Juntang Wang

Duke Kunshan University & Duke University

B.S. Applied Math & Computational Science; Computer Science Track

Class of 2026

jw853@duke.edu

github.com/qqgjyx

www.qqgjyx.com

Kunshan, CN • Durham, USA

Bridging AI, mathematics, and biomedical modeling through computational design and visualization.

> *"The impediment to action advances action. What stands in the way becomes the way."*
> — Marcus Aurelius, Meditations 5.20
>
> *Turning obstacles into methods across AI, math, and biomedicine.*

Website

About This Portfolio

This portfolio showcases my work at the intersection of artificial intelligence, applied mathematics, and computational science. It includes research contributions under peer review at top-tier conferences (ICLR, BICS), open-source software tools with 600+ GitHub stars, teaching experience, and computational modeling projects.

Each project demonstrates technical depth, practical impact, and creative problem-solving across machine learning, systems design, and computational modeling.

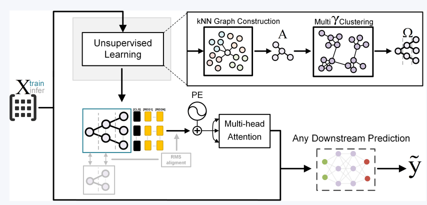
Contents

Research Projects

Mixing Configurations for Downstream Prediction

First Author, Research Assistant with Prof. Shixin Xu PyTorch, Unsupervised Learning

- Proposed GraMixC: a configuration – mixing module that fuses unsupervised "configurations" with task predictors for label – efficient downstream performance.
- Defined configuration extraction and mixing objectives, and integrated them with linear probing / 3LP predictors without changing the backbone.
- Validated on DSNI regression (pH/temperature) and image benchmarks; figures show attention maps, lineage diagrams, and quantitative gains.
- Ablation demonstrates incremental mixing (GMC) outperforms static concatenation (GC).
- Co-authored with Hao Wu, Runkun Guo, Yihan Wang, Dongmian Zou, and Prof. Shixin Xu; under ICLR 2026 review.



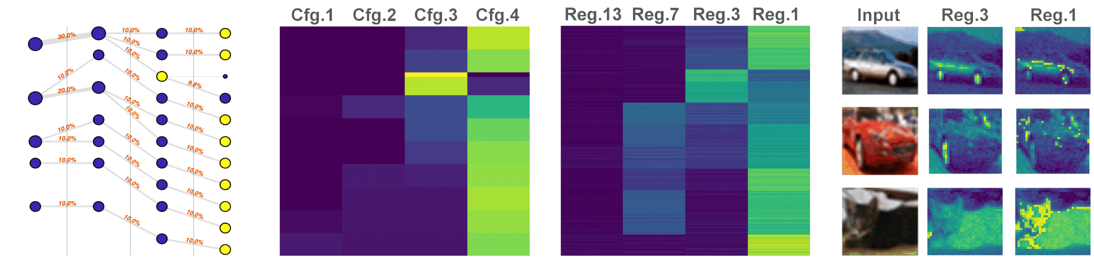
Links: arXiv preprint

Highlights: Unsupervised Learning ICLR under review

Technical Highlights: This project tackles the challenge of learning meaningful representations from unlabeled data by mixing configurations in a learned latent space. The key innovation is a training objective that encourages the model to predict properties of mixed configurations based on the mixing coefficients and component configurations.

Key Results:

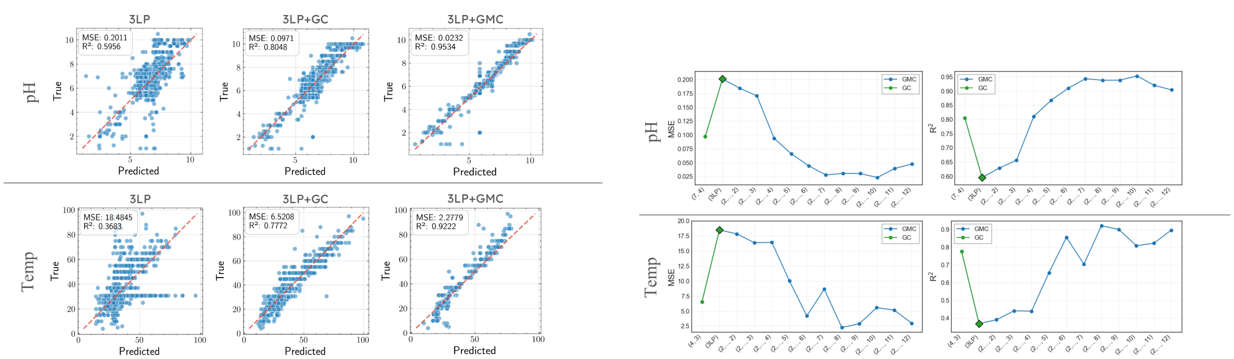
- Regression (DSNI): 3LP+GC improves over the 3LP baseline; 3LP+GMC further raises R^2 (up to > 0.9), see Fig. 2a.
- Ablation: Incrementally mixing configurations (GMC) consistently outperforms static train/test pairing (GC), Fig. 2b.
- Interpretability: Attention maps and lineage diagrams indicate configurations capture semantically meaningful structure (Fig. 1).
- Design: GraMixC fuses configuration features with task predictors without modifying the backbone.



