Workshop On
``Scientific Computing meets Machine Learning and Life Sciences''

October 7th - 9th, 2019
Texas Tech University
Lubbock, TX

TEXAS TECH UNIVERSITY™

This conference is partially supported by National Science Foundation, the Department of Mathematics & Statistics at Texas Tech University, and the Horn Professorship of Linda J. S. Allen.

SCIENTIFIC COMMITTEE

Wei Cai, Southern Methodist University

Carlos Castillo-Chavez, Arizona State University

Mac Hyman, Tulane University

Sudeep Sarkar, University of South Florida

Chunmei Wang, Texas Tech University

ORGANIZING COMMITTEE

Linda Allen, Mathematics & Statistics

Rattikorn Hewett, Computer Science

Jingyong Su, Mathematics & Statistics

Chunmei Wang, Mathematics & Statistics

Workshop Schedule

October 7, 2019 Monday

7:30am-8:45am Registration and Coffee & Refreshments

8:45am-9:00am Opening Remarks by vice president for Research & Innovation Dr. Joseph Heppert

Chair: Rattikorn Hewett
9:00am-10:00am Plenary Speaker: Qing Nie

10:00am-10:30am Group Photo and Coffee Break

Chair: Linda Allen
10:30am-11:00am Guowei Wei
11:05am-11:35am Xiu Ye
11:40am-12:10pm Gerardo Chowell

12:10-2:00pm Lunch

Chair: Xiu Ye
2:00pm-3:00pm Plenary Speaker: Doron Levy

3:00-3:35pm Coffee Break

Chair: Doron Levy
3:35pm-4:05pm Rays Jiang
4:10pm-4:45pm Pavan Turaga

6pm-8pm Reception at Matador Faculty Lounge

October 8, 2019 Tuesday

8:00am-9:00 am Registration and Coffee & Refreshments

Chair: Xiaobing Feng
9:00am-10:00am Plenary Speaker: Sudeep Sarkar

10:00am-10:30am Coffee Break

Chair: Chunmei Wang
10:30am-11:00am Chun Liu
11:05am-11:35am Joshi Shantanu

11:40am-1:00pm Lunch

1:00pm-2:30pm Poster Section at Matador Room

Chair: Chun Liu
2:30pm-3:00pm James Liu
3:05pm-3:35pm Justin Strait

3:35pm-3:45pm Coffee Break

Moderator: All Organizers
3:45pm-5:30pm Group Discussion

6pm-8pm Tutorial Session titled "Weak Galerkin Finite Element Methods for PDEs" lectured by Xiu Ye and Chunmei Wang at Math Building 010 (pizza provided)

October 9, 2019 Wednesday

8:00am-8:30 am Coffee & Refreshments

8:30am-9:30am *Chair: Anuj Srivastava*
Plenary Speaker: Todd Arbogast

Moderator: Linda Allen
9:35am-10:50am Distinguished Panelists:
Xiaobing Feng
Anuj Srivastava
Guowei Wei

10:50am-11:05am Coffee Break

Chair: Jingyong Su
11:05am-11:35am Juan Gutierrez
11:40am-12:10pm Yanyu Xiao
12:15pm-12:45pm Xiaoning Qian

12:50pm-2:00pm lunch

Moderator: All Organizers

2:00pm-3:00pm Discussion and Conclusion

Plenary Talks

Approximation of advection-diffusion equations using implicit WENO methods
Todd Arbogast, University of Texas at Austin

Abstract: Within the life sciences, researchers use models that often involve advective, diffusive, and reactive processes. These are often formulated as systems of nonlinear advection-diffusion-reaction equations. These equations are often advection dominated; moreover, the diffusion may degenerate (to zero). This means that the solutions to the equations can and often do develop steep fronts or even shock discontinuities. We consider approximation of nonlinear advection-diffusion equations using finite volume methods. These methods require two components: (1) reconstruction of a function from knowledge of its element averages and (2) time stepping (in the sense of the method of lines). We first discuss function reconstruction using the weighted essentially non oscillatory (WENO) framework, which allows accurate reconstruction of function values near function discontinuities. We develop a multi-scale, WENO method with adaptive order (WENO-AO), and analyze its performance theoretically and computationally. The idea is to create a high order accurate reconstruction where the function is smooth, and drop the order of the approximation and bias the reconstruction stencil to one side of a discontinuity. We next discuss time stepping procedures of the implicit type, because advection-diffusion equations are stiff. High order Runge-Kutta (RK) methods can fail near discontinuities (i.e., generate oscillations), so we develop a hybrid approach combining the RK method with a simple and stable backward Euler (BE) method. Finally, we discuss approximation of degenerate diffusion within the finite volume WENO framework.

Modeling the chemotherapy-induced selection of drug-resistant traits during tumor growth
Doron Levy, University of Maryland

Abstract: The emergence of drug-resistance is a major challenge in chemotherapy. In this talk we will present our recent mathematical models for describing the dynamics of drug-resistance in solid tumors. Our models follow the dynamics of the tumor, assuming that the cancer cell population depends on a phenotype variable that corresponds to the resistance level to a cytotoxic drug. We incorporate the dynamics of nutrients and two different types of drugs: a cytotoxic drug, which directly impacts the death rate of the cancer cells, and a cytostatic drug that reduces the proliferation rate. Through analysis and simulations, we study the impact of spatial and phenotypic heterogeneity on the tumor growth under chemotherapy. We demonstrate that heterogeneous cancer cells may emerge due to the selection dynamics of the environment. Our models predict that under certain conditions, multiple resistant traits emerge at different locations within the tumor. We show that a higher dosage of the cytotoxic drug may delay a relapse, yet, when this happens, a more resistant trait emerges. Moreover, we estimate the expansion rate of the tumor boundary as well as the time of relapse, in terms of the resistance trait, the level of the nutrient, and the drug concentration. Finally, we propose an efficient drug schedule aiming at minimizing the growth rate of the most resistant trait. By combining the cytotoxic and cytostatic drugs, we demonstrate that the resistant cells can be eliminated.

Data-driven multiscale modeling of cell fate dynamics
Qing Nie, University of California, Irvine

Abstract: Cells make fate decisions in response to dynamic environmental and pathological stimuli as well as cell-to-cell communications. Recent technological breakthroughs have enabled to gather data in previously unthinkable quantities at single cell level, starting to suggest that cell fate decision is much more complex, dynamic, and stochastic than previously recognized. Multiscale interactions, sometimes through cell-cell communications, play a critical role in cell decision-making. Dissecting cellular dynamics emerging from molecular and genomic scale in single-cell demands novel computational tools and multiscale models. In this talk, I will present our recent works on analyzing single-cell molecular data, and their connections with cellular and spatial tissue dynamics. Our mathematical approaches bring together optimization, statistical physics, ODEs/PDEs, and stochastic simulations along with machine learning techniques. By utilizing our newly developed computational tools along with their close integrations with new datasets collected from our

experimental collaborators, we are able to investigate several complex systems during development and regeneration to uncover new mechanisms in cell fate determination.

