Curriculum Vitae of Tanay Kumar Saha

Contact

835 Lockefield 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

Research Lab Experience

Aug'17- **CareerBuilder, Atlanta, GA** — Research Collaborator
NDA (Representation Learning/Graph)

05/17–08/17 **NEC Laboratories, Princeton, New Jersey** — System Research Group Intern
Direct Supervisor(s):- Jianwu Xu, Hui Zhang.
- Adding semantic component in NGLA (http://www.nec-labs.com/research/ngla-next-generation-log-analytics/554)
- Contextual log failure signature generation

01/16–06/16 **QCRI, Doha, Qatar (http://qcri.com/)** — Data Analytics Research Intern
Direct Supervisor:- Mourad Ouzzani. Worked on:
- Improving a 5-star rating system for Rayyan (https://rayyan.qcri.org/)
- Providing a duplicate detection pipeline for Rayyan (https://rayyan.qcri.org/)

12/15 – 8/16 **iControlESI, Dallas, Texas** — Research Collaborator
Designed an active learning system for predictive coding (http://www.icontrolesi.com/envize/)

08/12–05/18 **Data Mining Lab, IUPUI, Indiana, USA** — Graduate Research Assistant
Thesis Supervisor:- Mohammad Al Hasan. Worked on:
- Latent space Representation of Sentences; Link Prediction in Dynamic Network; Batch mode active learning algorithm for TAR (Corresponding Product: http://www.icontrolesi.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

Education

2012-2018 **PhD** Candidate — Purdue University, West Lafayette
Department of Computer Science
Thesis Title:- Latent representation and Sampling in Network: Application in text mining and biology

2012-2015 **MSc** — Indiana University - Purdue University Indianapolis (IUPUI)
Department of Computer and Information Science

2004–2009 **BSc** — Bangladesh University of Engineering and Technology (BUET)
Department of Computer Science and Engineering

Publications

Journal Papers

[1] Discovery of Functional Motifs from the Interface Region of Oligomeric Proteins using Frequent Subgraph Mining [Code: `https://gitlab.com/tksaha/func_motif`]
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] FS^3: 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] Android Malware Detection via Graphlet Sampling
Tianchong Gao, Wei Peng, Devkishen Sisodia, **Tanay Kumar Saha**, Feng Li, Mohammad Al Hasan
Transactions on Mobile Computing, 2018.

[2] Models for Capturing Temporal Smoothness in Evolving Networks for Learning Latent Representation of Nodes [Code: *https://gitlab.com/tksaha/temporalnode2vec.git*]
Tanay Saha, Thomas Williams, Mohammad Hasan, Shafiq Joty, Nicholas K. Varberg
KDD, 2018.

[3] 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.

[4] 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.

Certifications

2018-Now **Coursera certificates** Researcher
Neural Networks and Deep Learning, Improving Deep Neural Networks: Hyperparameter tuning, Regularization and Optimization, Structuring Machine Learning Projects, Sequence Models

Professional Services

2017-Now **PC Member** Researcher, Graduate Student
Pleased to serve as a PC Member of workshop on large-scale graph data mining and management (BigGraphs 2017, http://www.biggraphs.org)

2017-Now **Reviewer** Researcher, Graduate Student
Pleased to serve as a Reviewer of TKDE, TOIS

2014-Now **External Reviewer (Conference)** Researcher, Graduate Student
Pleased to serve as an external Reviewer of AAAI (2017), IEEE Big Data (2017), KDD (2017), ICDM (2016), DSAA (2014), ICIT (2015), SAC (2016)

2012-Now **External Reviewer (Journal)** Researcher, Graduate Student
Pleased to serve as an external Reviewer of TKDD

Awards

2012-now **IUPUI** Researcher, Graduate Student

- Roche Hypo Challenge, 2018 (`https://hypo-university-challenge.eu-gb.mybluemix.net/`)
- School of Science TA Award, IUPUI, 2017

2005-Now **BUET, BD GOVT.** Researcher, Graduate Student

- 2nd Prize in International Project Show organized by BUET
- Dean's List in CSE, BUET; Top-10 Merit List CSE, BUET
- 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
Nanyang Technological University (NTU)
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