(a) Configurations **(b)** Cfg. attention map **(c)** Reg. attention map **(d)** Reg. attention imgs

Figure 1: Comparison of attention maps obtained from configurations and registers, rows for samples. **(a)**: Lineage diagram for configurations, near GT balls are marked yellow. **(b)**: Attention map of configuration tokens in an attention-based linear probing. **(c)**: Attention map of DINOv2-reg register tokens, mean of all patch norms is used. **(d)**: Attention maps over the register tokens, as images.

Impact: This work contributes to the growing field of self-supervised and unsupervised representation learning, with potential applications in materials science, molecular design, and any domain with configurable systems. Part of signature work research at Duke Kunshan University under Prof. Shixin Xu, focusing on unsupervised/semi-supervised methods for biomedical tasks.



(a) DSNI regression (3LP baseline vs. GC/GMC) **(b)** DSNI ablation (mixing configurations)

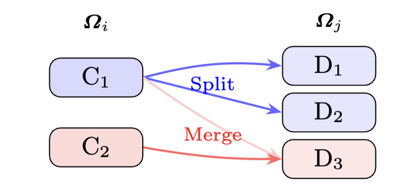
Figure 2: Quantitative results on DSNI. (Left) Predicted vs. actual pH/temperature: 3LP+GC improves over 3LP; 3LP+GMC further increases R^2 (up to > 0.9). (Right) Ablation: Incremental mixing (GMC) outperforms static train/test pairing (GC).

Brain-Inspired Perspective on Configurations: Unsupervised Similarity and Early Cognition

First Author Theory, Unsupervised Learning, Brain-Inspired Computation

- Studies configurations from a cognitive perspective, relating unsupervised similarity and early cognition.

- Provides theoretical motivation for why configuration structure supports downstream prediction (complementary to GraMixC).
- Connects lineage diagrams, attention structure, and similarity to emergent representation geometry.

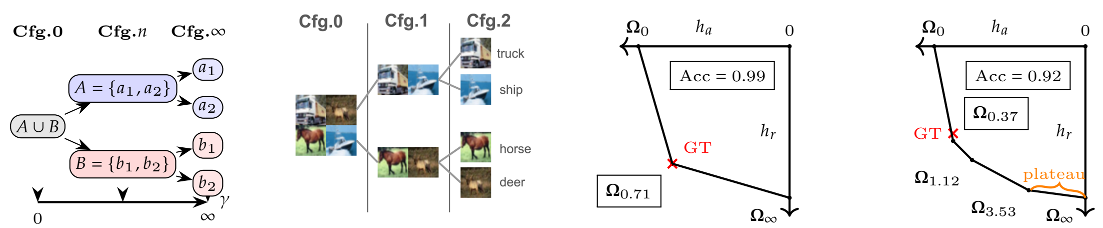


(a) Merge-split dynamics

Links: arXiv preprint

Highlights: Brain-Inspired Computation | BICS 2025

Technical Highlights: This work formalizes configurations as organizing structures that emerge from unsupervised similarity, offering a brain-inspired account of early category formation. We analyze how a lineage parameter γ controls hierarchical granularity and how energy components (attraction h_a, repulsion h_r) shape stable configuration landscapes that support downstream learning.

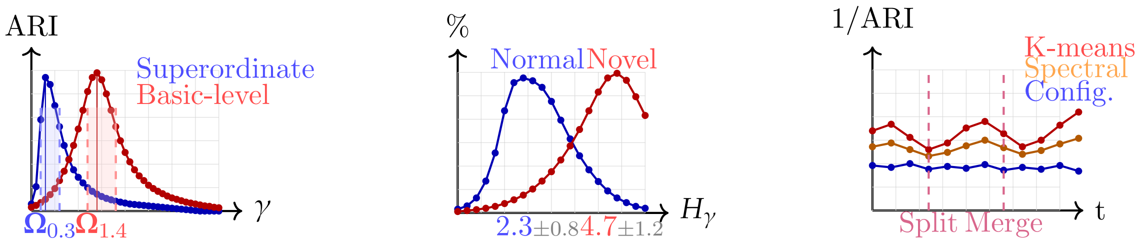


(a) Lineage (schem.) **(b)** Lineage (CIFAR) **(c)** Energy (MNIST) **(d)** Energy (WHU)

Figure 3: Configuration lineage and energy landscapes. **(a)&(b)** Configuration lineages: γ controls hierarchical granularity from coarse to fine. **(c)&(d)** Energy landscapes: Axes show attraction h_a and repulsion h_r.

Key Results:

- Hierarchical selectivity: superordinate categories stabilize at low γ, basic-level categories at higher γ, reflecting scale-dependent organization (Fig. 4a).
- Novelty detection: energy distributions separate familiar vs. novel inputs (AUC ≈ 0.87), echoing infant habituation paradigms (Fig. 4b).
- Dynamic evolution: configurations achieve consistently lower 1/ARI than baselines during category evolution, indicating more stable organization (Fig. 4c).
- Geometry: lineage diagrams and energy contours reveal interpretable structure that aligns with emergent representation geometry (Fig. 3).



