

Madhurima Nath

madhurima.nath@icloud.com || mnath@vt.edu || +1-540-235-2449

[Google Scholar](#) || [GitHub](#) || [LinkedIn](#)

Data scientist with PhD in computational physics, specializing in machine learning, NLP, and data engineering. Experience delivering end-to-end solutions across consulting, spanning data pipeline development, predictive analytics, and GenAI applications. Academic background in complex systems analysis, network science, and statistical physics.

POSITIONS

Senior Data Scientist , Slalom, Inc., New York, New York, USA	Sep 2024 – Present
Data Scientist , Slalom, Inc., New York, New York, USA	Jul 2021 – Aug 2024
Associate Data Scientist , Slalom, Inc., New York, New York, USA	Jan 2020 – Jun 2021
Post-doctoral Research Assistant , Virginia Tech, Virginia, USA	Feb 2019 – Dec 2019
Instructor, Department of Physics , Virginia Tech, Virginia, USA	Aug 2018 – Dec 2018
Graduate Research Assistant , Virginia Tech, Virginia, USA	May 2014 – Dec 2018

EDUCATION

Ph.D., Physics (Computational) , Virginia Tech, Virginia, USA	Dec 2018
<ul style="list-style-type: none"><i>Research Focus</i>: Complex systems analysis, Statistical physics, Network science, Community detection<i>Dissertation</i>: Application of Network Reliability to Analyse Diffusive Processes on Graph Dynamical Systems<i>US Patent</i> (US20210286859A1): System, method and computer readable medium for sensitivity of dynamical systems to interaction network topology	
M.S., Physics , Virginia Tech, Virginia, USA	May 2017
M.Sc., Physics , Indian Institute of Technology Delhi, India	May 2012
<ul style="list-style-type: none"><i>Thesis</i>: Study of Cold Atomic Condensates by Atomic Photon Interactions<i>Award</i>: Best Master of Science Thesis 2011-2012	
B.Sc. (Hons.) , Physics, University of Calcutta, India	May 2010
<ul style="list-style-type: none"><i>Minors</i>: Mathematics and Chemistry	

DATA ENGINEERING PROJECTS

Data Lake Implementation on Snowflake (12 months) || Nonprofit Organisation

- Maintained and upgraded Snowflake data lake processing 2M daily JSON bundles of statewide patient health information from 6 regional health information exchange networks.
- Built Snowflake task orchestrations and parameterised stored procedures for batch ingestion and export pipelines, ensuring compliance with data protection regulations for sensitive patient information.
- Refactored legacy monolithic JavaScript into modular Python and SQL components, improving maintainability and code quality.

Network Predictive Maintenance Solution (4 weeks) || Telecom Client

- Developed statistical analysis combining 6 RF performance metrics with geographical clustering to proactively identify faulty network infrastructure from user device data.
- Implemented automated stored procedures on Google BigQuery processing 90-day network performance data across New York metropolitan area, collaborating with network engineers to validate preventive maintenance insights.
- Projected ~US \$8M annual operational cost savings through reduction of approximately 30K customer support calls and 6K service visits, with methodology validated for nationwide scaling.

Data Migration to Salesforce (5 months) || Nonprofit Organisation

- Migrated 1TB of historical data to Salesforce through custom SQL stored procedures, maintaining foreign key relationships across interdependent datasets ranging from hundreds of thousands to billions of rows.
- Configured parameterised Azure Data Factory pipelines supporting full historical loads and incremental delta updates with table partitioning for large-scale dataset performance.
- Built validation framework including row count verification, source-to-target comparison, and duplicate detection to ensure data integrity across migrations.

GENERATIVE AI & LARGE LANGUAGE MODEL (LLM) PROJECTS

Product Catalogue Discovery Assistant (2 months) || US Fortune 500 Financial Services Client || Lead Data Scientist

- Designed and developed a RAG system on Databricks integrating Llama with LangChain to enable internal advisors to search and retrieve product information from internal catalogues.
- Configured MLflow for experiment tracking and model versioning, with Unity Catalog managing the vector database for product embeddings and serving as the ML model registry.
- Coordinated with compliance teams to establish data governance controls and regulatory guardrails, demonstrating improved information retrieval over traditional keyword search.

