areas such as communications, information flow, and group and organizational affiliations [5]. SNA looks at groups of people and their interactions. This type of analysis provides a methodology that does a very good job at explaining much of the complex behavior of these social groups. SNA thus relies heavily on graph theory to model network structure.

Although there are other forms of academic collaborations but this paper defines collaboration as jointly co-authoring a paper and shows the use and evaluation of our approach on the identification of scientific co-authorship relationship. We use publications data to extract social networks of researchers. From the publication data, it is possible to know various attributes of a researcher like his research interests, collaborations, and even conferences attended recently. The publication data can be retrieved from various sources like journals, electronic databases, conference websites and proceedings, homepages of researchers and organizations, etc. Nowadays, it is common for research institutions and researchers to maintain a record of their publications and provide the same on their respective websites.

In this paper, we discuss the extraction of a collaboration network to study co-authorship collaborations in Computer

Science area of four of the IITs. The collaboration network is essentially a graph (*G*) where the vertices (*V*) represent authors and the link/edge (***e***) between them represents the fact that they are related by the relation of co-authorship. Each edge has certain weight that reflects the number of papers written by a pair of vertices (i.e. authors). In this paper, we use node, vertex and author interchangeably.

The paper is organized as follows. In section 2, we discuss background and related work in the area. We discuss our procedure for data collection in section 3. In section 4, we discuss various measures that can be used for analysis of the social networks. In section 5, we present the architecture of our social network extraction system. We present and discuss our experimental results in section 6. Finally, we conclude and give some future directions in section 7.

2. BACKGROUND AND RELATED WORK

Universities and other institutions of higher learning have been known for providing solutions to various problems confronting the society. Research has been answering to many such problems. Modern day research is faced with both extraordinary opportunities and challenges. A fast paced modern society turns to academics as public servants for immediate answers to an array of practical problems created by its own increasing needs and desires. Society is willing to invest in research as the basis of a knowledge economy as long as research proves to be responsive to its needs, productive and effective. Most of the questions science is required to answer are too complex to be addressed in the traditional disciplinary framework of academic research. Yet with the explosion of knowledge, research has become only more fragmented than ever. The lack of communication and coordination even within the faculties of a single university or even a single department leads to opportunities lost every single day. Lack of collaboration has aggravated the problem and even research groups within universities become specialized: long gone are the days where every subdiscipline within a scientific domain was equally represented at a university or other research institution.

Social networks have got a lot of focus from the research community long before the advent of the Web [5]. Social sciences made great strides in measuring and analyzing social networks between 1950 and 1980, at the same time when Vannevar Bush's proposed hypertext medium 'Memex' was gaining acceptance [9]. There are numerous examples of social networks formed by social interactions like co-authoring, advising, supervising, and serving on committees between academics; directing, acting, and producing between movie personnel; composing, and singing between musicians; trading and diplomatic relations between countries; sharing interests, making phone calls, and transmitting infections between people; hyperlinking between Web pages; and citations between papers.

The last decade has seen a rapid growth of research interests in Online Social Networks. Social network extraction is an emerging field of research with majority of research work concentrated towards late 2000s and the focus has been to construct efficient procedures and algorithms for the identification of community structure in a generic network. Construction of the researcher network by automating information extraction from Web can benefit many Web mining and social network applications [10]. For example, in this case, if all the profiles of researchers are correctly extracted, we will have a large collection of well-structured data about real-world researchers. The profiles extracted can help in expert finding for research guidance for new scholars, potential speakers and contributors for conferences, journals, workshops etc. The academic network so extracted can be used in many services, such as finding an appropriate person to introduce or negotiate someone, who one should talk in order to expand his/her network efficiently [11]. The extracted academic network may also be used for trend detection/prediction. Trend detection can help a researcher to analyze the thrust area of research in a particular field, what other researchers are doing in that or related field. Trend prediction can help research community to have an idea of the potential research topics/areas in a particular field. The size, reach, growth and diversity of these networks are their characteristic features challenging the research community.

Scientific social networks can be obtained by considering different scientific relations like project participation, co-authorship, thesis supervision, conference participation, technical production, etc. Social network of researchers can be constructed by using any or combination of these relations. However, the collaboration is normally established based on similar research interests. Of the various relations mentioned above, co-authorship relation is the most important measure [2] of collaboration among individuals and organizations.