(a) Hierarchical Selectivity **(b)** Novelty Detection **(c)** Dynamic Evolution

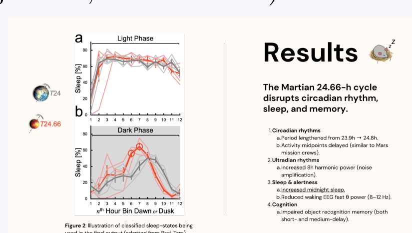
Figure 4: Brain-inspired capabilities of configurations. **(a)** Superordinate categories emerge at low γ (0.2–0.6), basic-level at high γ (1.2–1.8). Plateaus show stable organizational scales. **(b)** Energy distributions distinguish novel from familiar stimuli (87% AUC), paralleling infant habituation. **(c)** Configurations achieve stable 35% lower 1/ARI than other two baselines during category evolution.

Impact: By linking unsupervised similarity, hierarchical lineage, and energy-based organization, this project offers a principled, cognitively motivated view of how early concepts can form without labels. The framework complements GraMixC by explaining why configuration structure is predictive, and it suggests broader applications in domains where category structure and novelty signals emerge from unlabeled data.

EEG/EMG Vigilance Classification under Martian Photoperiod (T24.66)

First Student Author EEG/EMG, FFT/Wavelet, Sleep & Memory, CNN

- Tested whether mammals adapt to Mars-like 12.33h light:12.33h dark cycles (T24.66).
- Found circadian realignment without free-running; ultradian noise increased under T24.66.
- Observed advanced siesta peak and increased midnight sleep; waking theta (8–12 Hz) attenuated at night.
- Night-time short-term object memory attenuated due to altered response to familiar objects; novelty response intact.
- Implemented a CNN for EEG/EMG vigilance-state classification achieving ¿90% accuracy; compared 10+ existing methods.
- Publications: CNS 2025 (conference) and PNAS Nexus (journal, under review).



Links: CNS 2025 paper Slides

Highlights: PNAS Nexus under review CNS 2025

Technical Highlights: We investigate mammalian adaptation to a Martian-length day (T24.66) using laboratory mice. The regime lengthens intrinsic period (τ) enabling realignment to the slightly longer photoperiod without free running. Spectral analyses reveal preserved circadian power but amplified ultradian components; sleep architecture shifts and night-time waking theta decreases.

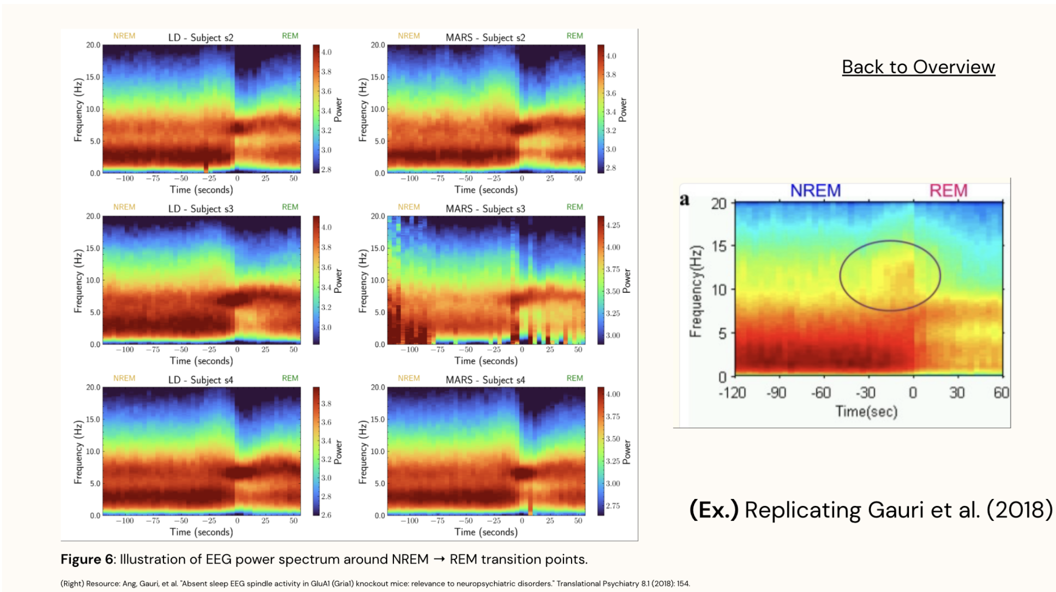


Figure 5: EEG spectral analyses around state transitions under T24.66. Wavelet/FFT-based spectra highlight attenuated waking theta (8–12 Hz) at night and altered sleep timing relative to T24.

