

Md Imrul Kayes, Ph.D.

CONTACT INFORMATION

2367 Nautical Way
Apt 111
Winter Park, FL, USA 32792

Voice: (813) 405-6320
E-mail: imrul@mail.usf.edu
WWW: <https://imrulkayes.github.io/>

RESEARCH INTERESTS

Mining large-scale (“big data”) online communities, such as social, community Q/A and blogging networks for limiting content abuse, creating user engagement and defining appropriate privacy settings; studying user behavior based on psychology and sociology theories.

EDUCATION

Ph.D. in Computer Science & Engineering December, 2015
Dissertation: Content abuse and privacy concerns in online social networks.
University of South Florida (USF), Tampa, Florida, USA

Master in Computer Science May, 2014
University of South Florida, Tampa, Florida, USA

Bachelor in Computer Science & Engineering September, 2009
Bangladesh University of Engineering and Technology, Dhaka, Bangladesh

EXPERIENCE

Sonobi Inc, Winter Park, Florida March, 2016 - Present
Data and Infrastructure Engineer

- Built predictionViz (Python, MySQL, R, Shiny), an interactive forecast evaluator which analyzes longitudinal prediction results on web publishers’ viewable ad spots.
- Created and deployed high availability HDFS, Hadoop, and Spark clusters using Ansible, Amazon AWS instances.
- Contributed to jetstream agency suite product: 1) real-time inventory forecasting using Python, Pandas and Statsmodels 2) devised and implemented algorithms for targeting audience based on users’ browsing behavior and spatial distribution of browsing 3) implemented metrics that enable online advertisers to target audiences from high quality publishers.
- Developed ETL processes to consolidate data from Amazon Redshift to the target (MySQL) database.
- Built ETL data pipeline for forecasting using Amazon S3, HDFS, Spark, Druid and Airflow.

Yahoo Research Labs, Barcelona, Spain September, 2013 - December, 2013
Research Intern

- Built a data pipeline to transform raw data of Yahoo Answers for analysis using Hadoop, Pig Latin, Python, and R.
- Characterized content abusers in the Yahoo Answers dataset (more than one year recorded activity logs of 1.5 million users), created exploratory graphs, and conducted statistical tests.
- Built machine learning models and detected content abusers in Yahoo Answers with 83% accuracy.
- Collaborated on writing papers using ShareLatex, Git and published the paper in the World Wide Web’15 conference.

Distributed System Labs, Tampa, Florida, USA May, 2012 - December, 2015
Research Assistant

- Designed and Implemented Java and Jena based “Aegis” privacy framework for social ecosystems.
- Analyzed user retention in social blogs and predicted top retained users with 93.62% accuracy using machine learning.
- Proposed a network centrality aggregation method to identify influential bloggers and detected influential bloggers.

- Analyzed privacy concerns and user behavior in Q/A sites and built machine learning models to predict privacy.
- Showed that cross-cultural variations exist in community Q/A sites analyzing user activity logs.
- Published research works in ASONAM'15, Hypertext'15, SocialCom'14, ICC'13, PST'13, and SocInfo'12 conferences.

SoftwarePeople, Copenhagen, Denmark

Software Engineer

March, 2010 - July, 2011

- Worked as a Scrum team member for the development of Agile-based software.
- used C#, ASP.NET, MS SQL server to build web-based business applications and service oriented systems.
- Researched on software quality and testing, wrote research papers, awarded a grant.
- Published research in the ICECT'11 conference; awarded a travel grant to present the paper.

Delta Life Insurance Co. Limited, Dhaka, Bangladesh

Software Developer

September, 2009 - December, 2009

- Developed “Xfer” (using C#, Oracle) to synchronizes data between remote and central servers.
- Worked on the development of the official website of Delta Life Insurance Co. Limited.

Binary Solutions, Dhaka, Bangladesh

Software Developer

November, 2008 - August, 2009

- Developed “EasyShare” using C# and MySQL, which collects data from the Dhaka Stock Exchange server and shows comparative market situation.