There are several studies which have been conducted for extraction of social network from various information sources like WWW, e-mail, instant messenger logs, search engines, etc. Referral Web [12] was the first attempt of this kind to develop an automated interactive tool for social network extraction from a specified domain and finding shortest referral chains to experts. It uses a search engine (Altavista) to extract social networks through co-occurrence of names in close proximity in any document e.g. personal homepages, lists of co-authors in technical papers, citations, and organizational charts publicly available on the WWW taken as evidence of a direct relationship. The network obtained is an egocentric network, in that it is focused on a specific person. Referral Web [12] has influenced many studies for automatic extraction of social networks. Tombe et al. [11] proposed a system for social network extraction of conference participants from the Web. The idea behind this study is that: at academic conferences, a participant registers a brief profile with fields like Name, E-mail, Affiliation, etc. well before the conference which means that there is enough time to gather information about the participants from the Web. P. Mika in the year 2005 developed Flink [13], a system for extraction, aggregation, and visualization of online Semantic Web community. The Web mining module of Flink obtains as in [12] hit count from a search engine (Google) for both the persons X and Y individually as well as hit count for co-occurrence of these two names with the target being the Semantic Web community.

Constructing a co-authorship relation based social network is straight forward, as various authors of a publication are explicitly stated in the publication itself. In SNA, co-authorship networks are also called as "*affiliation networks"* where vertices represent authors and edges represent the relationship (i.e. co-authored publication) between these authors. If two vertices are connected to each, it implies that the authors represented by those two vertices have co-authored a paper. The strength of relationship between two authors is directly proportional to the number of publications authored together.

Collaboration networks in the scientific communities are a well-studied subject for its inherent complexity and motivation to predict or analyze certain features among the persons involved. Various studies [14, 15], have investigated into various network parameters like small-world, betweenness centrality, vertex centrality etc. to interpret the obtained data. The analysis of these parameters provides a good insight of the health of the research community and the institution. Good health indicates that the institution is alive, full of activity, publishing more papers, and attracting more projects, collaborations and grants. This can help one to find potential researchers for collaboration, project funding, guidance etc. and could prove beneficial for institutions, students, and funding agencies as well.

The co-authorship network can be considered as a true a proxy to the social collaboration network of the researchers and has attracted research attention in recent years. The idea of co-authorship networks started with the Erdos number project (www.oakland.edu/enp/). The greater the distance between Paul Erdos and another researcher (*r*), the greater will be his Erdos number and vice versa. Many studies [14, 15, 16, and 17] on co-authorship network analysis across several domains were made by Newman. These studies answered a broad variety of questions about collaboration patterns by analyzing co-authorship networks across several domains: Biology, Computer Science, Mathematics, and Physics. The parameters analyzed in these studies were: the number of papers authors write, how many people they write them with, the typical distance between researchers through the network, etc. Few other studies also have analyzed co-authorship network. For example, social science collaboration networks were investigated in [18], digital library community in [19, 20], and database community in [21].

3. DATA COLLECTION

There are various online sources for collection of public data. These include publication databases like DBLP, CiteSeer, Google Scholar, etc., journal homepages, conference proceedings (websites), homepages of individual researchers as well as organizational/institutional websites. Since this work focuses on collaboration between faculty members of the institutions under consideration, institutional websites are the best and authentic source of the publication data of their faculty members. The data which we have used in this study has been obtained from the websites of the four IITs under consideration i.e. namely IIT Delhi, IIT Kanpur, IIT Kharagpur, and IIT Madras.

The data analyzed in this work pertains to 2005-2011 period. We obtained data about faculty members and associated publications of Computer Science departments of these four IITs. The data was highly unstructured as is the case with any Web data. Data on web pages can be found in different formats. HTML is designed for unstructured data which contains information in several formats, e.g., text, image, video and audio. It is known that web pages in HTML format are 'dirty' because their contents are ill-formed and 'broken' [22]. Moreover, different IITs have maintained the required data (faculty list and publications list) in different formats.

We considered only those faculty members for the purpose of social network extraction who were currently occupying a teaching position in the department. Altogether, we analyzed publications of 107 researchers from the Computer Science area of four IITs resulting in 1017 co-authored publications, including journal papers, conference papers, invited papers and technical reports, and 2375 co-authorship relationships.

To analyze the data under consideration and extract social networks, it was necessary to extract the co-author relationship for the researcher under consideration from publications data. The problem with publication data is that the author names of a particular publication are written in different formats in different publications. Moreover, the formats for name list and publications were different in different IITs. In order to map names in the faculty list to the name in publication data, we developed an algorithm to extract similar/same name from these publication data. We have used this algorithm for the purpose of name resolution and extraction of relationships. The algorithm (not discussed in this paper) worked with great precision and the relationships extracted and used in this work were obtained using it.

4. NETWORK ANALYSIS METRICS

Social Network Analysis is a branch of sociology that is formalized to a great extent based on the mathematical notions of graph theory. This formal model captures the key observation of Social Network Analysis, namely that to a great extent social structure alone determines the opportunities and limitations of social units and ultimately effects the development of the network as well as substantial outputs of the community.