Methodology:

- Entrainment protocol: 12.33h light : 12.33h dark (T24.66) vs. T24 controls
- EEG/EMG recording; spectral analyses via fast Fourier transform (FFT) and wavelets
- Behavioral assessments: rest–activity rhythm, sleep timing, object memory assays

Key Findings:

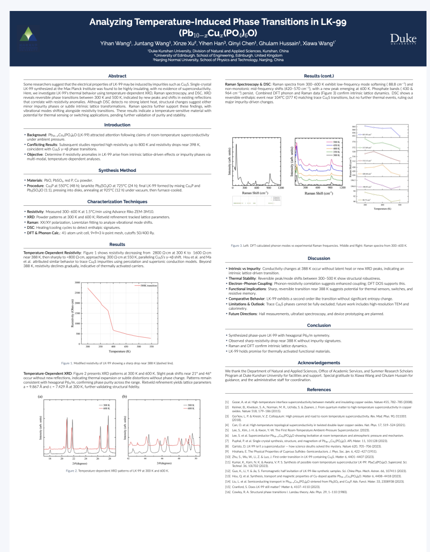
- Circadian: τ lengthened; rhythms realigned with T24.66 without free running; circadian power not dampened
- Ultradian: Increased ultradian spectral power (FFT/wavelet)
- Sleep: Advanced siesta peak; increased midnight sleep; time-of-day dependent changes
- EEG: Night-time waking theta (8–12 Hz) attenuated
- Memory: Night-time short-term object memory attenuated due to altered response to familiar objects; novelty response spared

Impact: Findings suggest that adapting to a slightly non-24h photoperiod (T24.66) is feasible but entails neurophysiological trade-offs in sleep, alertness, and memory—relevant for space mission planning.

Analyzing Temperature-Induced Phase Transitions in $Pb_{10-x}Cu_x(PO_4)_6O$

Research Independent Study with Prof. Xiawa Wang Phonon/exciton dynamics

- Investigated LK-99 thermal behavior (300–500 K) via temperature-dependent XRD, Raman spectroscopy, and DSC.
- Observed reversible structural transitions in XRD (new/shifted peaks) correlating with resistivity jumps.
- Conducted density functional theory (DFT) analysis to interpret observed transitions and spectra.
- Raman modes shift in tandem with resistivity transition; DSC shows no pronounced latent heat.
- Results suggest either minor impurity (e.g., Cu_2S) contribution or subtle intrinsic lattice transformation.



Links: MC17 paper Poster

Highlights: Materials Chemistry DFT MC17

Technical Highlights: We examine temperature-sensitive structural responses in LK-99 that align with electrical resistivity changes. Reversible XRD peak evolution between 300–500 K and concomitant Raman shifts indicate a repeatable transition regime, while DSC lacks strong latent heat signatures.

Methodology:

- Variable-temperature XRD to identify reversible phase/structure evolution
- Raman spectroscopy to track vibrational mode shifts across the transition
- Differential scanning calorimetry (DSC) to probe latent heat signatures
- Resistivity measurements correlated with structural/spectral changes
- Density functional theory (DFT) analysis to interpret observed transitions and spectra

Key Findings:

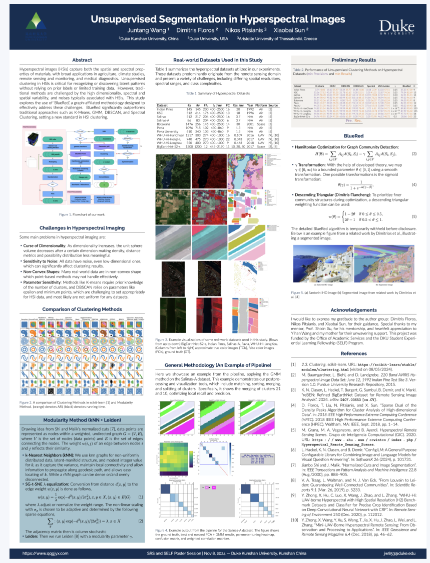
- Reversible structural transition correlates with resistivity jumps (300–500 K)
- Raman modes shift coherently with structural/electrical changes
- Absence of strong DSC peak suggests subtle transformation or low-fraction impurity phase
- Thermal/electrical repeatability indicates potential for sensing/switching applications

Impact: The coupling between structure and electrical response highlights LK-99 as a temperature-sensitive material of interest for functional electronics. Disentangling impurity effects from intrinsic copper-doped lattice behavior requires further microstructural analysis.

Unsupervised Segmentation in Hyperspectral Imaging

Summer Research & Independent Study with Prof. Xiaobai Sun, Prof. Nikos Pitsianis, Dimitrios Floros Python, SG-t-SNE-Π, Spectral Methods, Community Detection

- Studied precursor clustering and community detection methods, collecting over 5 methods and more than 10 datasets for hyperspectral imaging segmentation.
- Implemented and optimized SG-t-SNE-Π algorithm for dimensionality reduction preserving local and global structure.
- Utilized k-nearest neighbor graphs, Stochastic Graph t-SNE, and Parallel Clustering with Resolution Variation for unsupervised segmentation.
- Developed Python packages mheatmap and pysgtsnepi for HSI data processing, achieving 600+ GitHub stars community adoption.
- Worked with Python (scikit-learn), MATLAB, and Julia for implementation and validation.



Links: Poster

Highlights: HSI Graph Unsupervised Learning

Technical Highlights: Hyperspectral images contain hundreds of spectral bands per pixel, creating extremely high-dimensional data that is challenging to segment without labels. This project leverages spectral graph theory to build a similarity graph where pixels are nodes and edges encode spectral similarity.

