Scientific Collaborations

Aaloke Mozumdar
aaloke19004@iiitd.ac.in
IIIT-Delhi
India

Gitansh Raj Satija
gitansh19241@iiitd.ac.in
IIIT-Delhi
India

Karan Abrol
karan19366@iiitd.ac.in
IIIT-Delhi
India

Karanjot Singh
karanjot19050@iiitd.ac.in
IIIT-Delhi
India

Rohan Jain
rohan19095@iiitd.ac.in
IIIT-Delhi
India

CCS CONCEPTS

• **Information Retrieval**; • **Collaborative Network Analysis**;

KEYWORDS

Collaborative Networks, Social Network Analysis

ACM Reference Format:
Aaloke Mozumdar, Gitansh Raj Satija, Karan Abrol, Karanjot Singh, and Rohan Jain. 2018. Scientific Collaborations. In *Proceedings of Make sure to enter the correct conference title from your rights confirmation emai (Conference acronym 'XX)*. ACM, New York, NY, USA, 3 pages. https://doi.org/XXXXXXX.XXXXXXX

1 INTRODUCTION AND MOTIVATION

Collaboration in any domain of life tends to increase the productivity of the task. Different scientific fields have seen many new discoveries in the past century, all of which have been because of the contribution of numerous researchers building upon each other's research. When researchers collaborate, they bring different skills to the table and pave the way for new scientific developments. Studies in the past have proven that collaboration between authors leaves a positive influence on Research. This brings about a need to connect researchers to increase collaboration. The relational tie between authors may help identify long-term collaborations, common research interests, preferred conferences, and research groups under formation. Furthermore, as social ties evolve, new research interests and new collaborations can be identified. This can also help in identification of possible hidden collaboration nets. [4]

2 PROBLEM STATEMENT

Analyse the scientific collaborative network among researchers at IIIT Delhi, by studying publication data and using it to create a web platform to link researchers (both faculty and students). Further, define metrics to determine the impact of researchers on the overall collaborative network and use these to identify budding researchers and subsequently incentivize them.

Conference acronym 'XX, June 03–05, 2018, Woodstock, NY

ACM ISBN 978-1-4503-XXXX-X/18/06…$15.00
https://doi.org/XXXXXXX.XXXXXXX

3 LITERATURE REVIEW

Collaborative Network analysis has been a growing field amongst the researchers. We look at various works which provide key insights and help us formulate the scope our research.

In Cooperative Authorship Social Network [4] The authors propose an approach to build a co-authorship research social network that disseminates new publications and research connections over a multi layered architecture. The data is collected from the DBLP repository with around 677,345 authors. This service disseminates new publications and research connections to individuals who subscribe to research topics or researcher names. The process of creation of such a dissemination service takes place in six phases. First the information about research groups and researchers are mined from the Web or provided by individuals. Next, Their publications are organized in semi-structured data in a Digital Library. In the next steps, A DL interactive process feeds relevant information to build the network and a dissemination service evaluates & processes this data. The paper also presents approaches to determine two types of associations namely Collaboration in Co-authorship and Collaboration in Research Areas. The performance of this Automatic SN, based on occurrence of co-authorship between authors is compared with a Manual SN designed by a specialist It is observed that the Manual SN covers 93.44% of the Automatic SN. The Automatic SN covers 83.82% of the Manual SN.
In the research paper [3], the authors propose an approach to analyse the structure of the co-authorship network among researchers at the Italian Institute of Technology. The paper uses multiple metrics like centrality measures, density, degree distribution, degree centralisation, clustering coefficient, connectivity, and diameter. The metrics were defined as follows:

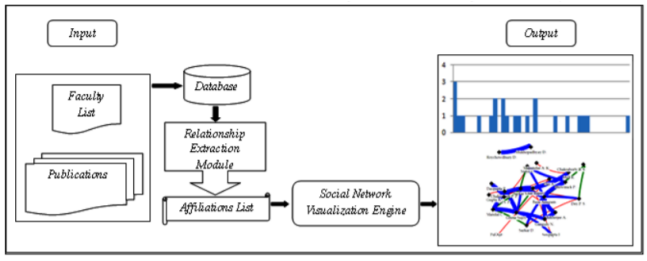

Figure 1: Social Network Extraction System Architecture [1]