Towards Open World Video Event Understanding - Flexible Representations, Commonsense Priors, and Self-Supervised Learning
Sudeep Sarkar, University of South Florida

Abstract: Events are central to the content of human experience. From the constant stream of the sensory onslaught, the brain segments, extracts, represents aspects related to events, and stores in memory for future comparison, retrieval, and re-storage. Contents of events consist of objects/people (who), location (where), time (when), actions (what), activities (how), and intent (why). Many deep learning-based approaches extract this information from videos. However, most methods cannot adapt much beyond what they were trained and are incapable of recognizing new events beyond what they were explicitly programmed or trained. The main limitation of current event analysis approaches is the implicit closed world assumption. The ability to support open world inference is limited by three main aspects: the underlying representation, the source of semantics, and the ability to continuously learn or adapt.
In this talk, I will focus on flexible representations, amenable for open-world, and self-supervised learning that is not dependent on the existence of a large amount of training data. We will see how Grenander's pattern theory-based canonical representation offers an elegant, flexible, compositional mechanism. It naturally models semantic connections between what is observed directly in the image and prior knowledge in large-scale commonsense knowledge bases, such as ConceptNet. The use of knowledge bases such as ConceptNet allows expanding the set of primary objects and actions to a very large (not infinite) set without the need for massive annotated training sets. And finally, if we have time, how predictive learning can be used to continuously learn how to segment a video into elementary event segments, again without training annotations.

Distinguished Panelist Talk

Recent developments in mixed finite element methods for stochastic Stokes and Navier-Stokes equations
Xiaobing Feng, University of Tennessee

Abstract: Besides the mathematical interests, stochastic Stokes and Navier-Stokes equations have been proposed to study turbulence flow under random forcing. Even in the simplest setting,their PDE solutions have very low regularity in time (and in space), which then poses a significant challenge for developing efficient and convergent numerical methods for the stochastic Stokes and Navier-Stokes equations. In particular, the most natural and popular class of numerical methods for those equations, namely mixed finite element methods, had not been proven to work. In this talk I shall present some recent developments in mixed finite element methods for the Stokes and Navier-Stokes equations with multiplicative noise. I shall highlight the establishment of the continuous and discrete stochastic inf-sup conditions and the strong convergence not only for the velocity approximation but also for the pressure approximation, as well as the new analysis techniques used to obtain these results. Numerical experiments will also be presented to validate the theoretical results.

Statistical Morphometrics in Life Sciences
Anuj Srivastava, University of South Florida

Abstract: Ever since D'Arcy Thompson's monumental work "On Growth and Form",
scientists have sought techniques to quantify shapes of biological objects and to help understand their roles in larger biological systems. This quest has been accelerated by a revolution in techniques that probe the chemical, structural, and dynamical nature of molecules, cells, tissues, and organs across scales. Such structured data defies past techniques from Euclidean statistics, and requires careful constructions of representations, models, and analyses. I will provide some examples of modeling and analyzing biological structures using a combination of geometry, statistics, and computational solutions.

Trends and opportunities in life sciences
Guo-Wei Wei, Michigan State University

Abstract: Life science or biological science is believed to be the last forefront of natural sciences. It became molecular in the 1960s and assumed an omics dimension around the dawn of the millennium. The exponential growth of biological data has paved the way for biological sciences to undertake another historical transition from qualitative, phenomenological and descriptive to quantitative, analytical and predictive. This transition gives rise to unprecedented opportunities for mathematicians. Mathematics is indispensable for understanding the rules of life. For example, differential geometry, algebraic geometry, algebraic topology, knot theory, combinatorics, topological graph, and spectral graph are powerful tools for the abstraction and simplification of complex biomolecules, such as proteins, DNA, and RNA, and for revealing their structure-function relationship. Dynamical systems and multiscale analysis are essential techniques for analyzing and elucidating molecular, organelle, and cellular motion and dynamics. Partial differential equations are vital for understanding the transport of ions, molecules, organelles, and cells in the biological environment. Additionally, mathematics is all-known for its strong presence in evolution, population, neuroscience, ecology, virology, immunology, and physiology, which are becoming molecular and data-driven nowadays. Scientific computing facilitates various mathematical approaches to biological sciences. Mathematics underpins machine learning that is the fourth paradigm of biological sciences. Currently, mathematicians are working more closely with experimentalists than ever before.

Invited Talks

Sub-epidemic modeling framework for short-term forecasting epidemic waves
Gerardo Chowell, Georgia State University

Abstract: Simple phenomenological growth models can be useful for estimating transmission parameters and forecasting epidemic trajectories. However, most existing phenomenological growth models only support single-peak outbreak dynamics whereas real epidemics often display more complex transmission trajectories. Methods: We develop and apply a novel sub-epidemic modeling framework that supports a diversity of epidemic trajectories including stable incidence patterns with sustained or damped oscillations to better understand and forecast epidemic outbreaks. We describe how to forecast an epidemic based on the premise that the observed coarse-scale incidence can be decomposed into overlapping sub-epidemics at finer scales. We evaluate our modeling framework using three outbreak datasets: Severe Acute Respiratory Syndrome (SARS) in Singapore, plague in Madagascar, and the ongoing Ebola outbreak in the Democratic Republic of Congo (DRC) and four performance metrics.
Results: The sub-epidemic wave model outperforms simpler growth models in short-term forecasts based on performance metrics that account for the uncertainty of the predictions namely the mean interval score (MIS) and the coverage of the 95% prediction interval. For example, we demonstrate how the sub-epidemic wave model successfully captures the 2-peak pattern of the SARS outbreak in Singapore. Moreover, in short-term sequential forecasts the sub-epidemic model was able to forecast the second surge in case incidence for this outbreak, which was not possible using the simple growth models. Furthermore, our findings support the view that the national incidence curve of the Ebola epidemic in DRC follows a stable incidence pattern with periodic behavior that can be decomposed into overlapping sub-epidemics.
Conclusions: Our findings highlight how overlapping sub-epidemics can capture complex epidemic dynamics, including oscillatory behavior in the trajectory of the epidemic wave. This observation has significant implications for interpreting apparent noise in incidence data where the oscillations could be dismissed as a result of overdispersion, rather than an intrinsic part of the epidemic dynamics. Unless the oscillations are appropriately modeled, they could also give a false positive, or negative, impression of the impact from public health interventions. These preliminary results using sub-epidemic models can help guide future efforts to better understand the heterogenous spatial and social factors shaping sub-epidemic patterns for other infectious diseases.

Machine Learning in Biomedical Research – Early Detection of Disease via Telemetry Signals
Juan B. Gutiérrez, University of Texas at San Antonio

Abstract: Machine learning can solve problems that traditional methods in mathematics are simply unable to tackle. The context for this talk is malaria, an infectious disease caused by members of the Plasmodium genus. There is a lapse of between one and two weeks after infection, called the liver stage, in which detection of malaria is not possible with current technology. Once malaria symptoms begin, the disease might turn deadly within 24 hours, hence the urgency to develop early detection methods. In this talk, I will present high-frequency physiological data captured from telemetry devices surgically implanted in Macaca mulatta and M. fascicularis prior to and during infection with Plasmodium knowlesi sporozoites. The challenges of analyzing this type of data are enormous, and include but are not limited to: selection of method of analysis, data capture, data storage and transmittal, reproducibility of results, etc. Particularly, we will talk about deep learning can be used for this problem. Our results show for the first time that host physiological perturbations can be detected while malaria parasites are multiplying in the liver, a step that precedes blood-stage infections and clinical symptomology. These findings demonstrate that machine learning can be deployed successfully to address problems for which we lack mathematical tools, and although we obtain knowledge without understanding, it is precisely our inability to comprehend the characteristics of the phenomenon detected by machines what opens pathways for the advance of mathematics.