The field of Social Network Analysis as it is today is the result of the convergence of several streams of applied research in fields like sociology, social psychology and anthropology. Many of the concepts and theories of network analysis have been developed independently by various researchers often through empirical studies of various social settings. For example, many social psychologists of the 1940s found a formal description of social groups useful in depicting communication channels in the group when trying to explain processes of group communication. Already in the mid-1950s anthropologists have found network representations useful in generalizing actual field observations, for example when comparing the level of reciprocity in marriage and other social exchanges across different cultures.

An important aspect of the structure of social networks came from a remarkable experiment conducted by the American psychologist Stanley Milgram [23]. Milgram in 1967, went out to test the common observation that no matter where we live, the world around us seems to be small: we routinely encounter persons not known to us who turn out to be the friends of our friends. Milgram's experiments not only wanted to test whether we are in fact all connected but was also interested in what is the average distance between any two individuals in the social network of the American society. Milgram calculated the average of the length of the chains and concluded that the experiment showed that on average Americans are no more than six steps apart from each other.

Social network metrics such as degree, betweenness, closeness and network centrality are often the subject of academic research. Understanding social networks and their metrics is important as these networks form the underlying structure, which allows for rapid information distribution [24].

Several network analysis measures as proposed in [25] can be used to indentify influential nodes and discover community structures of the extracted social networks. We are interested in capturing the internal connectivity as well as attributes of key nodes in the network. In order to identify the leaders in the network, the quantity of interest in many social network studies is the "*betweenness centrality*" of an actor '*i*'. *Centrality* is a measure of the information about the relative importance of nodes and edges in a graph. Several centrality measures like *betweenness centrality*, c*loseness centrality*, and *degree centrality* have been proposed in [25] to identify the most important actors (leaders) in a social network. In addition to other aspects of analysis, we are also interested in answering the following four questions:

(i) Who are the hub/leaders?
(ii) Who has more connections?
(iii) How strong are the collaboration ties?
(iv) How collaborative the authors are?

We may use four measures namely (i) *Betweenness centrality*, (ii) *Degree centrality*, (iii) *Clustering coefficient*, and (iv) *average degree* to answer the above four questions efficiently.

Betweenness centrality measures the fraction of all shortest paths that pass through a given node. Nodes with high *betweenness centrality* play a crucial role in the information flow and cohesiveness of the network and are indispensable to the network due to the information flow they assist in. Nodes with the high *betweenness* act as gate keeper [26].

Degree centrality of node in the network is the number of links incident on it and is used to identify nodes that have highest number of connections in the network. A more sophisticated version of degree centrality is *eigenvector centrality*. It not only depends on the number of incident links but also the quality of those links [26]. We use *eigenvector centrality* in our experiments.

Clustering coefficient signifies how well a node's neighbourhood is connected. The more connected the neighbours are with one another, the higher the clustering coefficient. This is because the neighbourhood graph is heading towards becoming a clique i.e. a complete graph where every node is connected to one another [27]. The clustering co-efficient of a network as given in [28] is the average of the clustering co-efficient of all the nodes in the network. It indicates the degree to which nodes in a network tend to cluster together and it is therefore considered to be a good measure if a network demonstrates "*small world*" behaviour [28]. Stanley Milgram's [23] theory of the "*6 Degree of Separation*" utilises the *average path length* metric.

The *average degree* of all the nodes in the network is a measure of how collaborative the authors are.

5. SOCIAL NETWORK EXTRACTION SYSTEM ARCHITECTURE

Figure 1 shows the architecture of the system that we used for extraction and visualization of co-authorship social network from the obtained publication data. The publication data obtained from the websites of four IITs under consideration is cleaned before further processing. We parse the input to store it in the database in a suitable format. The extracted data is stored in a database. The relationship extraction module takes publication and faculty information as input from the database. The relationship extraction module extracts affiliation lists from the publication data which serves as input to the social network visualization engine.

We use *NodeXL* [29], an open source graph analysis and visualization tool, which works as an add-on to Microsoft Excel, as social network generation and visualization engine for visualization and analysis of the extracted co-authorship relations. The social network visualization engine generates the social (academic) graphs and other network statistics based on the affiliation (co-authorship) list provided.

6. RESULTS AND DISCUSSION

Two types of co-authorship social (academic) network graphs viz. affiliation networks and internal collaboration networks were obtained using the graph generation and visualization engine. The affiliation networks shows the general co-authorship relationships (internal as well as external) whereas internal affiliation networks shows the co-authorship collaboration among the faculty members within the department (of a particular IIT) itself. Figure 2(a), 3(a), 4(a) and 5(a) show the extracted affiliation networks and Figure 2(b), 3(b), 4(b) and 5(b) show the extracted internal collaboration networks. In these graphs, we have labelled edges blue, green, and red for highest collaborative strength, moderate collaborative strength, and weak collaborative strength, respectively. In addition to the colours which we have used to distinguish the strength of relationship, the width of the edges is directly proportional to the strength of relationship. The more the width of a particular edge the stronger the relationship i.e. more papers authored together. We extracted and analyzed various graphs and tried to come up with meaningful interpretations of the co-authorship networks from our study, which are given as follows.

