Syed Arefinul Haque

syedarehaq@gmail.com
linkedin.com/in/syedarehaq Google Scholar

Summary

Data scientist with a background in network science, machine learning, and natural language processing. In today's connected world, macro patterns of critical infrastructures such as the Internet, power-grid even our society can be modeled using networks. My training in network science provides me with the ability to extract, analyze and visualize vital nodes, hidden communities, vulnerabilities, and diffusion dynamics in large scale networks.

Technical Strengths

Analytical Skills: Network Analysis, Network Visualization, Natural Language Processing, Agent Based Modelling, Simulation, Epidemiology, Clustering Techniques, Embedding Methods, Regression Analysis

Computing Skills:
Programming: Python, R, JavaScript, Bash
Database: SQL, BigQuery, MongoDB, neo4j
Visualization: Matplotlib, Seaborn, D3, Cytoscape, Gephi, GnuPlot, Adobe Illustrator
Network Analysis: NetworkX, Graph-tool, iGraph
Other: Unix, Git, SLURM, Google Cloud, LaTeX

Education

Northeastern University — Boston, MA, USA
Ph.D. in Network Science — 2020

United International University — Dhaka, Bangladesh
M.Sc. in Computer Science and Engineering — 2015
Thesis : "Virtual P2P client: accessing P2P applications using virtual terminals"

Institute of Business Administration, University of Dhaka — Dhaka, Bangladesh
B.B.A. in Finance (Minor in Marketing) — 2013

Experience

Research Assistant, Northeastern University — 2015-Present

Mapping organizational change for gender equity using NSF grant data

- Visualized co-occurrence of gender equity related keywords found the grant data. NLP, Network Visualization
- Managed large scale bibliographic databse using Google Cloud Platform Google-BigQuery

Flocking behavior in science

- Implemented embedding methods to find the similarity between scientific fields to see how different fields cross-pollinate with each other over time. NLP Word2Vec; UMAP; HPC
- Extracted keywords from the abstract and of 120 million papers collected from Microsoft Academic Graph. NLP tf-idf; Google-BigQuery

Reconstructing pathways of Zika virus epidemic in Americas

- Collected genomic inferences and surveillance data on the Zika virus and applied statistical techniques to compare them with model generated data to learn how the disease spread throughout Americas. Epidemiology, Simulation Google-BigQuery; Python Cross Correlation, Linear Regression
- Developed a web based interactive visualization which illustrates the simulated imported Zika cases in more than 3000 urban areas throughout the world. D3; MongoDB; ExpressJS

Research Assistant, United International University — 2013-2015

- Developed a virtualized Peer-to-Peer (P2P) client of BitTorrent protocol that was deployed in web server to measure how well P2P clients perform in the edge clouds like AWS and Linode. Amazon AWS; WebRTC; ExpressJS

Business Development Executive, Mukto Software Limited 2013 - 2015

– Served as a liaison between the corporate customer and the software development team by outlining requirements of enterprise resource planning (ERP) software projects. Project Management; Kanban

Self Directed Projects

Gender stereotype in media 2019-Present

Collected all the English movie subtitles from OpenSubtitles. Exploring if there is an evolution of gender stereotypes in the media from the word embedding created from this dataset. NLP; Word Embedding;

Bias in newspaper portrayal 2018

Crawled and curated newspaper data from six Bangladeshi newspapers and used named entity recognition (NER) tools to find actors in those news articles. Through this analysis we were able to show the bias towards political actors in news reporting. NLP; NER

Selected Course Projects

Bayesian and Network Statistics 2017

Analyzed dataset from Second Life online platform to explore the factors behind formation of trust relationship between users. Social Network Analysis; ERGM

Dynamical Processes on Complex Networks 2016

Developed a predictive epidemiological model of 2016 US presidential election using social media and poll data. Mechanistic Modeling; SIR Process

Network Science Data 2016

Crawled GitHub social network and analyzed how the reputation of a user increases based on their collaboration in diverse projects. Network Analysis; Web Crawling

Publications

- Chowdhury, S. S., Saquib, N., Zawad, N., Mandal, M.K. & **Haque, S. A.** (2018), Statement networks: a power structure narrative as depicted by newspapers. *Proceedings of NeurIPS 2018 workshop on Machine Learning for the Developing World*
- Hassan, M. K., Islam, L. & **Haque, S. A.** (2017), Degree distribution, rank-size distribution, and leadership persistence in mediation-driven attachment networks. *Physica A: Statistical Mechanics and its Applications*, 469, 23-30
- **Haque, S. A.**, Islam, S., Islam, M. J., & Grégoire, J. C. (2016). An architecture for client virtualization: A case study. *Computer Networks*, 100, 75-89.

Working Papers

- **Haque, S. A.**, Sun, K., Zhang, Q.,Pastore y Piontti, A., Chinazzi, M., Mistry, D., Dean, N. E., Rojas, D. P., Merler, S., Poletti, P., Rossi, L., Halloran, M. E., Longini Jr., I. M. & Vespignani, A. Uncovering the hidden paths of Zika in the Americas. (In preparation)
- Mistry, D., Litvinova, M., Pastore y Piontti, A., Mu, K., Xiong, X., **Haque, S. A.**, Quan-Hui, L., Gomes, M. F. C., Fumanelli, L., Longini Jr., I. M., Halloran, M. E., Merler, S., Ajelli, M. & Vespignani, A. Inferring high-resolution disease-specific human mixing patterns. (Under review)

Advanced trainings and Certificates

- 9th Summer Institute in Statistics and Modeling in Infectious Diseases (SISMID), University of Washington 2017
- Complex System Summer School (CSSS 2016), Santa Fe Institute, New Mexico 2016

Awards

- Travel scholarship for SISMID 2017, University of Washington 2017
- NULab travel grant, Northeastern University 2016