Harnessing the power of genomics data sciences with Machine Learning
Rays H.Y. Jiang, University of South Florida

Abstract: Biological science is in 'big data' era. Particularly, genomics data, such as 23&Me, have entered commercial sectors with projected exponential growth. This bio data revolution is enabled by the novel technology development of miniaturization, standardization and massive production. To face the challenge of outpouring data production, Artificial Intelligence (AI) approaches such as Machine learning (ML), knowledge representation, and Nature Language Processing (NLP) are poised to solve the genomics problems of large volume and high dimension data. At USF (University of South Florida) genomics, we have the unique capacity of integrating cutting-edge genomics, engineering, and domain knowledge to produce high volume of novel genomics data, in the field of infectious diseases, developmental biology and evolutionary biology. We currently are actively solving data problems of 1) very high dimensions of single-cell biology with Manifold Learning 2) unique evolutionary microbiota data with ML 3) large scale network analysis gene function data such as CRISPR and gene expressions. I will review the current status, challenges and future directions of translating genomics big data into actionable knowledge to propel biological and health sciences.

Elastic Hyper-alignment of fMRI Signals
Shantanu Joshi, University of California, Los Angeles

Abstract: This talk presents approaches for temporal matching of fMRI signals. It presents applications for functional magnetic resonance imaging (fMRI) time course and spectral alignment. We show elastic alignment of both amplitude and phase of the fMRI time courses as well as their power spectral densities. Experimental results show significant increases in pairwise node to node correlations and coherences following alignment. Additionally, we show results for task based fMRI signals, where we see improved power of detection of clusters and activations for single subject data.

Thermal Effects in General Diffusion with Biological Applications
Chun Liu, Illinois Institute of Technology

Abstract: All biological activities involve transport and distribution of ions and charged particles in specific biological environments. Moreover, the thermal effects are the key for these activities. In this talk, I will

introduce several extended general diffusion systems motivated by the study of ion channels and ionic solutions in biological cells. A general framework is established, which incorporates the energetic variational approaches (EnVarA) with various thermodynamics and kinematic conditions. In particular, we will focus on the interactions between different species, the boundary effects and the temperature effects.

Machine Learning and Transport Simulations for Groundwater Anomaly Detection
James Liu, Colorado State University

Abstract: In this talk, we present studies on models and algorithms for groundwater anomaly detection. Specifically, conductivity along with four other surrogates are used for identifying anomaly in groundwater, the one-class support vector machine (SVM) technique is utilized for model training, and real data from "Colorado Water Watch" is used for testing the models and algorithms. Design of code modules in Python is briefly discussed. Since groundwater contamination rarely happens in reality, we also use synthetic data from numerical simulations of flow and transport in porous media to test the anomaly detection code modules. This is based on the joint work with Ken Carlson, Jianli Gu, and Huishu Li at Colorado State University.

A Bayesian Gamma-Negative-Binomial Model for Single-Cell RNA-Seq Data
Xiaoning Qian, Texas A&M University

Abstract: A generative hierarchical gamma-negative-binomial (hGNB) model is developed to analyze single-cell RNA-sequencing (scRNA-seq) data, obviating the need for explicitly modeling zero inflation in many existing methods. Due to its generative nature, hGNB can account for covariate effects at both the gene and cell levels to identify complex latent representations of scRNA-seq data, without the need for commonly adopted pre-processing steps such as normalization. Efficient Bayesian model inference is derived by exploiting conditional conjugacy via novel data augmentation techniques. Experimental results on both simulated data and several real-world scRNA-seq datasets will be presented to show the potential of hGNB.

A parametric bootstrap region construction for the mean elastic shape
Justin D Strait, University of Georgia

Abstract: Visualization is an integral component of statistical shape analysis, where the goal is to perform inference on shapes of objects. When interested in identifying shape variation, one typically performs principal component analysis (PCA) to decompose total variation into orthogonal directions of variation. In many cases, shapes observe multiple sources of variation; using PCA to visualize requires decomposition into several plots displaying each mode of variation, without the ability to understand how these components work together. I propose a parametric construction (according to a model-based bootstrap) for a high density region for the elastic shape mean, with the goal of also producing a succinct visual summary of this region. The use of elastic shape representations allows for optimal matching of shape features, yielding more appropriate estimation of shape variation than some other approaches. Discussion of visualization issues is included. The proposed region is estimated for simulated data, as well common shapes from the well-known MPEG-7 dataset.

Geometric methods in human movement analysis
Pavan Turaga, Arizona State University

Abstract: In this talk, we will discuss problems involving human movement analysis, where we motivate the need for mathematical approaches rooted in differential geometry and topology. We discuss recent work in non-linear dynamical analysis as applied to human activity modeling where we propose to characterize dynamical attractors via their geometric and topological properties. We demonstrate results on activity recognition and scene analysis problems where these methods provide a well-grounded general purpose

framework. These methods are shown to work well in a variety of different dynamical time-series databases with minimal changes to the underlying framework. We also present preliminary interdisciplinary work and emerging opportunities at the intersection of geometric computing and real-time interactive feedback systems for mobility and balance-related disorders.

Can AI discover the drugs of the future?
Guowei Wei, Michigan State University

Abstract: The dominant win of Google's Alphafold in the latest Critical Assessment of Structure Prediction (CASP) competition has ushered a new era of scientific discovery. Researchers are excited about what the future may hold for drug discovery. Artificial intelligence (AI) might make new drug discovery significantly faster and cheaper. This could be particularly beneficial to patients with rare medical ailments, for whom drug discovery is currently not profitable, or for those whose medical ailments currently cannot be effectively treated with drugs, such as Alzheimer's disease. However, drug design is much more complex than protein folding prediction. Due to the structural complexity of protein-drug interactions, the high dimensionality of drug candidates' chemical space, and the involved molecular simulation and deep learning (DL), even all the world's computers put together do not have enough power to design drug automatically. In my lab, we tackle these challenges mathematically. Our work focuses on reducing the geometric complexity and degrees of freedom of protein-drug complexes for AI and DL, such as generative adversarial neural networks (GANs). We have introduced differential geometry, algebraic topology, and graph theory to obtain high-level abstractions of protein-drug complexes and thus enable GANs to handle excessively large datasets in drug discovery. Our mathematical AI has made us a top competitor in D3R Grand Challenges, a worldwide competition series in computer-aided drug design and discovery in the past three years.

Risk assessments on eradication of aquatic invasive species
Yanyu Xiao, University of Cincinnati

Abstract: In the work, we evaluate the performance of machine learning approaches for predicting successful eradication of aquatic invasive species and assess the extent to which eradication of an invasive species depends on the various specific ecological features of the target ecosystem and/or features that characterize the planned intervention.