TEACHING

University of South Florida, Tampa, Florida, USA

Teaching Assistant

Fall 2011 and Spring 2012

- Teaching assistant for Discrete Structures, Object oriented Design, and Program Design courses.
- Shared responsibility for exams, homework assignments, and grades.

University of Information Technology and Sciences (UITS), Dhaka, Bangladesh

Instructor in the Computer Science Department

Spring 2009

- Taught full time courses (C Programming, Algorithms, and System Analysis and Design).
- Assessed students’ performance with quizzes, exams, and programming assignments.

RESEARCH PROJECTS

Content Abusers in Community Question Answering [1, 15, 16]

Used crowdsourced abuse reports to characterize and detect content abusers in Yahoo Answers.

- Used Hadoop, Pig, Python, and R to analyze one year recorded activity logs of 1.5 million Yahoo Answers users.
- Showed that users are good at flagging content and flagged content are removed quickly.
- Showed that moderate deviance is not necessarily bad, but extreme deviant users are likely to be suspended.
- Built machine learning models that were able to detect content abusers with 83% accuracy.

Retention in Community Blogs [2, 6, 7]

Studied the factors that are associated with users’ continued participation.

- Wrote a multi-threaded and error-resistant crawler in Python to scrape an Ajax-based blogging platform, “Blogster”.
- Collected blogger profiles (contributed 91% blogs) and stored in the MongoDB database.
- Analyzed retention using NetwrokX and R: predicted top retained users with 93.62% accuracy.

Cultures in Community Question Answering [3, 16]

Studied the influence of national culture on people’s online questioning and answering behavior.

- Analyzed a sample of 200K users in Yahoo Answers from 67 countries using a number of cultural factors extracted from Geert Hofstede’s cultural dimensions and Robert Levine’s Pace of Life.

- Showed cross-cultural variations in the predictability of activities, contribution behavior, privacy, and power distance.

Privacy Concerns vs. User Behavior in Community Question Answering [4, 16]

Analyzed the association between users' privacy concerns and their activity in CQA platforms.

- Showed that privacy preference is correlated with behavior in the community in terms of engagement, retention, accomplishment, and deviance from the norm.

Aegis: Privacy as Contextual Integrity in Social Ecosystems [8, 9, 16]

Modelled user privacy as contextual integrity using semantic web tools.

- Proposed an ontology-based data model to capture users social data from diverse sources.
- Used semantic-web tools to generate default privacy policies based on contextual integrity.
- Designed an architecture and implemented the privacy model using Java and Jena.

Influential Bloggers in the Blogging Community [10]

Studied the problem of identifying influential bloggers in the BlogCatalog blogging community.

- Proposed a method that aggregates different network position measurements into an overall influence score that can be used to identify influential bloggers.
- Showed that bloggers have a core-periphery network structure, with the highly influential bloggers well connected with each others forming the core, and the non-influentials at the periphery.
- Showed that network centrality metrics are highly correlated.

Prioritizing Test Cases for Regression Testing [5, 11]

Prioritized test cases for regression testing based on the dependency network of faults.

- Modeled software fault dependencies as a directed graph and identified leading faults.
- Developed a subject software “Tarantula” (19,390 lines of code) using C#, Python and compared the effectiveness of the proposed approach with traditional techniques.

Product Backlog Rating (PBR): a Metric for Measuring Testing Quality in Scrum [13]

Presented a metric that measures the quality of the testing process in a Scrum process.

- Proposed a metric Product Backlog Rating (PBR) to assess the testing process in Scrum.
- PBR considered the complexity of the features to be developed in an iteration of Scrum, assessed test ratings and offered a numerical score of the testing process.

PUBLICATIONS

Books

[1] Online Social Networks: Privacy, Participation and Abuse, ISBN 978-3-659-97793-0, LAMBERT Academic Publishing, Saarbrucken, Germany 2016.