Enterprise Q&A Assistant (6 weeks) || US Fortune 500 Equipment Rental Firm || Lead Data Scientist

- Built a RAG chatbot using OpenAI GPT and LangChain to streamline sales team access to equipment manuals, pricing guides, and internal policies.
- Collaborated with business stakeholders to align solution with operational requirements and validate response accuracy through iterative testing.
- Delivered proof of concept demonstrating efficient handling of customer inquiries and reduced manual effort for complex queries.

Clinical Data Support Q&A Chatbot (6 weeks) || Healthcare Technology Startup || Data Scientist

- Developed a Q&A assistant using Anthropic Claude on AWS Bedrock to translate clustering algorithm outputs into natural language summaries for healthcare providers and researchers.
- Applied prompt engineering to combine clustering results with structured patient data including symptoms, health conditions, and drug classifications from JSON files.
- Partnered with design teams to enhance interface accessibility, delivering a proof of concept for extracting insights from complex bio-pharmaceutical datasets.

DATA SCIENCE PROJECTS

NLP for Safety Analytics (5 months) || US Fortune 500 Energy & Utility Client || Lead Data Scientist

- Developed and deployed end-to-end NLP pipeline on Azure Databricks to process five years of safety incident reports from multiple sources in a unified data lake.
- Implemented topic modelling and sentiment analysis to extract safety themes and trends from unstructured incident narratives across electricity and gas operations.
- Deployed production solution with automated weekly reports and interactive dashboards enabling executives to filter insights by business unit and monitor emerging themes.
- Solution received executive recognition and wide organisational adoption for providing actionable intelligence on safety policy and resource allocation.

NLP for e-Discovery Document Classification (6 weeks) || Legal Services Client || Lead data Scientist

- Developed proof of concept document classification system using Word2Vec embeddings and SVM classifiers on AWS SageMaker to automate legal document review processes.
- Designed retraining pipeline framework to enable incremental model updates from newly labelled documents.
- Demonstrated potential to reduce manual review hours and costs compared to traditional document review workflows.

NLP for Global Procurement Standardisation (8 months) || US Fortune 500 Consumer Packaged Goods Company

- Built text classification pipeline using Random Forest on Azure Machine Learning to automatically classify products into standardised global category templates across LATAM and North American procurement catalogues.
- Collaborated with engineering teams to incorporate MLOps best practices including automated pipeline updates and model retraining workflows for production deployment.
- Accelerated client's digital transformation roadmap by approximately one year by removing critical data standardisation bottleneck and improving data consistency across global operations.

Post-doctoral Research Assistant || Virginia Tech, USA || Supervisor: Prof. Stephen Eubank Feb 2019 – Dec 2019

- Developed hybrid computational approach combining Monte Carlo simulations with weak- and strong-coupling perturbative expansions to improve estimation of Moore-Shannon network reliability on graphs, addressing an NP-hard computational problem. [??]
- Extended weak- and strong-coupling perturbative methods from statistical physics to heterogeneous satisfiability problems, developing novel approach for constructing tight upper and lower bounds on approximation error for mostly monotonic probabilistic satisfiability problems. [??]
- Implemented statistical approaches for community detection in large-scale weighted directed networks, achieving significant improvements over traditional methods. [??]
- Applied computational methods to analyse international trade network dynamics from the United Nations (UN) Comtrade database, to identify crucial communities for preventing global pestilence distribution. [??]

Graduate Research Assistant || Virginia Tech, USA || Supervisor: Prof. Stephen Eubank May 2014 – Dec 2018

