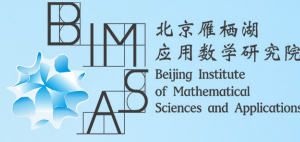




丘成桐数学科学中心
YAU MATHEMATICAL SCIENCES CENTER



数据分析及拓扑 国际数学会议

Data Analytics and Topology

February 17-21, 2025

Room A-110, TSIMF

组织者 Organizers

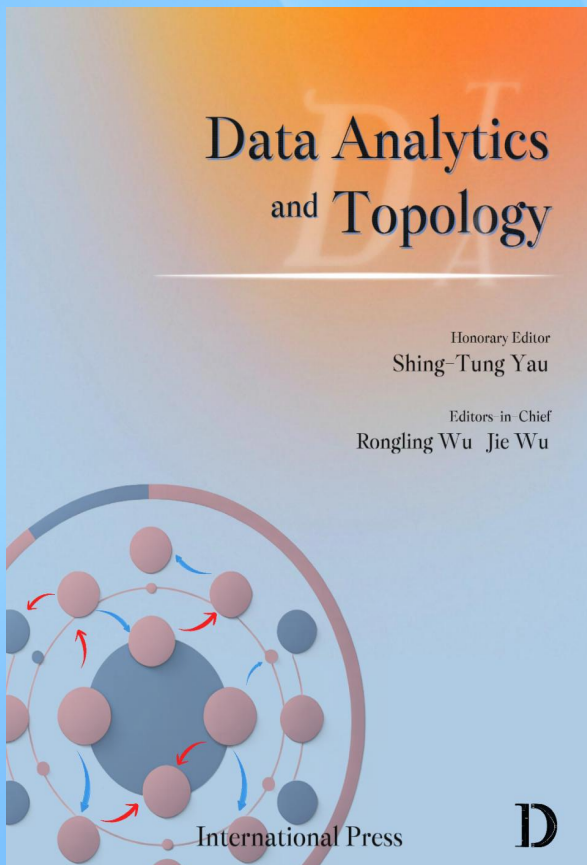
Rongling Wu(邬荣领), BIMSA & YMSC

Guowei Wei(魏国卫), Michigan State University

Jie Wu(吴杰), BIMSA

Data Analytics and Topology

Data Analytics and Topology is a peer-reviewed open access journal owned by Beijing Institute of Mathematical Sciences and Applications, under the sponsorship of the International Press inaugurated by Prof. Shing-Tung Yau.



Data Analytics and Topology has been created to meet the needs of the growing mathematical data science community as they create innovative concepts, theories, models, and tools that can best manipulate, interpret, and utilize big data. The aim of collecting big data is to uncover and extract fundamental principles and rules underlying scientific problems behind them. We anticipate that neither statistics nor even AI alone will suffice to achieve this goal in a sustainable way, given the heterogeneous, dynamical, interdependent, and high dimensionality of big data. Overcoming these challenges can be facilitated through the seamless integration of mathematics, particularly topology, with statistics.

Data Analytics and Topology endeavors to publish papers exploring the intersection between mathematics and statistics, especially as it applies to large-scale data sets. The journal welcomes contributions presenting new theories, methods, and interpretations of data within the realms of topological statistics or statistical topology. Its temporary website is available at <https://bimsa.net/journal/daat/>, where the scope of the journal is detailed.

We are currently seeking additional editors and associate editors for our journal. We need between 10 to 20 editors and approximately 20 associate editors. If you are interested in joining our team, please reach out to us at daatjournal@bimsa.cn.

We also invite you to submit your upcoming research papers or review articles to *Data Analytics and Topology*. We look forward to the opportunity of featuring your valuable work in our publication.

Contents

About the conference.....	1
Schedule	4
February 17, 2025, Monday	5
February 18, 2025, Tuesday	6
February 19, 2025, Wednesday.....	7
February 20, 2025, Thursday.....	8
February 21, 2025, Friday	9
Titles and Abstracts	10
Rongling Wu(邬荣领)	11
Carlile Lavor.....	11
Hongwei Lin(蔺宏伟)	11
Claudia Landi	11
Xiaopei Jiao(焦小沛)	12
Mingming Sun(孙明明)	12
Wenqi Pan(潘文祺)	13
Changjian Fa(法昌健)	13
Yangyang Bian(边洋洋)	13
Vassily Gorbounov.....	14
Yulia Gel.....	14
Theodore Papamarkou.....	15
Shailesh Lal	15
Congwei Song(宋丛威)	15
Grigory Papayanov	16
Rui Dong(董瑞)	16
Xiaomeng Xu (徐潇萌)	16
Mathieu Carriere	17
Wojciech Chacholski	17
Tamal Krishna Dey.....	17
Stephan Theriault	18
Wenting Zhao(赵雯婷)	18
Minh Quang Le.....	18
Haibao Duan(段海豹)	19