- Degree Centrality - It is the number of researchers that a particular researcher is collaborating with.
- Closeness Centrality - It signifies the fact that a central actor is also the one who has direct collaborations with many other researchers and thus can reach them through a very short path in the graph.
- Betweenness Centrality - It highlights the ability of a researcher to play a mediating role between other researchers thus playing a central role.
- Clustering Coefficient - It is a measure of the transitive nature of a researcher.
- Density - It is a measure of the number of edges in the graph compared to the total number of possible edges. The more dense a graph, the more collaborative is its nature
- Degree Centralization - It quantifies the range of variability of the degree of each individual actor. A lower degree centralization suggests that the graph has multiple central actors and as it increases the graph becomes more and centred around a few researchers.

These centrality metrics measure not only the author's productivity and impact, but also his role within the community and his influence in spreading information.
The authors concluded that researchers who played a bridging role between researchers published better quality papers. They also found that researchers collaborated within the institute more than external collaboration. The importance of seniority of a researcher was also highlighted and it was concluded that it is much more beneficial to retain a researcher than to replace them.
For future research, they emphasized the importance of human capital of the institute to calculate its growth. The most important limitation of the study was that they used only the citation and publication counts to evaluate research performance, other bibliometric indicators could have been used.

In Scientific Co-authorship Social Networks [1], the authors study the collaborative networks of the computer science departments of 4 major IITs - Delhi, Kanpur, Kharagpur and Madras and try to answer the following questions - Who are the hub/leaders? Who has more connections? How strong are collaboration ties? How collaborative are the authors? To answer these questions the authors use the following metrics - Betweenness Centrality, Degree Centrality, Clustering Coefficient and Average Degree respectively. The authors observed that in IIT Kanpur the central authors were also acting as hubs. However this was not the case within IIT Delhi. This implies that high degree centrality doesn't directly imply high betweenness centrality and vice versa. Further, concerning the strength of collaboration ties, it was observed that IIT Kanpur has the lowest clustering coefficient. This highlights lack of connectivity and that very few nodes are arranged into cliques (Complete graphs), this can further be validated by the large diameter of the network in IIT Kanpur and average path length between nodes. IIT Kharagpur has the highest clustering coefficient and majority of the nodes are clustered into cliques. Also it has the largest connected component. This implies that information flows much faster through the network in IIT Kharagpur than in Kanpur. In essence this paper gives us insight into measuring the strength of collaborations in institutes and in future studies we can study the effects of these on their overall research throughput.

In 2005, the Ministry of Health together with the Ministry of Science and Technology of Brazil, launched a program to study certain 'neglected' diseases (diseases present in developing countries which are ignored by the big pharmaceutical companies) that are prevalent in the poor and marginalized regions [5]. The program aimed to foster technological innovation in the research that it was funding. In this study, the authors recognized that evaluating the scientific productivity in these regions was of course a crucial task which would help the organizations better the allocation of their funds and traditional indicators like the number of articles would not be of much value. So they use SNA to develop new approaches and criteria to allow for a more comprehensive view of the productivity of the research being conducted in a region.
Publications on the seven diseases that were the interest of this program were retrieved from the 'Web of Knowledge' database of the Institute for Scientific Information (ISI) having at least 1 Brazilian author. They created symmetrical, co-occurrence matrices using the authorship data from these publications in rows and columns in order to map co-authorships between authors (authors - authors matrices) or institutions (institutions - institutions matrices).
The authors have majorly focused on studying the components and cut points formed in the network to analyze the productivity of research done in different regions and fields. Component analysis gives us an idea of the overall network structure. It reveals how separated or clustered it is and therefore providing valuable insights on possible collaborations. In the dengue researcher's network, each component in the network is identified by key words that the researchers had determined. The analysis of the 9 major components suggested, for instance, a collaboration between component III and VIII , which didn't have any edges connecting them, but the subject of both of their research interests was mainly around dengue vector control. Network cut points in the institute networks indicate actors responsible for connecting several other institutes in different Regions and are therefore essential to grow and encourage collaboration. This also throws light on one shortcoming of traditional indicators like H-index that don't factor in this aspect.