- Applied Moore-Shannon network reliability formalism (traditionally used for electronic circuit analysis) to biological and social network systems, establishing theoretical framework connecting electrical engineering concepts to network epidemiology and complex systems analysis. [GitHub repo](#) [??, ??]
- Developed computational approaches to evaluate Moore-Shannon network reliability formalism using Bernstein basis functions as polynomial basis set, providing convergence guarantees for sequential design and model selection in discrete finite systems. [GitHub repo](#) [??]
- Applied Moore-Shannon network reliability to predict final global states of graph dynamical systems, analysing how interactions between individual node states and their connections determine final outcomes in practical applications.[??, ??]
- Showed that synthetic network models generated using Exponential Random Graph Models cannot reliably predict epidemic outcomes on empirical contact networks. Analysis of National Longitudinal Study of Adolescent to Adult Health (Add Health) dataset demonstrated that models matching local network statistics can overestimate infection numbers by approximately 50%. Applied Birnbaum importance measures to quantify individual edge contributions to epidemic potential, enabling identification of critical connections for targeted. [??]
- Identified vulnerabilities within global food trade networks using commodity-specific data from UN Comtrade database, analysing where pest and pathogen contamination could cascade through global supply chains. Achieved approximately 96% precision in forecasting impact of mitigation strategies under various contagion scenarios. [??]
- Reduced time complexity from $O(n^2)$ to $O(n)$ for estimating energy states in interacting magnetic systems (Ising model) by reformulating the problem through network reliability framework. Established equivalence between network reliability and Ising partition function, enabling transfer of statistical physics computational methods to network science applications. [??]
- Developed parallel Markov-chain Monte Carlo scheme for estimating joint density of states in Ising model applications, addressing fundamental limitations of naive sampling approaches that yield poor approximations for partition function estimation. [??]

Master's Thesis || Indian Institute of Technology Delhi, India || Supervisor: Prof. Sankalpa Ghosh Jul 2011 – May 2012

Study of Cold Atomic Condensates by Atom-Photon Interactions [??] || Best Master of Science Thesis 2011-2012

- Developed computational framework in MATLAB for probing quantum many-body states of ultracold atoms in optical lattices using angle-resolved light transmission measurements, demonstrating that confined atoms act as quantum diffraction gratings.
- Calculated dispersive shifts in cavity resonance caused by atomic presence, establishing direct proportionality between resonance shift and atom count in illuminated sites for systematic quantum state characterisation.
- Generated transmission spectra across Mott Insulator and Superfluid phases by systematically varying cavity-lattice geometry, atom number, and illuminated site distribution parameters.
- Established non-destructive measurement approach enabling experimental exploration of Fock-space structure in few-body correlated quantum systems through optical cavity transmission analysis.
- Visualisation selected for American Physical Society (APS) Physical Review A [Kaleidoscope](#) (June 2012) based on aesthetic quality of graphics.

TEACHING EXPERIENCE

Instructor, Department of Physics, Virginia Tech, USA

Fall 2018

- Designed and delivered three undergraduate physics modules: Optics, Thermal Physics, Waves & Sound, to 90 students across all modules.
- Developed all teaching materials including lectures, problem sets, examinations, and assessment criteria.
- Managed course administration including student records and academic standards compliance.
- Achieved above-average student feedback scores for creating respectful learning environments and fair assessment practices.

Graduate Teaching Assistant, Department of Physics, Virginia Tech, USA

2013-2015

- Conducted weekly tutorial sessions for 10–15 undergraduate students in mechanics, electromagnetism, and optics, developing supplementary instructional materials and providing regular office hours support.
- Supervised four introductory physics laboratory courses in kinematics and electromagnetism, adapting teaching approaches for both engineering and non-engineering students.
- Received positive feedback for fostering inclusive learning environments and delivering effective student support.

SUMMER RESEARCH PROJECTS

Quantum Information Conservation: The No-Hiding Theorem || Completion letter

Jun 2012 – Jul 2012

Harish-Chandra Research Institute, India || *Supervisor:* Prof. Arun Kumar Pati

- Selected for prestigious Visiting Students Programme (VSP) in Physics at premier research institute funded by Department of Atomic Energy, Government of India, to conduct research in quantum information theory.
- Studied fundamental conservation laws governing information in quantum systems through theoretical analysis and numerical verification of the No-Hiding Theorem, establishing that quantum information cannot be created or destroyed but only redistributed between system and environment.
- Analysed perfect hiding scenarios where quantum information transfers completely to ancilla (environment) states without entanglement, demonstrating information conservation in the absence of quantum correlations.
- Quantified how information about input states distributes between system and environment in imperfect hiding cases using correlation measures implemented in MATLAB.