6.1. Distinguished Nodes (Hubs and Leaders)

Statistics obtained show that in IIT Kanpur (Figure 3-a), Kharagpur (Figure: 4-a) and Madras (Figure 5-a), Amitabha Mukherjee, P. P. Chakarbarti, and V. Kamakoti respectively have highest betweenness as well as highest eigenvector centrality; it means that faculty members occupying the central position in terms of information flow in the network were also those who are highly connected in the respective networks. This is not the case in IIT Delhi (Figure 2-a) as the nodes having highest betweenness and eigenvector centrality were Amitabha

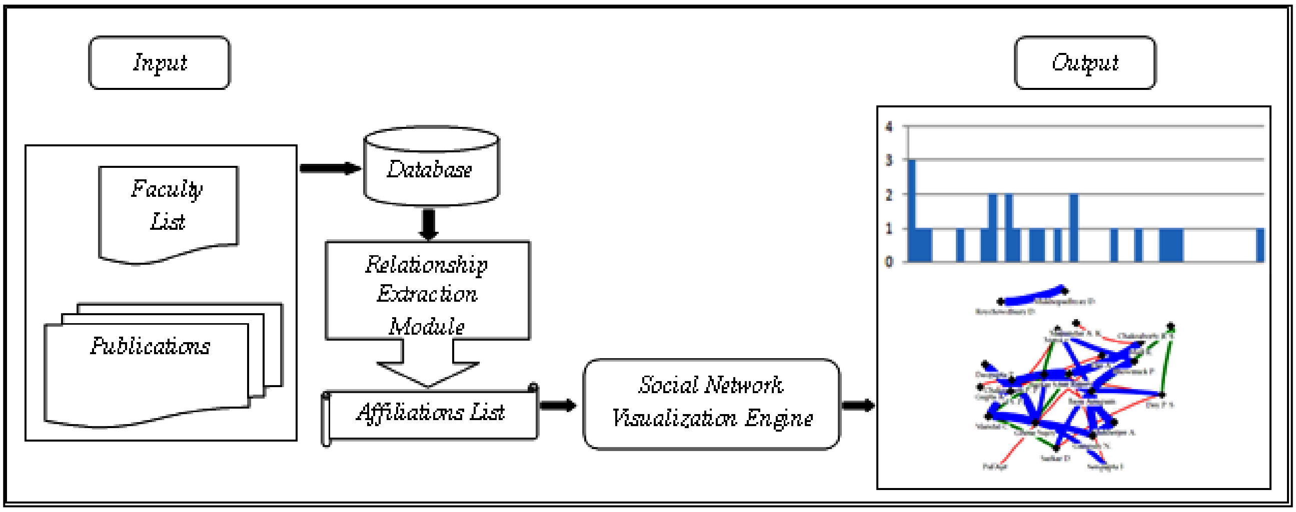

Figure 1: Social Network Extraction System Architecture

Bagachi and Subhashis Banerjee respectively; it means that although Amitabha Bagachi acts as gatekeeper in collaboration network, Subhashis Banerjee is the highest connected node in the network. This can be attributed to that fact that Subhashis Banerjee is connected to other highly ranked nodes in the network like Prem Kalra (as evident from Figure 2-a), whereas Amitabha Bagachi despite having highest betweenness have connections with low ranked nodes in the network.

6.2. Strength of Collaboration Ties

As per values of various network attributes that we have obtained from the graph visualization engine (Table-1), IIT Kanpur has the lowest clustering co-efficient. This highlights the lack of connectivity between the vertices (authors) in the affiliation graph of IIT Kanpur (Figure 3-a). Only few of the nodes are tightly clustered in *cliques* i.e. complete graph. This gets validated from the values of *diameter* and *average path length* for IIT Kanpur, as given in table-1. IIT Kharagpur has highest clustering co-efficient of all the four IITs and also the largest connected component of all the four affiliation graphs. This highlights great connectivity between the vertices in the affiliation graph of IIT Kharagpur and majority of the nodes in the affiliation graph are clustered in complete graphs i.e. cliques. This means that the flow of information is hard in IIT Kanpur, whereas it is easy in case of IIT Kharagpur.

All the four affiliation graphs (Figure 2-a, 3-a, 4-a and 5-a) have *average path length* less than 6 which means that these networks are like other social network graphs and are *small world*.