Finite element methods with discontinuous approximations
Xiu Ye, University of Arkansas at Little Rock

Abstract: In this presentation, different finite element methods with discontinuous approximations will be discussed including IPDG, HDG and specially WG finite element methods as well as the relations between them. In addition, new stabilizer free discontinuous finite element methods will be introduced.

Poster 1

Detecting Imprinting and Maternal Effects Using Monte Carlo Expectation Maximization Algorithm

Pooya Aavani
Department of Biological Sciences, Texas Tech University

Abstract: Genomic imprinting and maternal effects are both important epigenetic factors that are involved in many complex human diseases, including Prader-Willi and Angelman syndromes, and childhood cancers. Genomic imprinting (maternal or paternal) is an effect of an epigenetic process involving methylation and histone modifications in order to silence the expression of a gene inherited from a particular parent without altering the genetic sequence. A maternal effect, on the other hand, refers to a situation where the phenotype of an individual is influenced by the genotype of the mother regardless of one's own genotype. Though genomic imprinting and maternal effects arise from two different underlying epigenetic mechanisms, they can produce the same parent-of-origin patterns of phenotypic variation. As such, it is necessary to distinguish and study these two confounding effects together to avoid false positives and/or false negatives.

Numerous statistical methods have been developed to explore genomic imprinting and maternal effects. However, most of them either only model one of these two confounded epigenetic effects, or make strong yet unrealistic assumptions about the population to avoid over-parameterization, such as mating symmetry. A recent Likelihood inference method for detecting Imprinting and Maternal Effects (LIME) based on case-control family data does not require those assumptions. LIME uses only part of the full likelihood—partial likelihood—by exploiting the fact that the part of the likelihood containing the parameters of interest can be separated from that containing the nuisance parameters. It thus alleviates the need to make typically unrealistic assumptions and thus leads to a robust procedure with potentially greater power. On the other hand, because LIME obtains parameter estimation by maximizing partial likelihood, it is interesting to compare its efficiency with full likelihood maximizer, which has never been studied before.

To overcome the difficulty in over-parameterization when using full likelihood, in this study we propose a Monte Carlo Expectation Maximization (MCEM) method to detect imprinting and maternal effects jointly. The Expectation-Maximization (EM) algorithm is a way to find maximum-likelihood estimates for model parameters when one's data has unobserved latent variables. In the full likelihood based on case-control family data in detecting the two epigenetic effects, those unknown mating type probabilities, the nuisance parameters, can be considered as latent variables. However, ordinary EM algorithm cannot work here, because the E step consists of non-elementary integrations that cannot be solved algebraically. Therefore, we used Monte Carlo method to numerically approximate the expectation function value. Our results and simulations show that MCEM algorithm is promising to estimate the epigenetic effects properly.

Poster 2

A P0 - P0 WEAK GALERKIN FINITE ELEMENT METHOD FOR SOLVING SINGULARLY PERTURBED REACTION-DIFFUSION PROBLEMS

Ahmed Al-Taweel, Saqib Hussain, Xiaoshen Wang, and Brian Jones
Department of Mathematics and Statistics, University of Arkansas at Little Rock

Abstract: This work investigates the lowest-order weak Galerkin finite element (WGFE) method for solving reaction-diffusion equations with singular perturbations in two and three space dimensions. The system of linear equations for the new scheme is positive definite, and one might readily get the well-posedness of the system. Our numerical experiments confirmed our error analysis that our WGFE method of the lowest order could deliver numerical approximations of the order O(h1=2) and O(h) in H1 and L2 norms, respectively.

Poster 3

Adaptive finite element methods on polygonal meshes

Shuhao Cao
Department of Mathematics, UC Irvine

Abstract: Both residual-based and equilibrated flux-based posteriori error analyses for various discretizations applied to a model diffusion interface problem are presented. An h-refinement strategy for polygonal meshes and hanging node automation are also discussed.This is a joint work with Long Chen.

Poster 4

Stoichiometric Knife-Edge Model on Discrete Time Scale

Ming Chen1, Lale Asik2, Angela Peace2
1School of Science, Dalian Maritime University, Dalian, Liaoning, China
2Department of Mathematics and Statistics, Texas Tech University, Lubbock, USA

Abstract: Ecological stoichiometry is the study of the balance of multiple elements in ecological interactions and processes. Modeling under this framework enables us to investigate the effect nutrient content on organisms whether the imbalance involves insufficient or excess nutrient content. This phenomenon is called the stoichiometric knife-edge". In this study, a discrete-time predator prey model that captures this phenomenon is established and qualitatively analyzed. We systematically expound the similarities and differences between our discrete model and the corresponding continuous analog. Theoretical and numerical analyses show that while the discrete and continuous models share many properties differences also exist. Under certain parameter sets, the models exhibit qualitatively different dynamics. While the continuous model shows limit cycles, Hopf bifurcations, and saddle-node bifurcations, the discrete-time model exhibits richer dynamical behaviors, such as chaos. By comparison the dynamics of the continues and discrete model, we can conclude that stoichiometric effects of low food quality on grazers are robust to the discretization of time. This study can possibly serve as an example for pointing to the importance of time scale in ecological modeling.

Poster 5

Bayesian two-sample mean hypothesis testing in high dimensions

Fan Chen
Department of Mathematics & Statistics, Texas Tech University

Abstract: A common problem in modern genetic research is comparing the mean vectors of two population means in high dimensional settings, that is, in ‘large-p-small-n’ settings. However, the classical Hotelling’s test cannot be applied since the sample covariance matrix is singular. To circumvent these limitations, we proposed a Bayesian two-sample test for the equality of population means in high dimension by clustering. Our proposed Bayesian factor is invariant under linear transformations of the marginal distributions and has closed form expression. We showed by simulation that our proposed Bayesian procedure has higher power than competing tests in the literature.

Poster 6

Gibbs-Importance Sampling Algorithm in Bayesian Quantile Regression

Mai Dao
Department of Mathematics & Statistics, Texas Tech University

Abstract: Bayesian quantile regression has gained great interest in both theoretical studies and practical applications due its versatile ability to estimate quantiles of the distribution of a response variable conditional on a set of explanatory variables. In this project, we extend the Zellner's prior to allow for a conditional conjugate prior with the likelihood-based approach of the working asymmetric Laplace distribution. The Gibbs - Importance sampling method is employed to both estimate the parameters of interest and perform variable

selection. The proposed method is compared with other Bayesian and frequentist procedures with both simulation and real data applications.

Poster 7

Assessing the combined impact of interventions on HIV and syphilis epidemic among gay, bisexual and other men who have sex with men in British Columbia: a co-interaction model

Jummy Funke David
Department of Mathematics, University of British Columbia

Abstract: We developed a mathematical transmission model to assess the impact of PrEP on HIV and syphilis infection and how the combination of testing and treating syphilis, HIV TasP, condoms and PrEP could eliminate both HIV and syphilis epidemic among gbMSM in BC over the next ten years. In addition, the World's Health Organization threshold for HIV elimination (one new case of HIV per 1000 susceptible gbMSM) and the control reproduction number (Rc) were examined. Different sensitivity analyses were performed and all aspect of the model were implemented in python.