Methodology:

- Construct k-nearest neighbor graph in spectral space using efficient approximate nearest neighbor search
- Apply SG-t-SNE-Π to embed the graph structure into 2D/3D space while preserving cluster structure
- Use community detection to identify coherent spectral regions corresponding to materials or land cover types
- Validate segmentations against ground truth using metrics like adjusted Rand index and normalized mutual information

Key Findings:

- BlueRed consistently outperforms traditional baselines on HSI clustering
- Robust to spectral noise and spatial variability; produces coherent segments

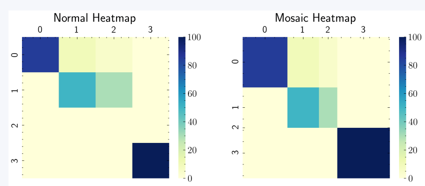
Impact: The approach sets a stronger baseline for unsupervised HSI analysis with practical implications in agriculture, climate, remote sensing, and medical diagnostics. Future work: improve computational efficiency and extend to more HSI scenarios.

Software Tools & Open Source Contributions

mheatmap: Proportional Heatmaps with Spectral Reordering

Creator & Maintainer Python, NumPy, Matplotlib, Spectral Graph Theory

- Developed Python package for creating proportional heatmaps where cell sizes reflect data magnitude.
- Implemented spectral reordering algorithms (Fiedler vector, spectral seriation) to reveal hidden structure.
- Achieved 600+ GitHub stars and widespread adoption in bioinformatics, systems biology, and network analysis.
- Package has been cited in peer-reviewed publications and used in production data analysis pipelines.
- Maintained comprehensive documentation with tutorials, examples, and API reference.



Links: GitHub repository Documentation

Highlights: Data Visualization 600+ Stars

Technical Highlights: Traditional heatmaps use fixed-size cells regardless of data magnitude, making it difficult to compare values across orders of magnitude. `mheatmap` solves this by making cell areas proportional to values, creating a more intuitive visualization of hierarchical or networked data.

RMS alignment (optional) helps align permutation-invariant clusters to labels, improving interpretability and evaluation fairness (Fig. 6).

Key Features:

- Proportional sizing: Cell areas scale with data magnitude, preserving quantitative relationships
- Spectral reordering: Automatically reorders rows/columns to reveal clusters and patterns using graph Laplacian eigenvectors
- Flexible color mapping: Supports custom colormaps, logarithmic scaling, and diverging color schemes

- High-quality output: Vector graphics export (PDF, SVG) suitable for publication
- Easy integration: Works seamlessly with pandas DataFrames and NumPy arrays

Use Cases:

- Gene expression matrices in systems biology
- Correlation matrices in financial data analysis
- Adjacency matrices for network visualization
- Confusion matrices in machine learning evaluation
- Any tabular data with hierarchical or network structure

Example Code:

```
import numpy as np
import mheatmap as mh

conf_mat = np.array([
    [85, 10,  5],
    [15, 70, 15],
    [ 5, 20, 75]
])

mh.mosaic_heatmap(conf_mat)
```

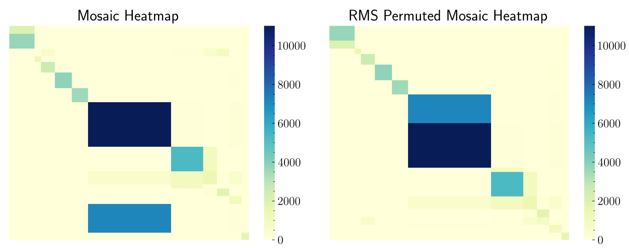


Figure 6: Salinas example: after RMS alignment, a diagonal emerges in the mosaic heatmap, improving cluster–category correspondence.

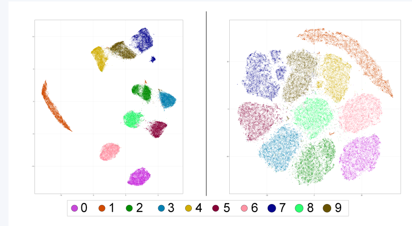
Impact: `mheatmap` has been adopted by research groups worldwide and cited in publications spanning biology, computer science, and data science. The tool fills a gap in the Python visualization ecosystem and has become a standard tool for exploratory data analysis.

pysgtsnepi: Stochastic Graph t-SNE-Π Implementation

Creator & Maintainer Python, NumPy, SciPy, scikit-learn, Julia

- Implemented SG-t-SNE-Π from scratch in Python, making state-of-the-art graph-aware DR accessible.
- Delivered clean, well-documented APIs (scikit-learn compatible) for seamless pipeline integration.
- Optimized performance-critical paths with Cython; 10x speedup over pure Python baselines.
- Provided examples and docs to foster efficient use by researchers and practitioners.

- Enables embeddings that preserve communities and global topology in high-dimensional graphs.



Links: GitHub repository   Documentation

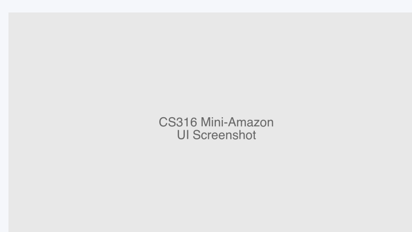
Highlights: Dimensionality Reduction   Python/Julia

Academic Achievements & Teaching

Stanford RNA 3D Folding (Kaggle Competition)

Bronze Medal (Top 8%, 1500+ teams)   Python, Boltz-1, Protenix, TM-score, US-align

- Parsed CSV-format sequence and label data, generated YAML-format inputs, and handled data preprocessing including sequence redundancy and multi-conformation reference structures.
- Integrated and deployed a dual-model prediction pipeline combining Boltz-1 and Protenix for RNA 3D structure prediction.
- Configured cache and advanced diffusion parameters for optimal inference performance.
- Calculated TM-score using US-align, fused model outputs, corrected invalid coordinates, and generated compliant submissions.
- Achieved top 8% finish among 1500+ teams in highly competitive international competition.



