

CHUAN (SOPHIE) DU

+1 (217) 417-8936 | chuandu2019@u.northwestern.edu | <https://chuandu2.github.io/>

EDUCATION

NORTHWESTERN UNIVERSITY

Master of Science in Analytics

EVANSTON, IL

EXPECTED: SEP 2018 - DEC 2019

UNIVERSITY OF ILLINOIS AT URBANA – CHAMPAIGN

Bachelor of Science in Applied Mathematics, Statistics; Minor in Informatics, Computational Science & Engineering

URBANA – CHAMPAIGN, IL

AUG 2014 – MAY 2018

- Graduated with High Distinctions, LAS Dean's List, Pi Mu Epsilon National Mathematics Honor Society GPA: 3.86
- Meritorious Winner of 2017 the Mathematical Contest in Modeling (MCM)

SKILLS

- **Data Science:** Data Analysis, Predictive Analytics, Mathematical Modeling, Data Management, Data Visualization, Statistical Learning, Probabilistic Programming, Analytics for Big Data
- **Programming & Software:** R, Python, SQL, SAS, JAVA, LaTeX, AWS Redshift, Tableau

EXPERIENCE

ILLINOIS GEOMETRY LABORATORY – DISCRETE MORSE THEORY & VECTOR FIELDS

Undergraduate Researcher

CHAMPAIGN, IL

JAN 2016 – FEB 2018

- **Publication:** *RGB image-based data analysis via discrete Morse theory and persistent homology*, (first author) with C. Szul, A. Manawa, N. Rasekh, R. Guzman, R. Davidson, [arXiv:1801.09530](https://arxiv.org/abs/1801.09530)
- Designed converters using Python to convert RGB images into grayscale, used Australian National University's source code to create discrete Morse functions (DMF) to extract key topological information from images, and constructed custom models of DMF vector fields in a cubical complex
- Generated data-informative persistence diagrams to describe the birth and death time of persistence pairs, enabling users to predict future image-based data behavior with the life-span information of topological features in the input images
- Applied the theory to perform analysis on open-source heat maps of water scarcity variability and crime rates variability data

FUNCTION CAPITAL – SILICON VALLEY OFFICE

Investment Analyst Summer Intern

REDWOOD CITY, CA

JUN 2017 – JUL 2017

- Consolidated and analyzed financial data of target companies with Excel to generate reports for investment decision-making
- Investigated initial financing of target projects, mapped competitive landscape of each competitor to analyze technical barriers and competitiveness, and attended industry-related events to grasp industry prospects
- Completed trends reports for four projects in the fields of machine learning, deep learning, AI-AGV and robotics

CHINESE ACADEMY OF SCIENCES – RANDOM WALK IN MICRORNA – DISEASE ASSOCIATIONS

Research Assistant

BEIJING, CHINA

AUG 2017

- Constructed mathematical models based on corresponding integrated similarity of diseases and miRNA including Gaussian interaction profile kernel similarity, semantic similarity and functional similarity. Developed random walk algorithm for microRNA-disease associations using Python, and mastered validation methods of LOOCV and k-fold cross validation

BP NORTH AMERICA – INDUSTRY PRACTICUM AT NORTHWESTERN UNIVERSITY

Data Science Student Consultant

EVANSTON, IL

NOV 2018 – PRESENT

- Developing comprehensive automated price evaluation tool to enable a more efficient and standardized approach to structuring pricing for commercial sales contracts. Integrating data science tools to find patterns in historical changes along with predictive analytics for future changes to enhance risk management of price volatility

COMPETITION & PROJECT

WAITING IN AIRPORT SECURITY CHECKPOINT: A PERSPECTIVE FROM QUEUEING THEORY – 2017 MCM/ICM

JAN 2017

- Led the group to conduct empirical analysis using R to establish a mathematical model for increasing checkpoint throughput and reducing variance in wait time, proposed and implemented a feasible method to estimate passenger volume at various times
- Developed two modifications to the current procedural model, with the first one incorporating a Bifurcation System and the second designing a Circular Line-up System

SURGERY PAIN LEVEL ANALYSIS – NORTHWESTERN UNIVERSITY

OCT 2018 – DEC 2018

- Analyzed the reported pain levels from patients across different race, gender and surgery type to identify pain patterns of different patient groups through k-means cluster analysis with 70% data explained. Examined factors that affected patient pain levels through significance tests. Predicted patient pain levels with regression models (KNN, Elastic Net, Random Forest)