Soumya Sarkar

Curriculum Vitae

19 Y, Abinash Chandra
Banerjee Lane, Kolkata 700010
(+91) 8585027859
soumya.sarkar.work@gmail.com
Sam131112

Research Interests

Natural Language Processing, Data Mining, Machine Learning

Education

2015 – 2020 **Doctor of Philosophy (Ph.D.)**, *Computer Science and Engineering*,
Indian Institute of Technology, Kharagpur,
Advisor: Prof. Animesh Mukherjee.

2013 – 2015 **Master of Technology (M.Tech.)**, *Computer Science and Engineering*,
Indian Institute of Technology, Patna.
8.9/10

2008 – 2012 **Bachelor of Technology (B.Tech.)**, *Information Technology*,
RCC Institute of Information Technology, Kolkata.
8.31/10

Present Affiliation

Jan 2021 – present **PostDoc Researcher**, *UKP Lab*, TU Darmstadt,
Advisor: Prof. Dr. Iryna Gurevych.

Work Experience

Oct 2020 – Dec 2020 **Research Intern**, *Ben Gurion University of the Negev*.

June 2019 – Sep 2019 **Applied Scientist Intern**, *Amazon India Machine Learning*.

May 2017 – July 2017 **Internship at**, *TCS Innovation Labs*.

Sep 2015 – May 2016 **Externship at**, *Flipkart Pvt. Ltd.*.

May 2015 – Aug 2015 **Sponsored Reseach and Industrial Consutancy, IIT Kharagpur**,
Project Fellow, sponsored by, ITRA Media Labs Asia.

Aug 2012 - Dec 2012 **Cognizant Technologies Pvt. Ltd**,
Graduate Trainee.

Conference Publications

1. Das, P., Reddy, B. P., Chakraborty, D., **Sarkar, S.** and Mukherjee, A. (2021). When expertise gone missing: Uncovering the loss of prolific contributors in Wikipedia. In ICADL 2021, Virtually (online mode) *Full Paper*
2. Bhanu Prakash Reddy*, Sasi Bhushan*, **Soumya Sarkar***, Animesh Mukherjee (*equal contribution) *NwQM: A neural quality assessment framework for Wikipedia* In Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP) (pp. 8396-8406) *Full Paper*
3. **S. Sarkar**, B. P. Reddy, S. Sikdar, and A. Mukherjee. *StRE: Self attentive edit quality prediction in Wikipedia.In Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, pages 3962-3972, Florence, Italy, July 2019. Association for Computational Linguistics.Full Paper*
4. **Soumya Sarkar**, Sanjukta Bhowmick, and Animesh Mukherjee. *On Rich Clubs of Path-Based Centralities in Networks. In Proceedings of the 27th ACM International Conference on Information and Knowledge Management,*

pp. 567-576. ACM, 2018. Full Paper

5. **Soumya Sarkar**, Aditya Bhagwat , Animesh Mukherjee *Core2Vec: A Core-Preserving Feature Learning Framework for Networks.* In 2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM), pp. 487-490. IEEE, 2018. *Short Paper*
6. Sandipan Sikdar, Tanmoy Chakraborty, **Soumya Sarkar**, Niloy Ganguly, and Animesh Mukherjee. 2018. ComPAS: Community Preserving Sampling for Streaming Graphs. In Proceedings of the 17th International Conference on Autonomous Agents and MultiAgent Systems (AAMAS '18). International Foundation for Autonomous Agents and Multiagent Systems, Richland, SC, 184-192. *Full Paper*
7. **Soumya Sarkar**, Suhansanu Kumar, Sanjukta Bhowmick, and Animesh Mukherjee. *Sensitivity and reliability in incomplete networks: Centrality metrics to community scoring functions.* In 2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM), pp. 69-72. IEEE, 2016. *Short Paper*
8. Ufimtsev, Vladimir, **Soumya Sarkar**, Animesh Mukherjee, and Sanjukta Bhowmick. *Understanding stability of noisy networks through centrality measures and local connections.* In Proceedings of the 25th ACM International on Conference on Information and Knowledge Management, pp. 2347-2352. ACM, 2016. *Extended Short Paper*