Journal/Conference (refereed)

[1] **Imrul Kayes**, Nicolas Kourtellis, Daniele Quercia, Adriana Iamnitchi, Francesco Bonchi, The Social World of Content Abusers in Community Question Answering. *ACM 24th International World Wide Web conference (WWW'15), Florence, Italy, 2015.* (Acceptance rate: 14%)

[2] **Imrul Kayes**, Jacob Chakareski, Retention in Online Blogging: A Case Study of the Blogger Community. *IEEE Transactions on Computational Social Systems*, Volume 2, Issue 1, 2015.

[3] **Imrul Kayes**, Nicolas Kourtellis, Daniele Quercia, Adriana Iamnitchi, Francesco Bonchi, Cultures in Community Question Answering. *ACM 26th Conference on Hypertext and Social Media (HT'15), Cyprus, 2015.*

[4] **Imrul Kayes**, Nicolas Kourtellis, Francesco Bonchi, Adriana Iamnitchi, Privacy Concerns vs. User Behavior in Community Question Answering. *IEEE/ACM Conference on Advances in Social*

Networks Analysis and Mining (ASONAM'15), Paris, France, 2015. (Acceptance rate: 18%)

[5] **Imrul Kayes**, Shafinaz Islam, Jacob Chakareski, The Network of Faults: A Complex Network Approach to Prioritize Test Cases for Regression Testing, *Springer Journal of Innovations in Systems and Software Engineering*. Volume 11, Issue 4, 2015.

[6] **Imrul Kayes**, Xiang Zuo, Da Wang, Jacob Chakareski, To Blog or Not to Blog: Characterizing and Predicting Retention in Community Blogs, *ACM/ASE 7th Conference on Social Computing (SocialCom'14), Beijing, China, 2014.* (Acceptance rate: 16%)

[7] **Imrul Kayes**, Xiang Zuo, Da Wang, Jacob Chakareski, Did You Blog Yesterday? Retention in Community Blogs (short paper), *ACM/ASE 3rd International Conference on Big Data Science (BigDataScience'14), Beijing, China, 2014.*

[8] **Imrul Kayes**, Adriana Iamnitchi, Aegis: A Semantic Implementation of Privacy as Contextual Integrity in Social Ecosystems, *IEEE 11th Conference on Privacy, Security and Trust (PST), Catalonia, Spain, 2013.* (Acceptance rate: 28%)

[9] **Imrul Kayes**, Adriana Iamnitchi, Out of the Wild: On Generating Default Policies in Social Ecosystems, *IEEE International Conference on Communications (ICC'13)-BSN, Budapest, Hungary, 2013.* (Acceptance rate: 35%)

[10] **Imrul Kayes**, Xiaoning Qian, John Skvoretz, Adriana Iamnitchi, How Influential are You: Detecting Influential Bloggers in a Blogging Community, *4th International Conference on Social Informatics, Lausanne, Switzerland, 2012.* (Acceptance rate: 25%)

[11] **Imrul Kayes**, Test Case Prioritization for Regression Testing Based on Fault Dependency, *Proc. of IEEE 3rd International Conference on Electronics Computer Technology (ICECT), Kanyakumari, India, 2011.*

[12] **Imrul Kayes**, Agile Testing: Introducing PRAT as a Metric of Testing Quality in Scrum, *ACM SIGSOFT Software Engineering Notes, Volume 36, Number 2, 2011.*

[13] **Imrul Kayes**, Mithun Sarker, Jacob Chakareski, Product Backlog Rating: A Case Study On Measuring Test Quality In Scrum, *Springer Journal of Innovations in Systems and Software Engineering*, December 2016, Volume 12, Issue 4.

Under review

[14] **Imrul Kayes**, Adriana Iamnitchi, Privacy and Security in Online Social Networks: A Survey. arXiv:1504.03342. (*Submitted*).