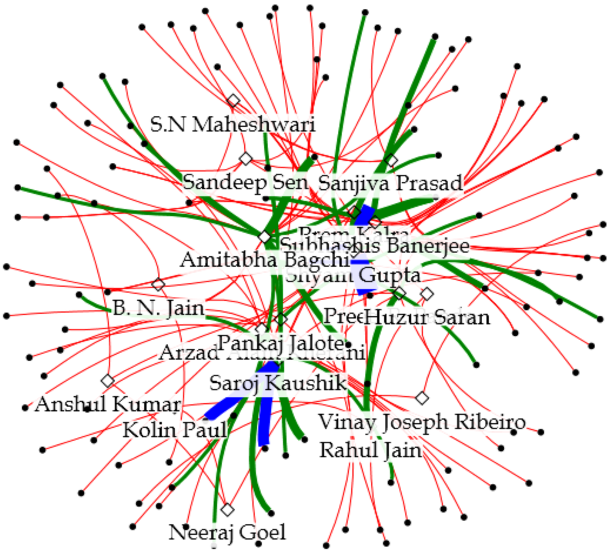

Figure 2(a): IIT Delhi Co-authorship Graph

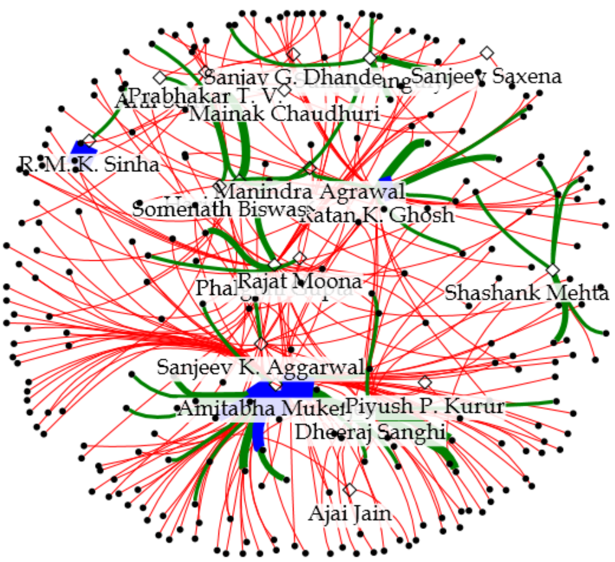

Figure 3(a): IIT Kanpur Co-authorship Graph

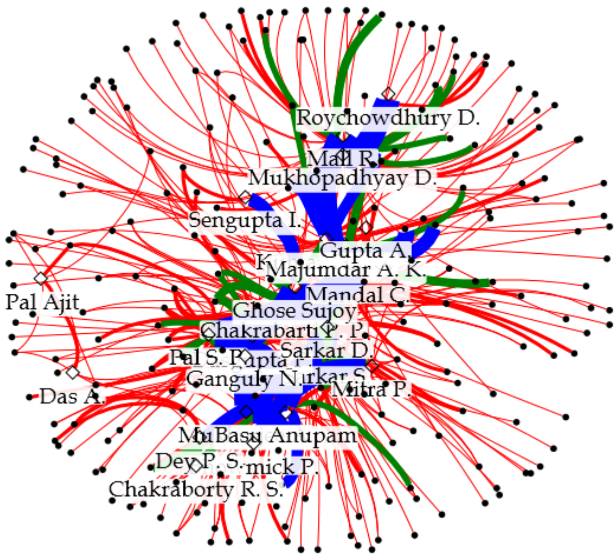

Figure 4(a): IIT Kharagpur Co-authorship Graph

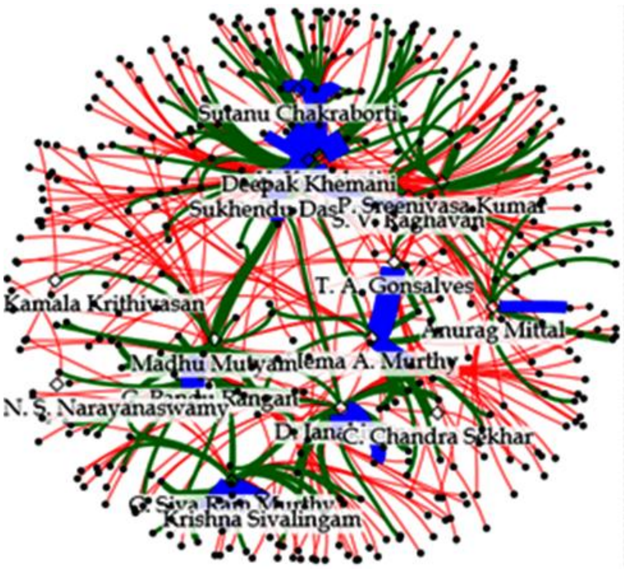

Figure 5(a): IIT Madras Co-authorship Graph

6.3. Network Attributes

Table-1shows various attributes of the co-authorship graphs that we have obtained by using *NodeXL* template for social network graph extraction and visualization. Various attributes of interest for the study are:

- ***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. Unconnected pairs are excluded.
- **Diameter (DI):** The longest distance between all pairs of nodes. It is a measure of "how far apart is the most distant pair".
- **Collaborators (CL):** The author's average collaborator.
- **Clustering Coefficient (CC):** Clustering co-efficient of the entire network.
- **Largest component (LC):** The percentage of the nodes that connect to the largest component (largest connected component).