Journal Publications

1. Adak, S., Chakraborty, S., Das, P., Das, M., Dash, A., Hazra, M., Mathew, B., Saha, P., **Sarkar, S.** and Mukherjee, A. (2021). Mining the Online Inphosphere: A survey. In WIREs Data Mining and Knowledge Discovery journal, (in press)
2. Binny Mathew, Anurag Illendula, Punyajoy Saha, **Soumya Sarkar**, Pawan Goyal, Animesh Mukherjee *Hate begets Hate: A Temporal Study of Hate Speech.* Proceedings of the ACM on Human-Computer Interaction (CSCW 2020) *Link*
3. **Soumya Sarkar**, Sandipan Sikdar, Sanjukta Bhowmick, and Animesh Mukherjee. *Using core-periphery structure to predict high centrality nodes in time-varying networks.* Data Mining and Knowledge Discovery 32, no. 5 (2018): 1368-1396. accepted in ECML PKDD 2018 Journal Track *Link*
4. **Soumya Sarkar**, Aditya Bhagwat, Animesh Mukherjee "A core-periphery structure based network embedding approach", Journal of Social Network Analysis and Mining (SNAM), Springer, April, 2021.

Book Chapters

1. **Soumya Sarkar**, Suhansanu Kumar, Sanjukta Bhowmick, and Animesh Mukherjee. *Centrality and Community Scoring Functions in Incomplete Networks: Their Sensitivity, Robustness, and Reliability.* In Machine Learning Techniques for Online Social Networks, pp. 135-154. Springer, Cham, 2018.*Link*
2. **S. Sarkar**, A. Karn, S. Bhowmick, A. Mukherjee. *An Empirical Study of the Effect of Noise Models on Centrality Metrics.* Dynamics on and off Complex Networks III (Editors: Ghanbarnejad, F., Saha Roy, R., Karimi, F., Delvenne, J.-C., Mitra, B.) ISBN: 978-3-030-14682-5

Workshop, Poster Publications

1. Vladimir Ufimtsev, **Soumya Sarkar**, Sanjukta Bhowmick, Animesh Mukherjee *Identifying Stable Networks* Poster in Proceedings of *SIAM Workshop on Network Science (NS16) Link*
2. Pandey, Arun, Roshni Chakraborty, **Soumya Sarkar**, and Joydeep Chandra. *Analyzing link dynamics in scientific collaboration networks: A social yield based perspective.* In Proceedings of the 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2015, pp. 1395-1402. ACM, 2015. *Full Paper*

Academic Achievements and Awards

- **IARCS** travel grant for attending *EMNLP 2020*
- **ACL Travel Grant, Microsoft Travel Grant** for attending *ACL 2019*
- **SIGIR Travel Grant, Microsoft Travel Grant** for attending *CIKM 2018*
- **Microsoft Travel Grant** for attending *ECML PKDD 2018*
- **IARCS** travel grant for attending *AAMAS 2018*
- **ACM Travel Award** for attending *ACM CODS-COMAD 2019* Kolkata
- Cleared **Graduate Aptitude Test 2013** with a *percentile of 99.8* (225,000 applicants)

Skills

Languages	Java, C++, Python, R
Tools	Intellij Idea, Matlab, Gephi, Weka, LaTeX
Operating System	Windows, Linux, Macintosh

Teaching Experience

Text Mining	TU Darmstadt, Winter 2021-22

Teaching Assistanship

Information Retrieval	IIT Kharagpur, Autumn 2019
Approximation Algo	IIT Kharagpur, Spring 2019
Information Retrieval	IIT Kharagpur, Autumn 2018
Machine Learning	IIT Kharagpur Spring 2018
FADML	IIT Kharagpur Autumn 2017
PDS Lab	IIT Kharagpur Spring 2017
Social Computing	IIT Kharagpur Autumn 2016
Complex Network	IIT Kharagpur Spring 2016
Discreet Structures	IIT Kharagpur Autumn 2015

Declaration

I hereby declare that all information above are correct to the best of my knowledge and belief.

Soumya Sarkar
Kolkata,
September 23, 2021