[15] **Imrul Kayes**, Nicolas Kourtellis, Daniele Quercia, Adriana Iamnitchi, Francesco Bonchi, The Good, the Bad and the Deviant in Community Question Answering. (*Submitted*)

Theses

[16] **Imrul Kayes**, Content abuse and privacy concerns in online social networks. PhD thesis, Department of Computer Science, University of South Florida, Tampa, FL, USA, 2015.

TECHNICAL
SUMMARY

- Languages & Frameworks: Python, C, C++, Java, C#, Ansible, ASP.NET, R, HTML, JavaScript, CSS, Bootstrap
- Machine Learning & Data Mining: Social, Community Q/A and Blogging network analysis and mining, Time-series prediction, Regression, Classification, Weka, R, Sklearn, Statsmodels

- Database: MongoDB, Oracle, MySQL, PostgreSQL
- Distributed Data Processing: Hadoop, Pig Latin, Spark, Druid
- Data Analysis and Visualization: R, Shiny, NetworkX, JUNG, SNAP, Gephi
- Deployment and Configuration Management: Ansible
- Applications: OmniGraffle, LATEX, Gnuplot, Git

HONORS AND AWARDS

- ACM travel grant to the 26th conference on Hypertext and Social Media, Cyprus, 2015.
- San Diego Supercomputer center scholarship to the Big Data summer school, San Diego, 2014.
- Google travel grant to the 4th SocInfo conference, Switzerland, 2012.
- SoftwarePeople travel grant to the IEEE ICECT conference, India, 2011.
- Dean's scholarship, Bangladesh Uni. of Engg. and Tech., 2009.
- CSE academic merit scholarship, Bangladesh Uni. of Engg. and Tech., 2006, 2007.

PROFESSIONAL ACTIVITIES

- Reviewer
 - Elsevier Computer Communications.
 - IEEE Transactions on Knowledge and Data Engineering (TKDE).
 - ACM Conference on Computer-Supported Cooperative Work and Social Computing, 2014.
- External reviewer
 - IEEE Transactions on Parallel and Distributed Systems.
 - ACM Cloud and Autonomic Computing Conference, 2013.
 - International Conference on Distributed Computing Systems (ICDCS), 2014
 - ACM Symposium on High-Performance Parallel and Distributed Computing, 2015.
- Summer schools: San Diego Supercomputer Center (SDSC): HPC Meets Big Data, 2014.

VOLUNTEER WORK

- Engineering EXPO, USF, 2013
- Engineering EXPO, USF, 2014

TALKS

- *Mitigating Content Abuse and Privacy Risk in Online Social Networks*
 - Department of Computer Science and Information Technology, Georgia Southern University, February, 2016.
 - Department of Computer Science, Texas A&M University at Corpus Christi, February, 2016.
- *The Social World of Content Abusers in Community Question Answering*
 - ACM 24th International World Wide Web conference (WWW), Florence, Italy, May, 2015.
 - Twitter, June, 2015
- *Aegis: A Semantic Implementation of Privacy as Contextual Integrity in Social Ecosystems*
 - IEEE 11'th Conference on Privacy, Security and Trust (PST), Catalonia, Spain, July 2013.
- *How Influential are You: Detecting Influential Bloggers in a Blogging Community*
 - 4th International Conference on Social Informatics, Lausanne, Switzerland, December, 2012.
- *Test Case Prioritization for Regression Testing Based on Fault Dependency*
 - IEEE 3rd Conf. on Electronics & Computer Technology, Kanyakumari, India, April, 2011.

REFERENCES

- Adriana Iamnitchi, Ph.D.
Associate Professor
Department of Computer Science and Engineering
University of South Florida
Tampa, FL, USA, 33620
anda@cse.usf.edu
+1 (813) 974-5357
- Nicolas Kourtellis, Ph.D.
Associate Researcher
Telefonica Research and Development
Barcelona, Spain, 08019

nicolas.kourtellis@telefonica.com
+34 913 12 87 00

- Jacob Chakareski, Ph.D.
Assistant Professor
Department of Electrical and Computer Engineering
Tuscaloosa, AL, USA, 35401
jacob@ua.edu
+1 (205) 348-2623