Fostering Cross-Disciplinary Collaboration in Biology, Medicine, and Computational Science
Poster Session Abstracts
July 29, 2025

Multiscale Topology-enabled Artificial Intelligence for Drug Discovery
Dong Chen, Michigan State University

Although natural language processing (NLP) models have achieved remarkable success across various domains, their application in computational biology has been constrained by the difficulty of integrating critical 3D structural information from biological sequences, which traditional NLP architectures cannot easily process. To address this limitation, we present TopoFormer, an innovative model that combines NLP with a multiscale topology framework called the Persistent Topological Hyperdigraph Laplacian (PTHL). PTHL translates intricate 3D protein-ligand interactions into sequences of topological invariants and homotopic shapes that are compatible with NLP methodologies, thereby capturing essential spatial interactions. TopoFormer surpasses conventional algorithms and state-of-the-art deep learning approaches, demonstrating superior performance in scoring, ranking, docking, and screening tasks across diverse benchmark datasets. Beyond computational biology, this method provides a versatile framework for transforming high-dimensional structured data into NLP-compatible formats, with promising applications in drug discovery, including efforts to combat COVID-19 and address drug-resistant mutations.

Data-Driven Modeling of Amyloid-beta Targeted Antibodies for Alzheimer's Disease
Lida (Kobra) Rabiei, Pennsylvania State University

Alzheimer's disease (AD) is characterized by the accumulation of amyloid beta, which is strongly associated with disease progression and cognitive decline. Despite the approval of monoclonal antibodies targeting Aβ, optimizing treatment strategies while minimizing side effects remains a challenge. This study develops a mathematical framework to model Aβ aggregation dynamics, capturing the transition from monomers to higher order aggregates, including protofibrils, toxic oligomers, and fibrils, using mass-action kinetics and coarse-grained modeling. Parameter estimation, sensitivity analysis, and data-driven calibration ensure model robustness. An optimal control framework is introduced to identify the optimal dose of the drug as a control function that reduces toxic oligomers and fibrils while minimizing adverse effects, such as amyloid-related imaging abnormalities (ARIA). The results indicate that Donanemab achieves the most significant reduction in fibrils. These findings provide a quantitative basis for optimizing AD treatments, providing valuable insight into the balance between therapeutic efficacy and safety.

Preventing future zoonosis: SARS-CoV-2 mutations enhance human-animal cross transmission
JunJie Wee, Michigan State University

The COVID-19 pandemic has significantly influenced the evolution of the SARS-CoV-2 virus, leading to the emergence of subvariants with increased ability to infect humans. However, these evolutionary advantages may not necessarily apply to transmission between animals and humans. This study proposes that mutations enhancing the virus's ability to infect animal hosts do not always result in greater infectivity in humans. It also explores the possibility that certain mutations acquired in animals could later enhance the virus's adaptability in humans. To investigate this, the researchers focus on mutations in the virus's receptor-binding domain (RBD) that may promote transmission between humans and animals. They developed a multitask Topological Laplacian deep learning model, MT-TopLap, trained on various deep mutational scanning datasets, to predict how mutations affect the binding strength between the RBD and ACE2 receptors in different species,

including humans, cats, bats, deer, and hamsters. Their analysis highlights specific RBD mutations—such as Q498H in the original SARS-CoV-2 and R493K in the BA.2 variant—that may increase the likelihood of cross-species transmission.