Poster 8

Comparative analysis of phenomenological growth models applied to epidemic outbreaks.

Leidy Yissedt Lara Díaz
Universidad de Concepción- Chile

Abstract: Phenomenological models are particularly useful for characterizing epidemic trajectories because they often offer a simple mathematical form defined through ordinary differential equations (ODEs) that in many cases can be solved explicitly. Such models avoid the description of biological mechanisms that may be difficult to identify, are based on a small number of model parameters that can be calibrated easily and can be utilized for efficient and rapid forecasts with quantified uncertainty. These advantages motivate an in-depth examination of 37 data sets of epidemic outbreaks, with the aim to identify for each case the best suited model to describe epidemiological growth. Four parametric ODE-based models are chosen for study, namely the logistic and Gompertz model with their respective generalizations that in each case consists in elevating the cumulative incidence function to a power $p\in [0,1]$. This parameter within the generalized models provides a criterion on the early growth behavior of the epidemic between constant incidence for $p=0$, sub exponential growth for $0<p<1$ and exponential growth for $p=1$. Our systematic comparison of a number of epidemic outbreaks using phenomenological growth models indicates that the GLM model outperformed the other models in describing the great majority of the epidemic trajectories. In contrast, the errors of the GoM and GGoM models stay fairly close to each other and the contribution of the adjustment of p remains subtle in some cases. More generally, we also discuss how this methodology could be extended to assess the ``distance'' between models irrespective of their complexity.

Poster 9

The Number of Pollinators Really Drive The Stability Of Stoichiometric Plant-Pollinator-Herbivore Models

Dilini Fonseka and Angela Peace
Department of Mathematics and Statistics, Texas Tech University

Abstract Plant-pollinator interactions play an important role in the maintenance of the balance of nature. All organisms living in the environment are composed of different ratios of chemical elements. By considering the balance of essential chemical elements in nature, we can formulate mathematical models to study their role in the dynamics of the system as well as nature. We formulate and analyze stoichiometric plant-pollinator and stoichiometric herbivore-plant- pollinator models. Our models include three dimensional and four-dimensional systems of ordinary differential equations to represent the plant, pollinator, herbivore populations, as well as the varying nutrient levels of the plant. We analyze the dynamics of the systems such as non- negativeness and

boundedness of solutions, as well as the existence and stability of boundary equilibria. We perform a bifurcation analysis of the models and also a parameter sensitivity analysis of stoichiometric plant-pollinator model using Latin hypercube sampling and partial rank correlation coefficient technique. LHS show that the search rate and the carrying capacity of pollinators are most important parameters to the stoichiometric plant-pollinator model. Bifurcation analysis shows the existence of critical thresholds of number of pollinators for plants to survive and for herbivores to die.

Poster 10

Generative network complex (GNC) for drug discovery

Christopher Grow, Kaifu Gao, Duc Duy Nguyen, and Guo-Wei Wei
Department of Mathematics, Michigan State University

Abstract: A generative network complex (GNC) is proposed to design novel compounds, predict their physical and chemical properties, and select potential drug candidates that fulfill various druggable criteria such as binding affinity, solubility, partition coefficient, etc. Currently, even the largest databases of chemical compounds remain small compared to space spanned by all energetically stable stoichiometric combinations of electronic states, atoms, and topologies in molecules. Therefore, it is still a challenging task to identify a variety of novel compounds. Recently, many generative models have been developed to create new compounds. In this work, we combine a variational autoencoder and various deep neural network predictors to generate new compounds and predict their drug properties. Specifically, 2.08 million and 2.8 million novel compounds are generated respectively for Cathepsin S and BACE targets. These generated compounds are quite different from the seeds and cover a larger chemical space. Additionally, their drug properties were predicted by deep learning predictors. Finally, drug candidates that fulfill the threshold for druggable criteria are selected to construct their three-dimensional (3D) poses and further evaluate their chemical and physical properties. Performed on supercomputers, the whole process took less than one week. Therefore, our automatic GNC pipeline is very efficient for discovering new drug candidates.

Poster 11

Bayesian multi-domain learning for cancer subtype discovery

Ehsan Hajiramezanali
Department of Electrical & Computer Engineering, Texas A&M University

Abstract: Precision medicine aims for personalized prognosis and therapeutics by utilizing recent genome-scale high-throughput profiling techniques, including next-generation sequencing (NGS). However, translating NGS data faces several challenges. First, NGS count data are often overdispersed, requiring appropriate modeling. Second, compared to the number of involved molecules and system complexity, the number of available samples for studying complex disease, such as cancer, is often limited, especially considering disease heterogeneity. The key question is whether we may integrate available data from all different sources or domains to achieve reproducible disease prognosis based on NGS count data. In this paper, we develop a Bayesian Multi-Domain Learning (BMDL) model that derives domain-dependent latent representations of overdispersed count data based on hierarchical negative binomial factorization for accurate cancer subtyping even if the number of samples for a specific cancer type is small. Experimental results from both our simulated and NGS datasets from The Cancer Genome Atlas (TCGA) demonstrate the promising potential of BMDL for effective multi-domain learning without "negative transfer" effects often seen in existing multi-task learning and transfer learning methods.

Poster 12

A weak Galerkin harmonic finite element method for Laplace equation

Saqib Hussain
Department of Mathematics and Physics, Texas A&M International University

Abstract: In this article, a weak Galerkin finite element method for Laplace equation using harmonic polynomial space is proposed and analyzed. The idea of using P_k harmonic polynomial space instead of the full polynomial space P_k is to use a much smaller number of basis functions to achieve same accuracy when k >= 2. The optimal rate of convergence is derived in both H^1 and L^2 norms. Numerical experiments have been conducted to verify the theoretical error estimates. In addition, numerical comparisons of using P_2 harmonic polynomial space and using the standard P_2 polynomial space are presented.

Poster 13

An Empirical Comparison between Integer and Fractional Order SEIR Models of Measles

M. Islam, D. Medinab, A. Peace, T. Orabyb
Department of Mathematics and Statistics,Texas Tech University
School of Mathematical and Statistical Sciences, The University of Texas Rio Grande

Abstract: In this paper, we compare between the performance of systems of ordinary and (Caputo) fractional differential equations depicting the susceptible-exposed-infected-recovered (SEIR) models of diseases. In order to understand the origins of both approaches as mean-field approximations of fractional stochastic processes, we introduce the ordinary and fractional differential equations as approximations of some type of fractional nonlinear birth{death processes. Then, we examine validity of the two approaches against empirical courses of epidemics, we fit both of them to case counts of three measles epidemics that occurred during the pre-vaccination era in three different locations. We found that while the fractional order differential equations SEIR model gives a slightly better fit to some of the data, the ordinary differential equations SEIR model performs better overall.

Poster 14

Simulation of mean first passage time on embedded two dimensional domains

Sarafa Iyaniwura
Department of Mathematics, University of British Columbia

Abstract: We compute mean first passage time for a Brownian particle to escape a bounded 2-D domain containing absorbing traps, using the closest point method. These computations were done for both regular (unit disk) and irregular (star-shaped) domains with either stationary or moving traps. In addition, our numerical results are verified with analytical result obtained using the technique of strong localized perturbation. Furthermore, we find the optimal location of traps that minimizes the average mean first passage time. The mean first passage time approach has been used to study several problems in biological sciences, some of which include the time it takes for an immune system to respond to a new infection, the time it takes a surface-bound protein to arrive at a signalling region on a cell membrane, among others.