Renormalisation Group Study of Liénard Systems || Completion letter

May 2011 – Jul 2011

Jadavpur University, India || *Supervisor:* Dr. Dhruba Banerjee

- Applied renormalisation group methods to analyse limit cycle behaviour in nonlinear dynamical systems, working within the framework of Hilbert's sixteenth problem on limit cycles in polynomial differential equations.
- Derived amplitude equations for generalised Liénard systems using perturbative renormalisation group techniques, systematically incorporating higher-order damping effects through recursive expansions.
- Determined limit cycle existence and stability by analysing fixed points and eigenvalues of the amplitude equations.
- Extended the perturbative analysis to second order, examining the structure and convergence properties of the analytical approach for higher-order approximations.
- Validated theoretical predictions through computational analysis in Mathematica, generating phase portraits and limit cycle trajectories across parameter space.

TECHNICAL EXPERTISE

Computational & Data Processing

- Programming: Python, SQL, PySpark, R, Bash
- Distributed Computing: PySpark, Databricks, Snowflake || Cloud Platforms: AWS, Azure, GCP
- Version Control & CI/CD: GitLab, GitHub, Azure DevOps, BitBucket, Terraform (familiar)

Machine Learning Analytics

- ML Frameworks: MLflow, LangChain, TensorFlow (familiar) || Data Tools: Delta Lake, Delta Lakehouse
- GenAI Development: Prompt engineering, multi-agent frameworks, RAG (retrieval-augmented generation)

Project Management & Collaboration

- Agile/Scrum (JIRA), cross-functional team leadership, stakeholder communication
- Proposal development and client success story documentation for internal knowledge sharing

PUBLICATIONS

1. Mishra, R., Eubank, S., **Nath, M.**, Amundsen, M., and Adiga, A. (2022, November). [Community Detection Using Moore-Shannon Network Reliability: Application to Food Networks](#), In International Conference on Complex Networks and Their Applications (pp. 271-282). Cham: Springer International Publishing.
2. Eubank, S., **Nath, M.**, Ren, Y., and Adiga, A. (2022). [Perturbative methods for mostly monotonic probabilistic satisfiability problems](#). arXiv preprint arXiv:2206.03550.
3. **Nath, M.**, et. al. (2019). [Using network reliability to understand international food trade dynamics](#). In Complex Networks and Their Applications VII: Volume 1 Proceedings The 7th International Conference on Complex Networks and Their Applications COMPLEX NETWORKS 2018 7 (pp. 524-535). Springer International Publishing.
4. **Nath, M.**, Ren Y., and Eubank, S. (2019). [An approach to structural analysis using Moore-Shannon network reliability](#). In Complex Networks and Their Applications VII: Volume 1 Proceedings The 7th International Conference on Complex Networks and Their Applications COMPLEX NETWORKS 2018 7 (pp. 537-549). Springer International Publishing.
5. **Nath, M.**, Ren, Y., Khorramzadeh, Y., and Eubank, S. (2018). [Determining whether a class of random graphs is consistent with an observed contact network](#). Journal of theoretical biology. 440, 121-132.
6. **Nath, M.**, and Eubank, S. (2018). [Model selection for sequential designs in discrete finite systems using Bernstein kernels](#). arXiv preprint arXiv:1807.06661.
7. Ren, Y., Eubank, S., and **Nath, M.** (2016). [From network reliability to the Ising model: A parallel scheme for estimating the joint density of states](#). Physical Review E, 94(4), 042125.
8. **M. Nath**, et. al. (2015). [A two-parameter method to characterise the network reliability for diffusive processes](#). In Complex Networks VI: Proceedings of the 6th Workshop on Complex Networks CompleNet 2015 (pp. 139-148). Springer International Publishing.
9. Agarwala, A., **Nath, M.**, Lugani, J., Thyagarajan, K., and Ghosh, G. (2012). [Fock-space exploration by angle resolved transmission through a quantum diffraction grating of cold atoms in an optical lattice](#). Physical Review A, 85(6), 063606.

