

Curriculum Vitae of Tanay Kumar Saha

Contact

742 Blake Street, Apt F, Indianapolis, IN 46202, USA

Phone: +1 (317) 991 2340, Email: tksaha@purdue.edu

[Per. Webpage]; [Google Scholar]; [Semantic Scholar]; [Scopus Author]; [ResearchGate]; [FigShare's]

Research Interest

Machine Learning: Representation Learning, Deep Learning, Reinforcement Learning, Graphical Models

Natural Language Processing: Representation Learning of Textual units, Cross-Lingual and Multilingual Learning, Text Summarization, Compositional Semantics, Open Domain Question Answering, Total Recall

Network Analysis: Representation Learning of Network units, Link Prediction, Mining Higher order relation

Teaching Experience

2012–2018	IUPUI, Indiana, USA Tutor and grader undergraduate level Discrete Mathematics, Data Mining, and Artificial Intelligence course and graduate level Data Mining and Algorithm in Bio-informatics course. (Earned TA Award in 2017)	Graduate Teaching Assistant
2011–2012	Jagannath University, Dhaka, Bangladesh (http://jnu.ac.bd/) Conducted undergraduate level Data Structure and Algorithm courses	Lecturer
2010–2011	UITS, Dhaka, Bangladesh (https://www.uits.edu.bd/) Conducted undergraduate level Data Structure, Algorithm and Programming Language course	Lecturer

Research Lab Experience

Aug'17-	CareerBuilder, Atlanta, GA NDA (Representation Learning/Graph)	Research Collaborator
05/17–08/17	NEC Laboratories, Princeton, New Jersey Direct Supervisor(s):- Jianwu Xu, Hui Zhang. <ul style="list-style-type: none">– Adding semantic component in NGLA (http://www.nec-labs.com/research/ngla-next-generation-log-analytics/554)– Contextual log failure signature generation	System Research Group Intern
01/16–06/16	QCRI, Doha, Qatar (http://qcri.com/) Direct Supervisor:- Mourad Ouzzani. Worked on: <ul style="list-style-type: none">– Improving a 5-star rating system for Rayyan (https://rayyan.qcri.org/)– Providing a duplicate detection pipeline for Rayyan (https://rayyan.qcri.org/)	Data Analytics Research Intern
12/15 – 8/16	iControlESI, Dallas, Texas Designed an active learning system for predictive coding (http://www.icontrolési.com/envize/)	Research Collaborator
08/12–05/18	Data Mining Lab, IUPUI, Indiana, USA Thesis Supervisor:- Mohammad Al Hasan. Worked on: <ul style="list-style-type: none">– Latent space Representation of Sentences; Link Prediction in Dynamic Network; Batch mode active learning algorithm for TAR (Corresponding Product: http://www.icontrolési.com/predictive-coding/); MCMC based Graph Mining, Network Motif Finding Algorithm; Directed graphlet sampling for Android malware detection; Name Disambiguation problem using only link data	Graduate Research Assistant

Education

2012-2018	PhD Candidate Department of Computer Science Thesis Title:- Latent representation and Sampling in Network: Application in text mining and biology Advisor:- Mohammad Al Hasan, Co-Advisor:- Jennifer Neville	Purdue University, West Lafayette
2012-2015	MSc Department of Computer and Information Science	Indiana University - Purdue University Indianapolis (IUPUI)
2004-2009	BSc Department of Computer Science and Engineering	Bangladesh University of Engineering and Technology (BUET)

Publications

Journal Papers

- [1] Discovery of Functional Motifs from the Interface Region of Oligomeric Proteins using Frequent Subgraph Mining
Tanay Kumar Saha, Ataur Katebi, Wajdi Dhifli, Mohammad Al Hasan
IEEE/ACM Transactions on Computational Biology and Bioinformatics (2017). 2017.
- [2] Name disambiguation from link data in a collaboration graph using temporal and topological features
Tanay Kumar Saha, Baichuan Zhang, Mohammad Al Hasan
Social Network Analysis and Mining 5.1 (2015) pp. 1–14. Springer Vienna, 2015, Springer Vienna.
- [3] FS³: A sampling based method for top-k frequent subgraph mining [Code: <https://github.com/tksaha/fs3-graph-mining>]
Tanay Kumar Saha, Mohammad Al Hasan
Statistical Analysis and Data Mining 8.4 (2015) pp. 245–261. Wiley Online Library, 2015, Wiley Online Library.

Conference Papers

- [1] Regularized and Retrofitted models for Learning Sentence Representation with Context (Acceptance Rate: 21%)
Tanay Kumar Saha, Shafiq Joty, Naeemul Hassan, Mohammad Al Hasan
Proceedings of the 26th ACM International Conference on Information and Knowledge Management, CIKM, 2017.
- [2] Con-S2V: A Joint Learning framework for incorporating Extra-Sentential Context into Sen2Vec (Acceptance Rate: 27%) [Code: <https://github.com/tksaha/con-s2v>]
Tanay Kumar Saha, Shafiq Joty, Mohammad Al Hasan
Machine Learning and Knowledge Discovery in Databases - European Conference, ECML PKDD, 2017.
- [3] ACTS: Extracting Android App Topological Signature through Graphlet Sampling (Acceptance Rate: 29%)
Wei Peng, Tianchong Gao, Devkishen Sisodia, **Tanay Kumar Saha**, Feng Li, Mohammad Al Hasan
IEEE Conference on Communications and Network Security, 2016.
- [4] Discovery of Functional Motifs from the Interface Region of Oligomeric Proteins using Frequent Subgraph Mining Method
Tanay Kumar Saha, Ataur Katebi, Mohammad Al Hasan
15th International Workshop on Data Mining in Bioinformatics (BIOKDD'16), 2016.
- [5] Finding network motifs using MCMC sampling [Code: <https://github.com/tksaha/motif-finding>]
Tanay Kumar Saha, Mohammad Al Hasan
Complex Networks VI, 2015, Springer International Publishing.
- [6] Batch-mode active learning for technology-assisted review
Tanay Kumar Saha, Mohammad Al Hasan, Chandler Burgess, Md Ahsan Habib, Jeff Johnson
IEEE International Conference on Big Data, 2015.
- [7] Name disambiguation from link data in a collaboration graph
Tanay Kumar Saha, Baichuan Zhang, Mohammad Al Hasan
2014 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM), 2014.