Poster 15

Extreme Dynamical complexity in Recurrent Neural Networks

Ian D. Jordan
Stony Brook University

Abstract: Gated recurrent units (GRUs) are specialized memory elements for building recurrent neural networks. Despite their incredible success in natural language, speech, and video processing, little is understood about the specific dynamics representable in a GRU network, along with the constraints these dynamics impose when generalizing a specific task. As a result, it is difficult to know a priori how successful a GRU network will perform on a given task. Using a continuous time analysis, we gain intuition on the inner workings of GRU networks. We restrict our presentation to low dimensions to allow for a comprehensive visualization. We found a surprisingly rich repertoire of dynamical features that includes stable limit cycles (nonlinear oscillations), multi-stable dynamics with various topologies, and homoclinic orbits. We contextualize the usefulness of the different kinds of dynamics and experimentally test their existence.

Poster 16

Audio Classification and Cover Song Identification

Nhat Le
Department of Computer Science**,** Texas Tech University

Abstract: Recently, sound classification in general and particularly cover song identification not only have numerous applications in field of digital signal processing or speech recognition but also are set of important and challenging tasks. Due to the high dimensionality of the audio data, features extraction is always required. In this work, we used size 13 of Mel-Frequency Cepstral Coefficients (MFCCs) which is one of the most popular feature extraction techniques used in speech recognition. The 39-elements feature vector for each sound was the concatenation of three things including mean, standard deviation and the mean first order difference between the successive feature frames. After that, t-SNE is used to reduce the dimensionality of the N x 39 features matrix to N x 2 (where N is the number of sound samples). This allows us to use the resulting 2d coordinates to visualize the feature of each sound sample, we observed that the similar sounds tend to group together. Furthermore, in this project, related to music information retrieval community, we also propose a novel method to identify cover song by representing the self-similarity matrix (SSM) of each song using MFCC feature extraction and applying the Smith Waterman algorithm on the binary cross-similarity matrix to score a match between the two songs.

Poster 17

Elastic Alignment of fMRI Signals

David Lee
Department of Neurology, UCLA

Abstract: We present a shape matching approach for functional magnetic resonance imaging (fMRI) signals. We use ideas from differential geometry and functional data analysis to define a functional representation for fMRI signals. The space of fMRI functions is then equipped with a reparameterization invariant Riemannian metric that enables elastic alignment of both amplitude and phase of the fMRI time courses as well as their power spectral densities. Our approach constructs an unbiased average of the signals, and warps the signals to the mean. As the warping is diffeomorphic, nonlinear and allows large deformations of time series if required, we term this approach as elastic functional alignment. We demonstrate application of this idea to resting state and task based fMRI of patients with major depressive disorder.

Poster 18

Impact of Discrete-time Malaria Transmission Model with constant releases of sterile mosquitoes

Yang Li
University of Cincinnati

Abstract: In this study, we first formulate a baseline discrete-time mathematical model for malaria transmission where the survival function of mosquitoes is of Beverton-Holt type. We then introduce sterile mosquitoes to the baseline model to explore the transmission dynamics with sterile mosquitoes. We derive formulas for the reproductive number R0 of infection and determine the existence and uniqueness of endemic fixed points as well, for the models with or without sterile mosquitoes. We then study the impact of the releases of sterile mosquitoes on the disease transmissions by investigating the effects of varying the release rates of the sterile mosquitoes. We use a numerical example to illustrate our results for all cases and finally give brief discussions of our findings.

Poster 19

Discrete-time Disease Model with Population Motion under the Kolmogorov Equation View and Application

Ye Li
Department of Mathematics & Statistics, Texas Tech University

Abstract: We introduce the Susceptible-Infected-Removed (SIR) model and the Susceptible-Exposed-Infected-Removed (SEIR) model coupled with a social mobility model (SMM). We discretize them by a Forward Euler Method, which can be viewed through a mean-field approximation from a discrete version. We calculate basic reproduction number R_0 using the next generation matrix method. Then we obtain hyperbolic forward Kolmogorov equations (high-dimensional PDEs) and show that its projected characteristics corresponding to these models coincide with population motivation. Finally, we use the Deep Galerkin Method (DGM) to solve the high order nonlinear PDEs. In this project, we can improve the global prediction of epidemics dynamics, which can provide suggestions on "how to control" epidemics.

Poster 20

A Phase Shift Deep Neural Network for High

Lizuo Liu
Southern Methodist University Department of Mathematics

Abstract: In this paper, we propose a phase shift deep neural network (PhaseDNN) which provides a wideband convergence in approximating high frequency solutions of wave equations. The PhaseDNN utilizes the fact that many DNN achieves convergence in the low frequency range first, thus, a series of moderately-sized of DNNs are constructed and trained in parallel for selected high frequencies. With the help of phase shifts in the frequency domain, implemented through a simple phase factor multiplication on the training data, each DNN will be trained to approximate the target function's higher frequency content over a specific range. Due to the phase shift, each DNN achieves the speed of convergence as in the low frequency range.

Poster 21

Temporal Transformer Networks: Joint Learning of Invariant and Discriminative Time Warping

Suhas Lohit
Electrical, Computer and Energy Engineering, Arizona State University

Abstract: Many time-series classification problems involve developing metrics that are invariant to temporal misalignment. In human activity analysis, temporal misalignment arises due to various reasons including differing initial phase, sensor sampling rates, and elastic time-warps due to subject-specific biomechanics. Past work in this area has only looked at reducing intra-class variability by elastic temporal alignment. In this paper, we propose a hybrid model-based and data-driven approach to learn warping functions that not just reduce intra-class variability, but also increase inter-class separation. We call this a temporal transformer network (TTN). TTN is an interpretable differentiable module, which can be easily integrated at the front end of a classification network. The module is capable of reducing intra-class variance by generating input-dependent warping functions which lead to rate-robust representations. At the same time, it increases inter-class variance by learning warping functions that are more discriminative. We show improvements over strong baselines in 3D action recognition on challenging datasets using the proposed framework. The improvements are especially pronounced when training sets are smaller.

Poster 22

Lowest Order Weak Galerkin for Stokes Flow on Polygonal Meshes

Nolisa Malluwawadu
Colorado State University Department of Mathematics

Abstract: This poster presents the lowest-order weak Galerkin (WG) finite element method for solving the Stokes equation on convex polygonal meshes. Constant vectors are used separately in element interiors and on edges to approximate fluid velocity, whereas constant scalars are used on elements to approximate the pressure. For the constant vector basis functions, their discrete weak gradients are established in a matrix space that is based on the CW_0 space [W. Chen and Y. Wang, Math. Comp., 86(2017), pp. 2053-2087], whereas their discrete weak divergences are calculated as elementwise constants. To circumvent the saddle-point problem, a reduced scheme for velocity is also established by using three types of basis functions for the discretely divergence-free subspace. A procedure for subsequent pressure recovery is developed. Error analysis along with numerical experiments on benchmarks are presented to demonstrate the accuracy and efficiency of the proposed new method.