CONFERENCE PRESENTATIONS

- Network reliability: a generic tool to explore diffusive processes on interacting systems, NASA Prebiotic Chemistry and Early Earth Environments (PCE3) Virtual Workshop 2022 Nano- to Cosmic- Studies of Complex Systems, University of Wisconsin Madison, Madison, WI, USA, 20 Oct, 2022.
- Perturbative methods for estimating relative contributions to network reliability, SIAM (Society for Industrial and Applied Mathematics) Workshop on Network Science, Virtual Workshop, 13-15 Sep, 2022.
- Statistical mechanical applications of graph dynamical systems, Condensed Matter Seminar, Department of Physics, Virginia Tech, Blacksburg, VA, USA, 30 Oct, 2017.
- Determining whether a particular contact network is consistent with a network model, 1st North American Social Networks Conference of the International Network for Social Network Analysis, Washington DC, USA, 26-30 Jul, 2017.
- Network reliability: A novel measure to study the effects of network topology on the diffusive dynamics, Symposium for the Society of Young Network Scientists, NetSci 2017, Indianapolis, IN, USA, 19-23 Jun, 2017.
- Network reliability: A measure to study diffusive dynamics on networks, Center for Soft Matter and Biological Physics Symposium 2017, Virginia Tech, Blacksburg, VA, USA, 17-18 May 2017.
- Effects of network structure on propagation of infectious diseases, 33rd Annual Graduate Student Assembly Symposium and Exposition, Virginia Tech, Blacksburg, VA, USA, 29 Mar 2017.
- Renormalisation group approaches for dynamics on irregular networks, American Physical Society (APS) March Meeting 2017, New Orleans, LA, USA, 13-17 Mar, 2017.
- Diffusive dynamics on a network, Southeastern Section of the American Physical Society (SESAPS) Conference 2016, Charlottesville, VA, USA, 9-12 Nov, 2016.
- Effects of network structure on epidemic modeling, Biocomplexity Institute Symposium 2016, Virginia Tech, Blacksburg, VA, USA, 1 Nov, 2016.
- A two-parameter method to characterise the network reliability for diffusive processes, CompleNet 2015, New York City, NY, USA, 25-27 Mar, 2015.
- Four-parameter characterisation of network reliability and analysis of critical point phenomenology, APS (American Physical Society) March Meeting 2015, San Antonio, TX, USA, 2-6 Mar, 2015.

PROFESSIONAL & COMMUNITY SERVICE

Peer Review

- NeurIPS: Machine Learning and Physical Sciences 2022, 2024
- ICML: Synergy of Scientific and Machine Learning Modeling 2023
- Physical Review E 2022–Present

Community Engagement

- Industry Ambassador, Women in Network Science Society 2022–2025
- Conference Organiser, Women in Network Science Networks 2021 2021

Invited Online Lectures & Panel Discussions

- Insights into Data Engineering (Parts 1 & 2), Women Who Code San Francisco [Part 1](#) || [Part 2](#) Jul 2021
- Exploring NLP Fuzzy Matching Algorithms, Women Who Code [YouTube](#), [Medium article](#) Apr 2021
- Statistics in Data Science Discussion, Women Who Code Statistics in Data Science Workshop Series [YouTube](#) Mar 2021
- Regression & Predictions, Women Who Code Data Science: Statistics Workshop Series - Statistics in Data Science [YouTube](#) [Medium article](#) Feb 2021

REFERENCES

- Prof. Stephen Eubank (Phd supervisor)
Professor, Department of Public Health Sciences, School of Medicine
Deputy Director, Network System Science and Advanced Computing Division, Biocomplexity Institute
University of Virginia, Charlottesville, Virginia, USA
Email: eubank@virginia.edu
Phone: +1-434-365-0000
- Prof. Madhav Marathe
Professor, Department of Computer Science
Executive Director, Network System Science and Advanced Computing Division, Biocomplexity Institute
University of Virginia, Charlottesville, Virginia, USA
Email: marathe@virginia.edu
Phone: +1-540-808-6220
- Maria Dela Vega-Mendoza
Senior Solutions Consultant
Databricks, New York, New York, USA
Email: maria.mendoza@databricks.com
Phone: +1-914-400-4368