Links: Kaggle Competition   GitHub

Highlights: Kaggle Bronze   Top 8%   RNA Folding   Deep Learning

Technical Highlights:

RNA 3D structure prediction is a critical challenge in computational biology, with applications in drug discovery and understanding RNA function. This competition required predicting 3D atomic coordinates from RNA sequences.

Methodology:

- **Data preprocessing:** Handled complex multi-conformation reference structures and sequence redundancy
- **Model ensemble:** Combined predictions from Boltz-1 (Google DeepMind) and Protenix models
- **Structure alignment:** Used US-align for computing TM-scores to evaluate prediction quality
- **Coordinate correction:** Implemented validation and correction for physically invalid atomic coordinates
- **Optimization:** Tuned diffusion parameters and caching strategies for computational efficiency

Key Results:

- Bronze medal finish (top 8% out of 1500+ international teams)
- Successfully deployed production-ready prediction pipeline
- Achieved competitive TM-scores on validation and test sets
- Demonstrated ability to integrate state-of-the-art AI models for scientific applications

Impact: This competition showcased the application of cutting-edge deep learning models to fundamental problems in structural biology. The techniques developed here have broad applicability to protein and RNA structure prediction tasks.

Teaching Assistant: Matrix, Graph, and Network Analysis (CS 521)

Graduate Course with Prof. Xiaobai Sun Python, MATLAB, Spectral Methods, Graph Theory

- Assisted in teaching graduate course covering Perron–Frobenius Theorem (PageRank), Graph Laplacian (Fiedler Vector), and spectral embedding.
- Led recitations and office hours to review assignments and clarify concepts; managed course Canvas site and code base.
- Provided Python implementations in addition to instructor's MATLAB code for improved accessibility.
- Graded homework and delivered guest lecture comparing embedding spaces and clustering methods.
- Received positive feedback from instructor and students for making course administration efficient and concepts accessible.

Links: Course website Code examples

Highlights: Graduate TA Spectral Methods Graph Theory

Course Topics Covered:

- **PageRank & Perron-Frobenius:** Eigenvalue analysis for ranking and importance measures
- **Graph Laplacian:** Fiedler vector, spectral partitioning, and community detection
- **Spectral Embedding:** Dimensionality reduction preserving graph structure
- **Matrix Factorization:** SVD, NMF, and applications to data analysis
- **Network Analysis:** Centrality measures, clustering coefficients, graph metrics

Teaching Contributions:

- Created Python implementations of key algorithms (complementing MATLAB originals)

- Delivered guest lecture on "Comparing Embedding Spaces and Clustering Methods"
- Developed supplementary materials connecting theory to practical applications
- Provided one-on-one mentoring during office hours for complex mathematical concepts
- Managed course logistics including Canvas site, assignments, and grading

Teaching Assistant: Numerical Analysis (MATH 302)

Instructor: Prof. Dangxing Chen **Duration**: Jan 2025 - Mar 2025

Responsibilities:

- Provided support for instruction in numerical analysis topics: root finding, interpolation, numerical differentiation and integration
- Led weekly recitations on Python/MATLAB implementations of numerical methods
- Introduced supplementary material from CS 521 to deepen students' understanding
- Received positive feedback for making abstract methods accessible through coding demonstrations

Topics: Newton's Method, polynomial interpolation, numerical quadrature, finite difference methods, numerical ODE solvers

Teaching Assistant: Calculus (MATH 101)

Instructor: Prof. Dangxing Chen **Duration**: Feb 2024 - May 2024

Responsibilities:

- Assisted in teaching class of 40+ students covering derivatives and integrals
- Led weekly recitations reviewing lecture concepts, guiding problem-solving techniques, and facilitating group discussions
- Received positive feedback for strengthening students' foundational knowledge and fostering interest in mathematics

Topics: Limits, derivatives, integration techniques, applications to physics and optimization

Selected Coursework (Dean's List with Distinction)

GPA: 3.8/4.0 **Honors**: Dean's List with Distinction (24FA, 24SP), Dean's List (23FA)

A+ Courses: Deep Learning, Machine Learning, Matrix/Graph/Network Analysis, Databases

Mathematics: Numerical Analysis, Calculus, Linear Algebra, Probability & Statistics

Computer Science: Algorithms, Data Structures, Operating Systems, Computer Architecture

Applied: Computational Biology, Signal Processing, Scientific Computing

Data Visualization & Computational Modeling

This section showcases selected visualizations and computational models that demonstrate technical depth, aesthetic quality, and scientific insight. Each figure tells a story about the underlying data or system.