Poster 23

Integration of deep learning and mathematics for drug discovery

Duc Nguyen, Zixuan Cang, Kaifu Gao, and Guo-Wei Wei
Department of Mathematics, Michigan State University

Abstract: Drug discovery is one of the most important tasks in life sciences and ultimately tests of our understanding of biological sciences and current methodologies. Although deep learning has had tremendous success in a wide variety of sciences, technologies, and industries, it encounters an enormous obstacle in drug design due to the intricate complexity of biomolecules structures and high dimensionality of biological datasets. We overcome these learning difficulties by constructing mathematical representations of diverse biomolecular data using differential geometry, algebraic topology, and graph theory. Our approaches have emerged as a top winner in D3R Grand Challenges, a worldwide annual competition series in computer-aided drug design, in the past few years.

Poster 24

EQSA: Earthquake Situational Analytics from Social Media

Huyen Nguyen and Tommy Dang
Department of Computer Science, Texas Tech University

Abstract: This work introduces EQSA, an interactive exploratory tool for earthquake situational analytics using social media. EQSA is designed to support users to characterize the condition across the area around the earthquake zone, regarding related events, resources to be allocated, and responses from the community. On the general level, changes in the volume of messages from chosen categories are presented, assisting users in conveying a general idea of the condition. More in-depth analysis is provided with topic evolution, community visualization, and location representation. EQSA is developed with intuitive, interactive features and multiple linked views, visualizing social media data, and supporting users to gain a comprehensive insight into the situation. In this work, we present the application of EQSA with the VAST Challenge 2019: Mini-Challenge 3 (MC3) dataset.

Poster 25

Testing for marginal independence between two categorical variables with multiple responses

Tong Nguyen
Texas Tech University

Abstract: Mind and body practices may be used to improve health and well-being or to help manage symptoms of health problems. The 2012 National Health Interview Survey collected information about Americans' top 3 commonly used modalities and whether using each modality is because of the recommendation from any doctor, family member, or friend. There are statistical methods to study the association between two multiple response categorical variables, i.e. modality selection and recommendation in the survey. However, due to the way the questions were designed in the survey, no information was collected about failure recommendation. No existing efficient statistical methods can be directly applied to the scenario. In this poster, we proposed a modified Pearson chi-square statistic to analyze the special data structure in the survey. Simulations were conducted to evaluate the proposed method.

Poster 26

A Severe Weather Index Based on the Historic National Oceanic and Atmospheric Administration (NOAA) Data

Thilini Mahanama
Texas Tech University

Abstract: Severe weather phenomena, such as tornadoes, floods, and wildfire pose risks to life and property, requires the intervention of authorities. One of the most visible consequences of changing climate is an increase in the intensity and frequency of extreme weather events. The relative strengths of the disasters grow far beyond the habitual seasonal maxima, often resulting in the subsequent increase in property losses. Thus, insurance policies should be modified to sustain regular catastrophic weather events. Our research work proposes a severe weather index for the total loss from natural disasters in the United States and studies how the insurance policies should be adapted to the increasingly volatile climate conditions. National Oceanic and Atmospheric Administration severe weather database is used for this study. It consists of millions of data over 74 types of severe weather events for every state within the United States from 1950 to 2018. To describe losses from extreme weather conditions, Empirical copulas have been used. Our work would help the assessment of risk attributed to severe weather events and improve communications between insurers and their clients.

Poster 27

A recurrent based machine learning approach for predicting learners' latent status in online learning

Bo Pei1, Zhuanzhuan Ma2, Wanli Xing1
1 College of Education, University of Florida
2 Department of Mathematics and Statistics, Texas Tech University,

Abstract: Student latent learning status is a significant indicator for their learning performances in online learning context. However, it is impossible for instructors to track every individual's learning status at each learning stage in online learning settings due to the large enrollments. In this paper, we proposed a new learning status prediction method based on the Logistic Regression inferred by Deep learning approaches. The model was applied on both demographical information and behavior information recorded by virtual learning environment (VLE). The final performance of the model was compared with some the baseline models such as Support Vector Machine, Decision Tree and Random Forest. The result indicates an excellent performance of the proposed model in terms of both identifying the early dropout students and predicting the final performances with an averaged f1-measure about 0.80 and Area under curve (AUC) around 0.80, respectively.

Poster 28

Assessing parameter identifiability in compartmental dynamic models using a computational approach: Application to infectious disease transmission models

Kimberlyn Roosa
Department of Population Health Sciences, School of Public Health, Georgia State University

Abstract: The successful application of mathematical models to guide public health interventions lies in the ability to reliably estimate model parameters and their corresponding uncertainty. Here, we present and illustrate a simple computational method for assessing parameter identifiability in compartmental epidemic models. We use a parametric bootstrap approach to generate simulated data of the incidence curve in order to derive the empirical distributions of the estimated parameters. These distributions are then used to quantify confidence intervals and relative biases of estimated parameters to assess structural parameter identifiability. To demonstrate this approach, we begin with a low-complexity SEIR model and work through examples of increasingly more complex compartmental models that correspond with real-world applications to pandemic influenza, Ebola, and Zika.

Poster 29

Nowcasting and short term forecasting the ongoing 2018-19 Ebola epidemic in the Democratic Republic of Congo (DRC)

Amna Tariq
School of Public Health, Division of Biostatistics and Epidemiology, Department of Population Health Sciences, Georgia State University

Abstract: The 10th and largest outbreak of the Ebola Virus Disease in the Democratic Republic of Congo (2018-19) occurring in an active conflict zone has spilled over to Uganda and is a global public health emergency. The spread of Ebola to Congolese urban centers along with deliberate attacks on the health care workers and health centers has hindered the epidemiological surveillance activities. In such circumstances of uncertainty, analyzing and forecasting the Ebola Virus Disease epidemic is challenged by the reporting of cases, specially the reporting delays (time lag between the date of symptom onset and the date of reporting). These reporting delays distort the epidemic incidence pattern misrepresenting estimates of the epidemic potential and outbreak trajectory. To assess the impact of reporting delays we conduct a real time assessment of the current dynamics of the ongoing Ebola outbreak in Democratic Republic of Congo using epidemiological data retrieved from the World Health Organization Situation Reports while correcting for the underreporting of cases. We utilize one of the nowcasting approaches that commonly employs modeling reporting delay distribution. Reporting delay adjustment is implemented using a nonparametric method that adapts survival analysis and life table techniques for use with right truncated data and employs estimated reverse time hazards to adjust for reporting delays. The national curve of crude incidence by week of symptoms onset is adjusted for the most recent time periods. Using these adjusted incidence curves, we also generate short-term forecasts of the epidemic using various phenomenological models in a statistical inference framework. Therefore, incorporating the information that we gain from real time nowcasting and forecasting of the sequential situation reports at the policy level can considerably improve the surveillance system in the Democratic Republic of Congo.