Table 1: Values of various network attributes of the co-authorship graphs.

IIT	P*	A#	APL	DI	CL	CC	LC
Delhi	138	131	1.76	4	2.03	0.092	16.79
Kanpur	225	271	4.37	10	2.02	0.017	85.24
Kharagpur	247	252	3.48	6	3.22	0.329	93.25
Madras	407	419	3.56	7	2.14	0.085	74.70

***Papers**
#Authors

The statistics presented in Table-1 shows various network attributes for co-authorship social network graphs that we have obtained for both internal as well as external author collaborations. From this statistics, it can be inferred that the faculty of the Computer Science department of IIT Madras has published the highest number of papers followed by IIT Kharagpur, Kanpur and Delhi. The strength of collaboration ties of the researchers of IIT Kharagpur is the highest whereas it is least in case of IIT Delhi. Mean separation (diameter) is largest in IIT Kanpur and smallest in case of IIT Delhi. This can be verified from the values of average path length.

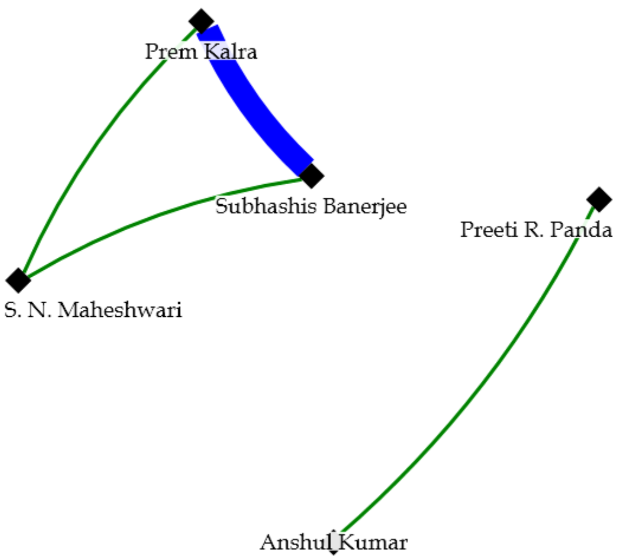

Figure 2(b): IIT Delhi-Internal Collaboration Graph

6.4. Inter-Departmental (Internal) Collaborations

IIT Kharagpur (Figure 4-b) lead the four IITs in internal collaboration with 85.16% of the total faculty members actually collaborating with each other, directly or indirectly, whereas in IIT Madras (Figure 5-b), Kanpur (Figure 3-b), and Delhi (Figure 2-b), it was 56.52%, 34.62%, and 16.13% respectively. The size of the largest connected component in IIT Madras,