Laplacian Eigenfunction-Based Neural Operator for Learning Nonlinear Reaction-Diffusion Dynamics
Jindong Wang, Pennsylvania State University

Learning reaction-diffusion equations has become increasingly important across scientific and engineering disciplines, including fluid dynamics, materials science, and biological systems. We propose the Laplacian Eigenfunction-Based Neural Operator (LE-NO), a novel framework designed to efficiently learn nonlinear reaction terms in reaction--diffusion equations. LE-NO models the nonlinear operator on the right-hand side using a data-driven approach, with Laplacian eigenfunctions serving as the basis. This spectral representation enables efficient approximation of the nonlinear terms, reduces computational complexity through direct inversion of the Laplacian matrix, and alleviates challenges associated with limited data and large neural network architectures---issues commonly encountered in operator learning. We demonstrate that LE-NO generalizes well across varying boundary conditions and provides interpretable representations of learned dynamics. Numerical experiments in mathematical physics showcase the effectiveness of LE-NO in capturing complex nonlinear behavior, offering a powerful and robust tool for the discovery and prediction of reaction--diffusion dynamics.

Topological Deep Learning for Predicting Mutational Impacts on COVID-19 Variants: Toward Supporting Vaccine and Antibody Drug Design
Rui Wang, New York University

Topological Data Analysis (TDA) offers a powerful framework for extracting insights from complex biological datasets. However, traditional tools like Persistent Homology (PH) are limited in capturing non-topological information, such as the shape evolution of biomolecules. To overcome this, we introduced Persistent Spectral Graphs (PSGs), a novel topological representation that encodes multidimensional shape dynamics from biological data. We also developed the open-source package HERMES to compute PSG-based features, which I later applied across diverse biomedical applications, including RNA motifs and protein–protein/peptide interactions.
During the COVID-19 pandemic, we applied PSG-derived features to train TopLapNet, a Topological Deep Learning (TDL) model for predicting binding free energy (BFE) changes in the SARS-CoV-2 Spike protein caused by mutations. This approach produced three key results: 1) Successful prediction of 2 crucial mutations on Spike residues L452 and N501 out of nearly 10,000 residues that could lead to more infectious strains; 2) Early report to FDA in March 2021 that two Eli Lilly's antibodies, Etesevimab and Bamlanivimab, were susceptible to escape due to several key RBD mutations found in Beta and Gamma variants; 3) Accurate forecast in May 2022 of the incoming dominance of BA.4 and BA.5 the time they comprised fewer than 3% of total sequences.
These findings underscore the potential of Topological Deep Learning to support predictive modeling for vaccine and antibody drug design.

Giving Virtual Mice a Good Death: Modeling Time of Death in Digital Twins Based on Tumor Dynamics
Chloe George, Pennsylvania State University

Drug development is expensive and slow, with nearly 90% of treatments failing during the preclinical phase. Virtual mice can help reduce these costs, but traditional approaches to survival analysis struggle to generalize time of death assignments. We present a method for generating virtual mouse populations by linking tumor growth dynamics to survival in a mechanistic framework. Our approach recovers individual-level patterns from homogeneous trial data and extrapolates to heterogeneous populations. We validate our model using independent time-course data from a separate mouse strain, showing that digital twins can generalize beyond their original dataset.

Optimal Control For Anti-Abeta Treatment in Alzheimer's Disease using a Reaction-Diffusion Model
Sun Lee, Pennsylvania State University

Alzheimer's disease (AD) is a progressive neurodegenerative disorder with limited treatment options to halt cognitive decline. While many mathematical models have been proposed to understand AD pathology, most rely on ordinary differential equations and overlook spatial heterogeneity in amyloid-beta accumulation.
In this work, we develop a spatially resolved reaction-diffusion model for Aβ dynamics, incorporating optimal control to design individualized treatment strategies. The objective is to minimize plaque burden while accounting for treatment-related side effects, such as amyloid-related imaging abnormalities (ARIA). We establish conditions for well-posedness and uniqueness of the optimal solution.
Numerical simulations, performed via the Finite Element Method using patient-specific PET imaging data, reveal that optimal dosing schedules significantly outperform constant regimens—reducing Aβ more effectively while mitigating adverse effects.
By integrating spatial dynamics and personalized treatment planning, our framework offers a novel approach to refining therapeutic interventions for Alzheimer's disease.

