

# Shubhangi Agarwal

🏠 [shubhangiat.github.io](https://shubhangiat.github.io) | [in shubhangi-agarwal](https://www.linkedin.com/in/shubhangi-agarwal)

✉ [sagarwal.mail@gmail.com](mailto:sagarwal.mail@gmail.com)

## SUMMARY

I am a researcher with an interest in **Graphs**. I have developed various algorithms based on statistical analysis for **Subgraph Querying** in large complex graphs. Some of the many areas that can benefit from the querying of subgraph structures are information extraction, recommendation systems, disease diagnostics, fraud detection. During my post-doc, I worked on **HyGraphs**, where as a team we explored development of a novel data model combining property graphs and time series data models, with an aim to provide a single platform for seamless, intuitive and efficient processing of large scale datasets which are dynamic in nature.

## EDUCATION

### Ph.D. in Computer Science and Engineering

Indian Institute of Technology Kanpur, Uttar Pradesh, India

CGPA: 8.25

2014 - 2023

### Bachelor of Technology Computer Engineering

Sardar Vallabhbhai National Institute of Technology, Surat, Gujarat, India

CGPA: 8.71

2010 - 2014

## PHD THESIS

### Subgraph Matching and Mining in Large Graphs

Supervisor: Arnab Bhattacharya

- Developed algorithms for Approximate Subgraph Matching in both deterministic and probabilistic graphs.
- Proposed a Graph Neural Network model for robust node embeddings with positional information.

## PUBLICATIONS

- “A Case Study of Cross-Lingual Zero-Shot Generalization for Classical Languages in LLMs”, V S D S Mahesh Akavarapu, Hrishikesh Terdalkar, Pramit Bhattacharyya, **Shubhangi Agarwal**, Vishakha Deulgaonkar, Chaitali Dangarikar, Pralay Manna, and Arnab Bhattacharya, 63rd Annual Meeting of the Association for Computational Linguistics (ACL), 2025, Austria.
- “Towards Hybrid Graphs: Unifying Property Graphs and Time Series”, Mouna Ammar, Christopher Rost, Riccardo Tommasini, **Shubhangi Agarwal**, Angela Bonifati, Petra Selmer, Evgeny Kharlamov and Erhard Rahm, 28th International Conference on Extending Database Technology (EDBT), 2025, Spain.
- “Transforming Time Series into Graphs and Back with HyGraph”, Mouna Ammar, **Shubhangi Agarwal**, Angela Bonifati and Erhard Rahm, Proceedings of the Workshops of the EDBT/ICDT Joint Conference, 2025, Spain.
- “VeNoM: Approximate Subgraph Matching with Enhanced Neighbourhood Structural Information”, **Shubhangi Agarwal**, Sourav Dutta and Arnab Bhattacharya, 7th Joint International Conference on Data Science and Management of Data (CODS-COMAD), 2024, India.
- “VerSaChI: Finding Statistically Significant Subgraph Matches using Chebyshev’s Inequality”, **Shubhangi Agarwal**, Sourav Dutta and Arnab Bhattacharya, Proceedings of the International Conference on Information and Knowledge Management (CIKM), 2021, pages 2812-2816, Australia.
- “GraphReach: Position-Aware Graph Neural Network using Reachability Estimations”, Sunil Nishad, **Shubhangi Agarwal**, Arnab Bhattacharya and Sayan Ranu, Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI), 2021, pages 1527-1533, Canada.
- “ChiSeL: Graph Similarity Search using Chi-Squared Statistics in Large Probabilistic Graphs”, **Shubhangi Agarwal**, Sourav Dutta and Arnab Bhattacharya, Proceedings of the International Conference on Very Large Data Bases (VLDB), 2020, pages 1654-1668, Japan.

## EXPERIENCES

### Post-doctoral Researcher

May 2024 - Apr 2025

- Worked on the ANR-DFG project ‘HyGraph: Querying and Analytics on Hybrid Graphs’ with a mid-sized team.

### Senior Student Research Associate

Aug 2022 - Dec 2023

- Worked on the project ‘Short and Long-term Fog Prediction for North Indian Cities’.

### Program Committee Member

2025 - current

- GRADES-NDA Workshop - SIGMOD/ PODS (2025), ICWE (2025), TGD Workshop - EDBT/ ICDDT (2025).

## External Reviewer

2020 - current

- WSDM (2024), CIKM (2021, 2022), DASFAA (2022), CoDS-COMAD (2020, 2021), KDD (2021).

## Intern with Flipkart Internet Private Limited

Jun 2015 - Jul 2015

- Worked in a project to prepare a model for identification of 'fake' listings.

## Summer Research Internship at IIT Gandhinagar

May 2012 - July 2012

- Worked under Prof. Pawan Lingras on Time Series Analysis

## TEACHING

---

### Instructor

Jan 2025 – Mar 2025

*Claude Bernard University Lyon 1*

- Program: M2 - Data and Intelligence for Smart Systems (DISS)
- Course: Data Processing and Analytics (Jointly with others)

### Teaching Assistant (IIT Kanpur)

Aug 2014 - Apr 2021

- Graded and evaluated projects for various courses of Computer Science.

### Senior Tutor (IIT Kanpur)

Aug 2017 - Apr 2018, Aug 2019 - Apr 2020

- Led teams of strength ~60; Assisted in backend management, paper-setting and grading.

### Teaching Assistant for MOOC (NPTEL - Remote)

July 2017 - Sep 2017

- Crafted objective questions and resolved student queries on Fundamentals of Database Systems.

## TECHNICAL SKILLS

---

**Languages** : C, C++, Python, Java, R, JavaScript, PHP, Shell scripting  
**Databases** : MySQL, Neo4j  
**Libraries** : PyTorch, Tensorflow, Scikit-learn, Numpy, Pandas  
**Tools** : git, LaTeX, Docker, Weka

## REFERENCES

---

### Arnab Bhattacharya

*Professor, Indian Institute of Technology Kanpur, India*

🏠 <https://cse.iitk.ac.in/users/arnabb/>

✉ [arnabb@iitk.ac.in](mailto:arnabb@iitk.ac.in), [arnabb@cse.iitk.ac.in](mailto:arnabb@cse.iitk.ac.in)

☎ +91-512-259-7650

### Sourav Dutta

*Chief NLP Research Scientist, Huawei Research Centre, Dublin, Ireland*

🏠 <https://sites.google.com/view/homesouravdutta/>

✉ [sourav.dutta2@huawei.com](mailto:sourav.dutta2@huawei.com)

### Angela Bonifati

*Distinguished Professor, Lyon 1 University, France*

🏠 <https://perso.liris.cnrs.fr/angela.bonifati/index.shtml>

✉ [angela.bonifati@gmail.com](mailto:angela.bonifati@gmail.com)

### Riccardo Tommasini

*Associate Professor, Lyon 1 University, France*

🏠 <https://www.riccardotommasini.com>

✉ [riccardo.tommasini@insa-lyon.fr](mailto:riccardo.tommasini@insa-lyon.fr)