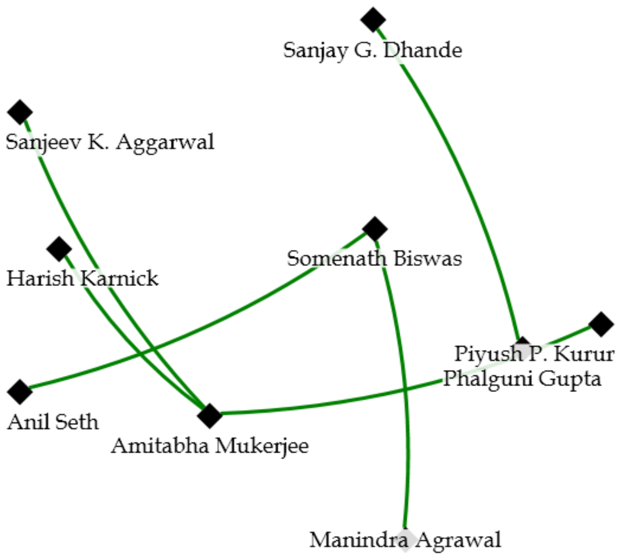

Figure 3(b): IIT Kanpur-Internal Collaboration Graph

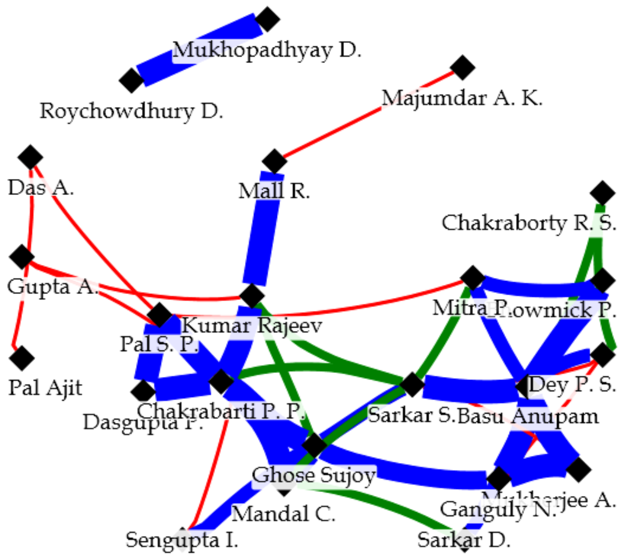

Figure 4(b): IIT Kharagpur-Internal Collaboration Graph

Kharagpur, Delhi, and Kanpur was 100%, 91.31%, 60.00%, and 44.44% respectively. This means that all the faculty members in the local affiliation network of IIT Madras are connected to each other directly or indirectly, whereas in case of IIT Kharagpur, Delhi, and Kanpur it is 91.31%, 60.00%, and 44.44% respectively. From this, it can be inferred that faculty members in IIT Madras are very well connected and publish in collaboration with their colleagues in the department. In IIT

Madras, Kharagpur, and Kanpur, Sukhendu Das, Sarkar S., and Amitabha Mukherjee respectively, were highly connected and act as gatekeeper as well in their respective departmental co-authorship networks. These nodes (authors) are very much important for the flow of information in these networks.

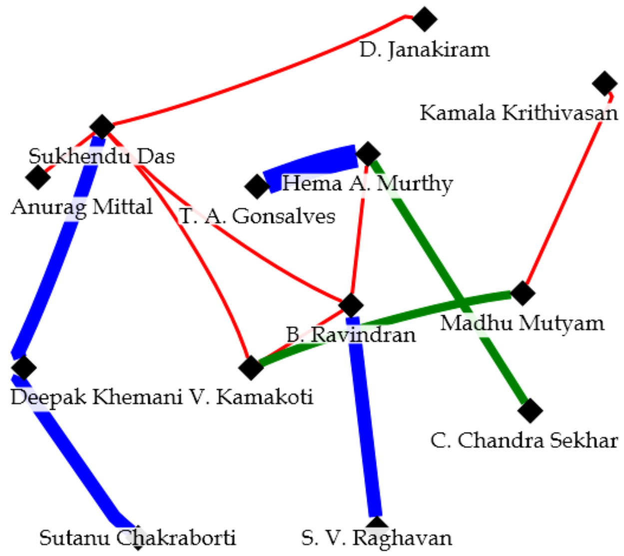

Figure 5(b): IIT Madras- Internal Collaboration Graph

7. CONCLUSIONS AND FUTURE DIRECTIONS

Co-authorship social networks are based on the co-authorship relationship (i.e. jointly conducting research or participating in a research study and presenting the results as a research publication together) and are a result of people collaborating to become co-authors. Analyses of these networks reveal the collaboration pattern and structure of the scientific community. Publishing patterns and trends of a particular group or institution can be analyzed by studying these networks. In addition, co-authorship networks also provide a platform for studying network evolution and dynamics. In this paper, we studied co-authorship networks (both internal and external) of computer science departments of the four IITs under consideration.

We demonstrated the experiments conducted with publication data for the analysis of various network parameters. We have shown the collaboration pattern in these institutions of higher and technical learning. Prominent researchers in all the four IITs were identified including gatekeepers, hubs, leaders. Detailed analysis of the generated social (academic) networks i.e. graphs and the values of various network attributes show that internal collaboration ties in departments under consideration was highest in IIT Kharagpur and lowest in IIT Delhi. IIT Madras has published the highest number of papers whereas IIT Delhi has lowest number of publications during the period under investigation.

Future extensions of this work could be the analysis of collaboration at the institutional level. This can help understand the collaboration pattern and ties among these institutions. Another extension could be the temporal analysis of the evolution pattern of these collaboration networks and their impact on the research and development activities in these institutions. The result of analysis of the collaboration ties can be improved by considering other collaboration relationships like co-supervision, project participation, etc.

8. REFERENCES

[1] Merriam-Webster's Collegiate Dictionary (1999). Tenth Edition. Springfield, MA: Merriam-Webster, Incorporated.

[2] Stroele, V., Oliveira, J., Zimbrao, G. and Souza, M.J. *"Mining and Analyzing Multirelational Social Networks."* In *Proceedings of 12th International Conference on Computational Science and Engineering,* Vancouver, Canada, 2009, pp 711-716.