Modeling Early-Onset Cancer Trends: Integrating Tumor Size at Diagnosis to Distinguish Risk from Detection
Navid Mohammad Mirzaei, Columbia University

Recent studies have reported a rise in early-onset cancer cases (diagnosed before age 50), prompting debate over whether this increase reflects earlier detection due to non-specific medical testing—suggested by decreasing tumor size at diagnosis—or a genuine rise in underlying cancer risk, or both. Traditional Multi-Stage Clonal Expansion (MSCE) models assume cancer is detected at the emergence of the first malignant cell, though later adaptations have introduced lag times or stochastic detection to represent delays in diagnosis. Here, we present a new approach, explicitly incorporating tumor-size-at-diagnosis into the MSCE framework. This extension accounts for improvements in cancer detection over time and helps distinguish between apparent and true increases in early-onset cancer incidence. We show the updated model is structurally identifiable and yields more accurate parameter estimates than the classic version.
We applied this model to colorectal, breast, and thyroid cancers across three birth cohorts (1950–1954, 1965–1969, and 1980–1984) to evaluate shifts in underlying risk while adjusting for changing detection dynamics. Our results indicate faster progression through carcinogenic stages and shorter mean sojourn times (the interval between malignant transformation and diagnosis) in more recent cohorts.
Additionally, we used the model to assess the effects of established screening programs for breast and colorectal cancer. Our results align with well-documented differences in screening effects between these cancers. Together, these findings highlight the importance of incorporating tumor-size-at-diagnosis into

cancer models and provide support for a true rise in early-onset cancer risk in recent generations for colorectal, breast, and thyroid cancers.

Persistent Directed Flag Laplacian (PDFL)-Based Machine Learning for Protein–Ligand Binding Affinity Predic on

Mushal Zia, Michigan State University

Direc onality in molecular and biomolecular networks plays a significant role in the accurate representa on of the complex, dynamic, and asymmetrical nature of interac ons present in protein-ligand binding, signal transduc on, and biological pathways. Most tradi onal techniques of topological data analysis (TDA), such as persistent homology (PH) and persistent Laplacian (PL), overlook this aspect in their standard form. To address this, we present the persistent directed flag Laplacian (PDFL), which incorporates directed flag complexes to account for edges with direc onality originated from polariza on, gene regula on, heterogeneous interac ons, etc. This study marks the first applica on of the PDFL, providing an in-depth analysis of spectral graph theory combined with machine learning. Besides its superior accuracy and reliability, the PDFL model offers simplicity by requiring only raw inputs without complex data processing. We validated our mul-kernel PDFL model for its scoring power against other state-of-art methods on three popular benchmarks, namely PDBbind v2007, v2013, and v2016. Computa onal results indicate that the proposed PDFL model outperforms compe tors in protein-ligand binding affinity predic ons, indica ng that PDFL is a promising tool for protein engineering, drug discovery, and general applica ons in science and engineering.

A Systematic Computational Framework for Practical Identifiability Analysis

Shun Wang, Pennsylvania State University

Practical identifiability is a fundamental challenge in the data-driven modeling of biological systems, as many model parameters cannot be directly measured and must be estimated from experimental data. We propose a novel mathematical framework for practical identifiability analysis in dynamic models. Starting from a rigorous mathematical definition, we prove that practical identifiability is equivalent to the invertibility of the Fisher Information Matrix (FIM). We further establish the relationship between practical identifiability and coordinate identifiability, introducing an efficient metric that simplifies and accelerates identifiability assessment compared to traditional profile likelihood methods. To address non-identifiable parameters, we incorporate new regularization terms, enabling uncertainty quantification and improving model reliability. To support experimental design, we propose an optimal data collection algorithm that ensures all model parameters are practically identifiable. Applications to Hill functions, neural networks, and biological models demonstrate the effectiveness and computational efficiency of the proposed framework in uncovering critical biological processes and identifying key observable variables.