Poster Papers

- [1] Finding network motifs using MCMC sampling
Tanay Kumar Saha, Mohammad Al Hasan
18th Annual International Conference on Research in Computational Molecular Biology (RECOMB), 2014.

Working/Submitted Papers

- [1] Effective Feature Representation for Link prediction in Dynamic Networks
Mahmudur Rahman, **Tanay Kumar Saha**, Mohammad Al Hasan, Kevin S. Xu, Chandan K. Reddy
Machine Learning Journal, 2017.
- [2] Study of Methods for Abstract Screening in a Systematic Review Platform
Tanay Kumar Saha, Mourad Ouzzani, Hossam Hammady, Ahmed K. Elmagarmid, Mohammad Al Hasan
Journal of Biomedical Informatics, 2017.

Provisional Patent Application

- [1] Method and System for Log Based Computer Server Failure Diagnosis
Jianwu Xu, Hui Zhang, Pranay Anchuri, **Tanay Kumar Saha**
2017.
- [2] Apparatus and Method of Implementing Batch-Mode Active Learning for Technology-Assisted Review of Documents
Jeffrey A Johnson, Md Ahsan Habib, Chandler L Burgess, **Tanay Kumar Saha**, Mohammad Al Hasan
US Patent App. 15/260,444, 2016.
- [3] Apparatus and Method of Implementing Enhanced Batch-Mode Active Learning for Technology-Assisted Review of Documents
Jeffrey A Johnson, Md Ahsan Habib, Chandler L Burgess, **Tanay Kumar Saha**, Mohammad Al Hasan
US Patent App. 15/260,538, 2016.

Professional Services

2017-Now	PC Member Pleased to serve as a PC Member of workshop on large-scale graph data mining and management (BigGraphs 2017, http://www.biggraphs.org)	Researcher, Graduate Student
2014-Now	External Reviewer (Conference) Pleased to serve as an external Reviewer of AAAI (2017), IEEE Big Data (2017), KDD (2017), ICDM (2016), DSAA (2014), ICIT (2015), SAC (2016)	Researcher, Graduate Student
2012-Now	External Reviewer (Journal) Pleased to serve as an external Reviewer of TKDD	Researcher, Graduate Student

Voluntary Services

2012-Now	IUPUI Pleased to serve as a President of Bangladesh Student Association (BDSA IUPUI); an International Graduate Welcome Volunteer (IGWV); a Treasurer for Asian Student Union (ASU) and a Member in School of Science Graduate Student Council (SOSGSC)	Researcher, Graduate Student
----------	---	------------------------------

Awards

2005-Now

BUET, IUPUI, BD GOVT.

Researcher, Graduate Student

- School of Science TA Award, IUPUI, 2017; Graduate Research Assistantship for pursuing PhD in Purdue University, West Lafayette
- 2nd Prize in International Project Show organized by BUET; Dean's List in Level 2, 3, and 4 in CSE, BUET for maining CGPA > 3.75 in two consecutive terms; Top-10 Merit List in Level 2, and 3 in CSE, BUET; Imdad Sitara Khan Scholarship; Talent-pool Scholarship from Bangladesh Government in all Grades (from Grade 5-12)

Communication and Collaboration Skills

2014 - 2015

Oral Presentation

ASONAM (2014) , IEEE Big Data (2014), CompleNet (2015), IEEE Big Data (2015)

Presented the research I conducted— ASONAM (Name Disambiguation), 2014 IEEE Big Data (Graph Mining), CompleNet (Motif Finding), 2015 IEEE Big Data (Batch-Mode Active Learning)

2014

Poster

RECOMB (2014)

Presented the initial results on the Motif Finding Problem

2014 - 2017

Collaboration

eBay, CareerBuilder, NEC, QCRI, NIH, iControl ESI, IUPUI

Collaborated with a group of Skilled Researchers/Software Engineers from eBay, CareerBuilder, NEC Laboratories, iControl ESI, QCRI and NIH

References

Mohammad Al Hasan, Associate Professor
Department of Computer and Information Science
Indiana University - Purdue University Indianapolis, Indiana, IN 46202
Email: alhasan@iupui.edu

Shafiq Rayhan Joty, Assistant Professor
Office: Block N4, 2c-111
School of Computer Science and Engineering
Nanyang Technological University, Singapore
Email: srjoty@ntu.edu.sg

Mourad Ouzzani, Principal Scientist
Data Analytics, Qatar Computing Research Institute
Email: mouzzani@hbku.edu.qa

Sanjay Chawla, Director
Data Analytics, Qatar Computing Research Institute
Email: schawla@hbku.edu.qa

Prasenjit Mitra, Professor
College of Information Sciences and Technology
Department of Computer Science and Engineering
The Pennsylvania State University, University Park, PA 16802
Email: pmitra@ist.psu.edu