Poster 30

A De Novo Approach towards Protein folding via Reinforcement Learning (PfRL)

Niraj Verma, Mohamed Elsaied, Yunwen Tao, Eric Larson and Elfi Kraka
Southern Methodist University, Chemistry Department

Abstract: Proteins are one of the most important biological macromolecules, they serve for example as carriers of the message contained in the DNA. Proteins are composed of amino acids, which are arranged in a linear form and fold to form a unique 3D structure, which is imperative to their biological function. Often proteins are considered as the basic units of life. Detailed information about protein structure and function will lead to a better understanding of the complex processes that occur in a living organism.
As soon as a protein is synthesized as a linear sequence of amino acids, it folds in a matter of seconds to form a stable 3D structure, which is called the protein's native state. It is assumed that the information for the folding process is contained exclusively in the linear sequence of amino acids and that the protein in its native state has a minimum free energy.
In this work we approached the protein-folding problem via reinforcement learning. The reinforcement learning model predicts the protein sequence by adding one amino acid at a time (de novo) and matches the final generated protein with the original protein geometrically. Various molecular descriptors such as hydrophobicity, volume, alkalinity, hydrogen bonding possibility etc. for each amino acid were added to our model for better performance. The generated structure is then relaxed to form the final 3D protein structure. We will demonstrate how our work opens a new door towards approaching the protein-folding problem.

Poster 31

Block Preconditioning for Implicit Runge-Kutta Methods for Time-Dependent PDE Problems

Md Masud Rana*1, K. R. Long1, V. E. Howle1
1Department of Mathematics & Statistics, Texas Tech University, Lubbock, Texas, USA.

Abstract: Many important engineering and scientific problems require the solution of time-dependent PDE systems. Many of these systems have specific stability needs in order to compute realistic solutions such as needing A-stability or L-stability methods. For example, the reaction-diffusion equations of a spatiotemporal stoichiometric producer-grazer model is a stiff parabolic PDE, and benefits from an L-stable method. Certain classes of implicit Runge-Kutta (IRK) time-stepping methods, such as the Radau I and Radau II methods provide L-stability with higher-order accuracy, but one price of using an IRK method is needing to solve large linear systems at each time step. Suppose, for example, our PDE has been linearized and discretized with N degrees of freedom. Using an s-stage IRK method leads to an $sN \times sN$ linear system that must be solved at each time step. These systems are block- $(s \times s)$ systems, where each block is $N \times N$. Here, we investigate preconditioners for such systems.

Poster 32

A regularization approach for solving Poisson's equation with singular charge sources and diffuse interfaces

Siwen Wang
Department of Mathematics, University of Alabama

Abstract: In this poster, a simple Poisson's equation involving inhomogeneous media with a diffuse interface is studied. In particular, we will assume constant dielectric values inside each dielectric medium, while the dielectric function varies smoothly from one medium to another, through a narrow transition band. For Poisson's equation with singular charges and diffuse interfaces, a semi analytical method have already been proposed. The singular charges are treated analytically in this approach with diffuse interfaces. Nevertheless, this method is limited to simple geometries. Meanwhile, in regularization methods, a Poisson equation with the

same singular sources, the singular component can be analytically solved as Coulomb potentials or Green's functions. the other potential components can be accurately solved by finite difference or finite element methods. However, all existing regularization methods are designed for piecewise constant dielectric functions with sharp interfaces. It is unclear if regularization formulation could be established for diffuse interfaces. Thus, this poster presents the first regularization method in the literature that is able to handle diffuse interfaces. Besides a decomposition of potential function, the success of the new method lies in a decomposition of the inhomogeneous dielectric function. The singular charge sources containing in a complex domain can then be analytically treated. The details of the proposed regularization formulation and numerical validation of a simple example will be discussed.

Poster 33

Graph neural network for protein-ligand binding predictions.

Rui Wang
Department of Mathematics, Michigan State University

Abstract: Learning tasks often require dealing with graph data which contains rich information among graph nodes. Graph Neural Network (GNN) has become one of the most popular models for learning from graph inputs in various fields such as physics, chemistry, biology and linguistics. Our work focused on protein-ligand binding problems by using flexibility-rigidity index (FRI) of protein-ligand complexes as graph inputs and training GNN hyper-parameters automatically. We employ datasets CASF-2007 and CASF-2016 to validate the correlation, robustness and reliability of our GNN model.

Poster 34

Detect differentially methylated regions using non-homogeneous hidden Markov model

Shan Xue
Department of Mathematics & Statistics, Texas Tech University

Abstract: DNA methylation is an important epigenetic mechanism that is essential for normal development in mammals. Identifying differential Methylation (DM) between two biological conditions can help to study genomic regions associated with diseases. Homogeneous Hidden Markov Model (HMM) has been used to identify DM CpG sites accounting for spatial correlation across CG sites. However, as the CpG sites are not uniformly distributed throughout the whole genome, spatial correlation can be different between CpG sites with varying distances. In this paper, we explore a non-homogeneous HMM to determine the DM CpG sites. By using a heterogeneous transition matrix, the method can accommodate the uneven spacing of CpG sites while modeling hyper-, hypo-, and equal- methylated statuses together. Simulations were conducted to evaluate the performance of the method.

Poster 35

NETWORK CLUSTERING WITH ENTROPY-BASED MONTE CARLO METHOD

Qiannan Zhai
Department of Mathematics & Statistics, Texas Tech University

Abstract: A network clustering algorithm is useful for understanding the structure of network data. Clustering methods usually group vertices based on a certain similarity/distance measure, such that vertices assigned in the same clusters are more closely related than vertices in different clusters. Therefore, network clustering can be formulated as an optimal problem to maximize the contrast of within-cluster and between-cluster closeness. However, most of the existing clustering methods are greedy algorithm. In this poster, we explore the usage of a cross-entropy Monte Carlo method for solving such a combinatorial problem. Instead of placing a discrete uniform distribution on all the potential solutions, an iterative importance sampling technique is utilized "to

slowly tighten the net" to place most distributional mass on the optimal community structure and its neighbors. Simulation studies were conducted to assess the performance of the method.

Poster 36

Fast Multipole Method for Helmholtz Equation in Layered Media

Wenzhong Zhang
Mathematics Department, Southern Methodist University

Abstract: In this work, a fast multipole method (FMM) is proposed to compute long-range interactions of wave sources embedded in 3-D layered media. The layered media Green's function for the Helmholtz equation is decomposed into a free space component and four types of reaction field components arising from wave reflections and transmissions through the layered media. The proposed algorithm is a combination of the classic FMM for the free space component and FMMs specifically designed for the four types of reaction components, made possible by new multipole expansions (MEs) and local expansions (LEs) as well as the multipole-to-local translation (M2L) operators for the reaction field components. Moreover, equivalent polarization source can be defined for each reaction component based on the convergence analysis of its ME. The FMMs for the reaction components, implemented with the target particles and equivalent polarization sources, are found to be much more efficient than the classic FMM for the free space component due to the fact that the equivalent polarization sources and the target particles are always separated by a material interface. As a result, the FMM algorithm developed for layered media has a similar computational cost as that for the free space. Numerical results validate the fast convergence of the MEs and the O(N) complexity of the FMM for interactions of low-frequency wave sources in 3-D layered media.