Researchers from University of Southern California [2] recently conducted a study to for analyzing interactions between researchers and institutions. They were able to establish a superlinear relation between the number of active researchers and institute size. They also made an interesting observation that establishment of new institutions can 'trigger' even more potential institutions. IIIT-Delhi being a relatively new institution can be foundation for one such 'trigger'.

The above works motivated us to analyze the trends in co-authorship and research in our Institute and see if any fruitful and insightful results can be gained.

4 DATA

We will be using research publication data from Irins (scraping), Google Scholar platform (SerpApi Tool). IIIT-Delhi has 103 active

professors who have contributed in more than 2800 publications. We also have an active community of Undergraduate and Postgraduate students who are actively involved in the research domain. Thus, apart from the faculty we will also be extending our dataset to student publications.

The data will be represented using an undirected weighted graph G=(V, E,W). The vertices/nodes (V) will represent the researcher. The researchers will be connected with Edge (E) of Weight (W), where W is the number of collaborated publications the researchers have in common. Thus this graph will contain as many edges as the number of collaborations the IIIT-Delhi Researchers have made. The total number of nodes will be equal to the number of researchers from IIIT-Delhi and the number of external collaborators.

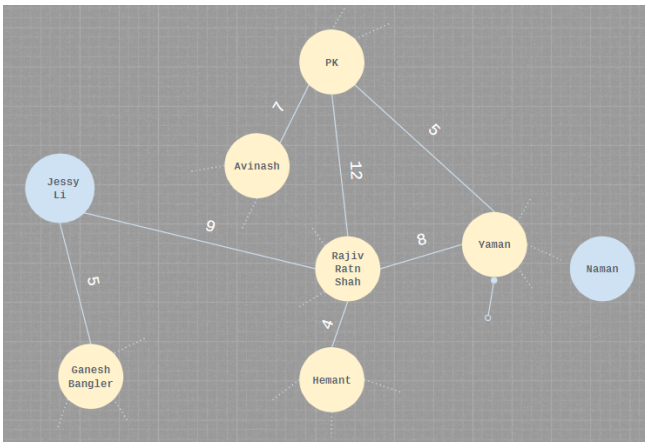

Figure 2: We see that the researcher nodes are connected with each other through edges of different weights, which represents the collaborations two authors have had. The yellow color nodes represent researchers from IIIT-Delhi, and the blue color nodes are external collaborators. Note: The values in the graph are for representational use only and may be different from actual values.

5 PLAN OF WORK

These are the 4 major milestones of our Project:

- Collecting Data: Publications indexed by the particular researcher's belonging to IIIT-Delhi. involved. (Ongoing)
- Co-Authorship network Assembly using that indexed Data (Ongoing)
- Defining important Metrics and using these metrics for Network Analysis
- Identifying a novel scoring system for co-authorship of the researchers
- Creating a Web Platform to link researchers and help visualize the collaborations that have happened over the years in IIIT-Delhi.

REFERENCES

[1] Tasleem Arif, Rashid Ali, and M. Asger. 2012. Scientific Co-authorship Social Networks: A Case Study of Computer Science Scenario in India. *International Journal of Computer Applications* 52 (08 2012), 38–45. https://doi.org/10.5120/8257-1790

[2] Keith A. Burghardt, Zihao He, Allon G. Percus, and Kristina Lerman. 2021. The Emergence of Heterogeneous Scaling in Research Institutions. arXiv:2001.08734 [physics.soc-ph]

[3] Enrico di Bella, Luca Gandullia, and Sara Preti. 2021. Analysis of scientific collaboration network of Italian Institute of Technology. *Scientometrics* 126, 10 (October 2021), 8517–8539. https://doi.org/10.1007/s11192-021-04120-

[4] Giseli Lopes, Mirella Moro, Leandro Wives, and José Palazzo Moreira de Oliveira. 2010. Cooperative Authorship Social Network. *CEUR Workshop Proceedings* 619.

[5] Carlos Medicis Morel, Suzanne J Serruya, Gerson de Oliveira Penna, and Reinaldo Guimarães. 2009. Co-authorship Network Analysis: A Powerful Tool for Strategic Planning of Research, Development and Capacity Building Programs on Neglected Diseases. *PLoS Neglected Tropical Diseases* 3 (2009).