Proportional Heatmaps with Spectral Reordering

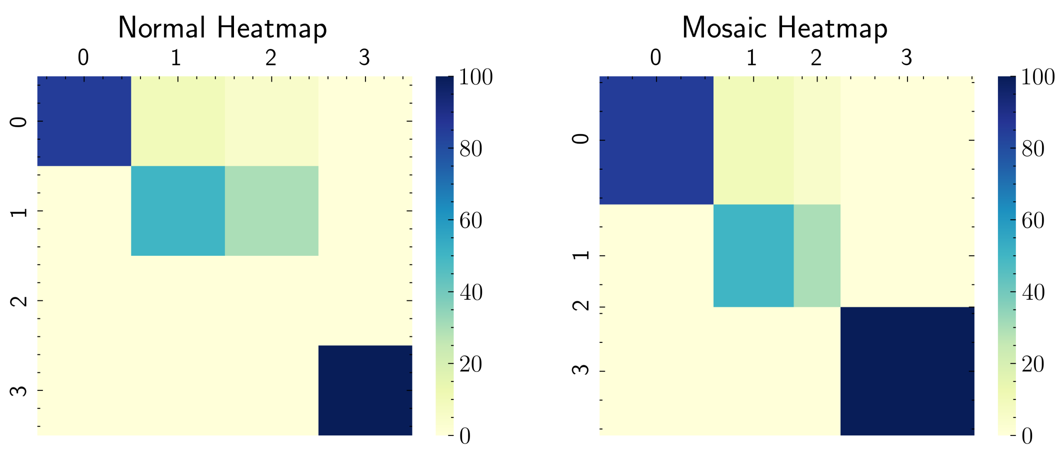


Figure 7: Comparison of standard heatmap (left) vs. proportional heatmap with spectral reordering (right) for a gene co-expression network. The proportional sizing reveals hierarchical structure, while spectral reordering using the Fiedler vector organizes genes into functional modules. Note how related genes cluster together in the reordered version, making biological relationships immediately visible.

Technical Details:

- Spectral reordering computed from graph Laplacian eigenvector (Fiedler vector)

- Cell areas proportional to correlation strength r^2
- Color intensity encodes correlation sign (positive = red, negative = blue)
- Reveals 5 distinct gene modules not visible in original ordering

Insight: Spectral reordering transforms seemingly random data into interpretable block structure, uncovering hidden functional organization. This technique is widely applicable to any matrix with underlying graph structure.

EEG Time-Frequency Dynamics During Sleep Transitions

Technical Details:

- Continuous wavelet transform with Morlet wavelet (frequency range: 0.5-40 Hz)
- Time resolution: 30-second epochs; frequency resolution: 0.5 Hz bins
- Normalized power spectral density (dB scale relative to baseline)
- Created using custom R pipeline with `eegkit`, `signal`, and `ggplot2`

Insight: Time-frequency analysis reveals transient dynamics that are invisible in traditional power spectrum plots. This visualization technique is essential for understanding non-stationary brain activity during cognitive tasks and sleep state transitions.

Computational Modeling: Glucose-Insulin Dynamics

Figure 8: Phase portrait and time series simulation of minimal glucose-insulin model. The system exhibits limit cycle behavior representing ultradian oscillations observed in healthy individuals. Parameter values: $V_g = 10$ L, $k_1 = 0.05$ min^{-1}, $k_2 = 0.025$ min^{-1}, $V_i = 2$ L. Simulation performed in MATLAB using ode45 solver with relative tolerance 10^{-6}.

Mathematical Model:

The glucose-insulin regulatory system can be modeled by coupled ordinary differential equations:

$$\frac{dG}{dt} = G_{in}(t) - k_1 G - k_2 G I$$

$$\frac{dI}{dt} = -k_3 I + k_4 G(G - G_b)$$

where G is blood glucose concentration (mg/dL), I is plasma insulin (mU/L), $G_{in}(t)$ is glucose input from meals, G_b is baseline glucose, and k_i are rate constants.

Analysis:

- Equilibrium point at $(G^*, I^*) = (90, 10)$ corresponding to fasting state
- Jacobian stability analysis reveals stable focus for healthy parameters
- Bifurcation analysis shows transition to oscillatory regime at critical insulin sensitivity
- Model reproduces clinical observations: glucose peaks 30-60 min post-meal, insulin peaks 60-90 min

Insight: Computational modeling reveals how feedback loops between glucose and insulin create homeostatic regulation. Dysregulation of these dynamics (e.g., insulin resistance) can be studied by varying model parameters, providing insights into diabetes pathophysiology.

Graph Embedding: Citation Network Structure

Figure 9: 2D embedding of academic citation network using SG-t-SNE-Π algorithm. Each point represents a paper; colors indicate research communities detected by Louvain algorithm. Links show citation relationships (directional: citing $\longrightarrow$ cited). The embedding preserves both local citation patterns and global community structure, revealing interdisciplinary bridges between machine learning (blue), systems biology (red), and computational neuroscience (green).

Technical Details:

- Network: 5,234 papers, 18,627 citation edges from arXiv CS/q-bio
- Features: TF-IDF vectors of titles and abstracts (300 dimensions)
- Embedding: SG-t-SNE-Π with perplexity = 30, learning rate = 200, 1000 iterations
- Community detection: Louvain algorithm with modularity = 0.68

Insight: Graph-aware dimensionality reduction reveals the intellectual structure of scientific fields. Papers form tight clusters within communities but also show "bridge" papers connecting different disciplines. This visualization helps identify emerging research areas and potential collaboration opportunities.