[3] Patel, N. *"Collaboration in the professional growth of American sociology."* Social Science Information, 6, 1973, pp. 77-92.

[4] Glanzel, W. and Schubert, A. *"Analysing Scientific Networks Through Co-authorship."* Handbook of Quantitative Science and Technology Research, 2004, Kluwer Academic Publishers.

[5] Wasserman, S. and Faust, K. *"Social Network Analysis: Methods and Applications."* Cambridge University Press, New York City, New York, 1994, U.S.A.

[6] Newman, MEJ. *"Co-authorship networks and patterns of scientific collaboration."* In *Proceedings of the National Academy of Sciences*, 101(1), 2001, pp. 5200-5.

[7] Davis, G.F. and Greve, H.R. *"Corporate Elite Networks and Governance Changes in the 1980s." The American Journal of Sociology*, 103(1), 1997, pp. 1-37.

[8] Stuart, T.E. *"Network Positions and Propensities to Collaborate: An Investigation of Strategic Alliance Formation in a High-technology Industry." Administrative Science Quarterly*, 43(3), 1998, pp. 668-98.

[9] Chakrabarti, S. *"Mining the Web: Discovering Knowledge from Hypertext Data."* Morgan Kaufmann Publishers, USA, 2003.

[10] Tang, J., Zhang, D., and Yao, L. *"Social Network Extraction of Academic Researchers."* In Proceedings of International Conference on Data Mining-ICDM'07, Nebraska, USA, October 2007, pp 292-301.

[11] Tomobe, H., Matsuo, Y. and Hasida, K. *"Social Network Extraction of Conference Participants."* In Proceedings of 12th International Conference on World Wide Web-WWW'03, Budapest, Hungary, May 2003.

[12] Kautz, H., Selman, B., and Shah, M. *"The Hidden Web."* American Association for Artificial Intelligence magazine, 18(2), 1997, pp 27–35.

[13] Mika, P. *"Flink: Semantic web technology for the extraction and analysis of social networks."* Journal of Web Semantics, 3(2), 2005, pp 211-223.

[14] Newman, MEJ. *"Scientific collaboration Networks-II. Shortest paths, weighted networks, and centrality."* Physical Review E, Volume 64,016132

[15] Newman, MEJ. *"The structure of scientific collaboration networks."* PNAS, January, 2001, 98(2), pp 404-409

[16] Newman, MEJ. *"Co-authorship networks and patterns of scientific collaboration."* In *Proceedings of the National Academy of Sciences*, 101(90001), 2004, pp.5200-5205.

[17] Newman, MEJ. *"Scientific Collaboration Networks-I. Network Construction and Fundamental Results." Physical Review E*, 64(1):16131, 2001.

[18] Moody, J. *"The structure of a social science collaboration network: Disciplinary Cohesion from 1963 to 1999." American Sociological Review*, 69(2), 2004, pp. 213-238.

[19] Liu, X., Bollen, J. Nelson, M.L. and Van de Sompel. H. *"Coauthorship networks in the digital library research community." Information Pro-cessing & Management*, 41(6), 2005, pp.1462-1480.

[20] Sharma, M. and Urs, S.R., *"Network Dynamics of Scholarship: A Social Network Analysis of Digital Library Community."* In *Proceeding of the 2nd PhD Workshop on Information and Knowledge Management*, New York, NY, USA, 2008, pp. 101-104.

[21] Elmacioglu, E. and Lee, D. *"On six degrees of separation in dblp-db and more." SIGMOD Rec*., 34(2), 2005, pp.33-40.

[22] Ma, L., Goharian, G. and Chowdhury, A. *"Automatic data extraction template generated web pages."* In Proceedings of the *International Conference on Parallel and Distributed Processing Techniques and Applications (PDPTA 03)*, Las Vegas, Nevada, USA, 2003, pp.642–648.

[23] Milgram, S. *"The Small World Problem." Psychology Today*, 2(1), 1967, pp.60-67.

[24] Newman, M.E.J. *"Community centrality." Phys. Reviews* 74 (2006).

[25] Chelmis, C. and Prasanna, V.K. *"Social Networking Analysis: A State of the Art and the Effect of Semantics."* In *Proceedings of 3rd IEEE Conference on Social Computing (SocialCom)*, Boston, MA, 2011, pp. 531-536.

[26] http://en.wikipedia.org/wiki/Centrality

[27] http://en.wikipedia.org/wiki/Clustering_coefficient

[28] Watts, D. J. and Strogatz, S. H. *"Collective dynamics of 'small-world' networks." Nature*, 393(6684), June 1998, pp. 440–442.

[29] Hansen, D., Shneiderman, B., and Smith, M. A. "*Analyzing Social Media Networks with NodeXL: Insights from a Connected World." Morgan Kaufmann*, Burlington, MA, USA, 2009.