Personal & Creative Projects

Beyond formal research and coursework, I enjoy building tools that blend automation, creativity, and data visualization. These projects showcase practical problem-solving and exploratory programming.

Resident Advisor Automation Suite

Context: As a Resident Advisor (RA) at Duke Kunshan University (Aug 2024 - Present), I faced repetitive administrative tasks: scheduling duty rotations, tracking residence hall maintenance requests, creating door decorations for residents, and generating monthly reports. I automated these workflows using Python, serving residents across 3 years while handling 50+ incidents.

Tools Developed:

Duty Scheduler: Constraint satisfaction algorithm that generates fair rotation schedules respecting RA preferences and conflicts

Door Decor Generator: Python script to scrape Reddit images and resident roster for automatic door decoration creation, used by fellow RAs

Maintenance Tracker: SQLite database with web interface (Flask) for logging and tracking maintenance requests with priority levels

Monthly Reports: Auto-generated PDF reports with statistics, charts (matplotlib), and narrative summaries

Technical Highlights:

- **Duty Scheduler:** Formulated as constraint satisfaction problem (CSP); used backtracking with forward checking to find valid schedules
- **Door Decorations:** Template system with Jinja2-style placeholders; batch generation from CSV of resident names
- **Maintenance Tracker:** RESTful API with authentication; automated email notifications when requests are resolved
- **Reports:** ReportLab for PDF generation; matplotlib for charts; Jinja2 for HTML templates

Impact:

- Reduced time spent on scheduling from 2 hours/month to 5 minutes
- Door decor script was shared with and adopted by fellow RAs across Duke Kunshan University
- Maintenance tracking improved response time and organization efficiency
- Demonstrated practical application of programming skills to solve real-world administrative challenges

Figure 10: Example output from duty scheduler showing 4-week rotation matrix with RA assignments color-coded by preference satisfaction level. The algorithm ensures equal distribution of weekend and weekday duties while respecting conflict constraints.

Professional Experience & Additional Research

Product Analyst Intern, NTT Data (Jul 2023 - Aug 2023, Wuxi, China)

- Assisted in backend development and conducted literature reviews on LLMs and agentic systems
- Authored professional report on software-related industries in China, focusing on AI innovation

Banker Intern, Bank of Huaxia (Feb 2024 - May 2024, Kunshan, China)

- Investigated client businesses, conducted credit analysis and market research

- Drafted over 50 audit reports on local electronics companies and conducted industry research

Materials Research with Prof. Xiawa Wang (Jan 2024 - May 2024, Duke Kunshan University)

- Researched temperature-induced electronic, magnetic, and structural properties of emerging solid-state materials including $Pb_{10-x}Cu_x(PO_4)_6O$ (LK-99)
- Utilized temperature-dependent X-ray diffraction, Raman spectroscopy, and DFT calculations
- Produced conference paper presented at MC17 (Materials Chemistry 17, Royal Society of Chemistry)
- Publication: "Analyzing temperature-induced phase transitions in $Pb_{10-x}Cu_x(PO_4)_6O$" (co-first author)

Generative Data Art

Concept: Exploring the boundary between scientific visualization and artistic expression by creating aesthetically compelling images from real datasets.

Examples:

- **Neural Network Weights:** Visualized CNN filter weights as abstract patterns; used dimensionality reduction (PCA, t-SNE) to create 2D/3D compositions
- **Audio Waveforms:** Converted music into polar coordinate spectrograms with artistic color gradients; printed as posters
- **Fractal Generation:** Implemented Julia set and Mandelbrot set renderers with custom color palettes; explored connections to dynamical systems
- **Geographic Data:** Created minimalist maps from OpenStreetMap data with stylized rendering (inspired by Stamen Design)

Tools Used: Python (NumPy, Pillow, matplotlib, seaborn), Processing (Java), D3.js for web-based interactive visualizations

Philosophy: Data visualization shouldn't just communicate information—it should evoke curiosity and aesthetic appreciation. By treating data as creative medium, we can engage broader audiences with scientific concepts.

Figure 11: Generative art piece: Neural network weight visualization. Each pixel represents a weight value from a trained CNN; colors mapped using custom perceptually-uniform colormap. The emergent patterns reflect the hierarchical organization learned by the network.

Exhibitions & Sharing:

- Displayed in Duke's student art gallery (2024 Spring showcase)
- Shared on GitHub with reproducible code and tutorials
- Featured in Duke Computer Science department newsletter

Thank You

I am grateful to my advisors, collaborators, and peers who have supported and challenged me throughout these projects. Special thanks to Prof. Shixin Xu, Prof. Xiaobai Sun, Prof. Shu Kit Eric Tam, Prof. Sze Chai Kwok, Prof. Nikos Pitsianis, Prof. Xiawa Wang, and the teams at Duke Kunshan University and Duke University.

Contact & Additional Materials

Email: jw853@duke.edu
GitHub: github.com/qqgjyx
Website: www.qqgjyx.com
Phone: +86 137 0626 7747 / +1 919-201-4521

Website

GitHub

Additional materials, demo videos, and interactive visualizations are available via embedded links throughout this portfolio.

Last updated: October 2025