Fengling Li(李风玲).....	19
Quanming Yao(姚权铭).....	19
Yaqing Wang(王雅晴)	20
Yunguang Yue(岳云光).....	20
Mattew Burfitt.....	20
Zehua Lai(赖泽华).....	21
Xin Fu(付鑫).....	21
Huiying Gong(龚慧莹).....	21
Shiping Liu(刘世平).....	22
Thomas Wetere Tulu	22
Welcome to TSIMF	24
About Facilities.....	26
Registration.....	26
Guest Room	26
Library	27
Restaurant.....	27
Laundry	28
Gym	28
Swimming Pool.....	28
Free Shuttle Bus Service at TSIMF	28
Playground.....	28
Contact Information of Administration Staff.....	29

About the conference

The 1st International Conference on Data Analytics and Topology

数据分析及拓扑国际数学会议

Date

February 17-21, 2025

Venue

Room A-110, TSIMF

Organizers

Rongling Wu(邬荣领), Beijing Institute of Mathematical Sciences and Applications (BIMSA)&Yau Mathematical Sciences Center, Tsinghua University

Guowei Wei(魏国卫), Michigan State University

Jie Wu(吴杰), Beijing Institute of Mathematical Sciences and Applications (BIMSA)

Abstract

Data analysis has a wide range of applications in science and engineering. Topological computational analysis has become an important analytical tool. As an emerging branch of algebraic topology, persistent homology integrates certain geometric features into topological invariants, thereby bridging the gap between traditional topology and geometry. Computational topology, particularly topological deep learning, holds the promise of providing potentially revolutionary approaches to scientific research. For example, persistent homology has achieved significant success in the extraction, simplification, and drug discovery of complex macromolecular structures. The topological descriptions serve as an excellent foundation for machine learning applied to large-scale complex datasets and images. In computational science, algebraic topology is also applied in concurrent computing, distributed computing, sequential computation, and networks. In summary, computational topology has found many valuable applications in fields such as physics, chemistry, biology, materials science, fluid dynamics, computer graphics, control theory, geometric design, shape analysis, and computational science. This success has greatly advanced the development of related mathematical fields, including computational geometry, differential geometry, spectral geometry, geometric topology, geometric algebra, combinatorial mathematics, partial differential equations, optimization theory, inverse problems, and statistics.

The upcoming seminar on Data Analytics and Topology will bring together researchers from mathematics, physics, chemistry, biology, and computational science to explore new methods for connecting different disciplines and to promote the application of topology in both mathematics and various applied fields.

The primary goals of the proposed seminar are as follows:

- To promote the development of mathematical analysis and topological tools that can effectively utilize current computational capabilities to advance our understanding of the complexities of chemical, biological, and computational systems.
- To inspire a flow of information from "experiment to mathematics," similar to how quantum physics in the last century contributed to the establishment of heuristic new mathematics.
- To foster new connections, interactions, and collaborations between mathematicians and data scientists.

-
- To provide a platform for exchanging ideas and sharing topological results related to data science and computational science research.
 - To introduce graduate students, postdoctoral researchers, and junior faculty members to this field and related disciplines, helping to train the next generation of researchers in computational topology.
 - To enhance the participation of women, underrepresented minorities, and individuals with disabilities in research related to computational topology and its applications.

Description of the aim

Currently, a significant obstacle for mathematicians in the fields of computational topology and its applications is the lack of knowledge in data science and/or computational science. Conversely, for data scientists and computer scientists, the primary barrier is the absence of knowledge in the newly developed mathematical tools and topological techniques. The upcoming seminar aims to help bridge the gap between scientists and mathematicians and to foster collaboration between these two communities.

Schedule

Time&Date	Monday (February 17)	Tuesday (February 18)	Wednesday (February 19)	Thursday (February 20)	Friday (February 21)
7:30~8:30	Breakfast (60 minutes)				
Chair	Jie Wu	Rongling Wu	Jianzhong Pan	Fengchun Lei	Kefeng Liu
9:00-9:30	Rongling Wu(Online)	Vassily Gorbounov	Wojciech Chacholski	Haibao Duan	Shiping Liu
9:30-10:00	Carlile Lavor(Online)	Yulia Gel(Online)	Tamal Krishna Dey(Online)	Fengling Li	Thomas Wetere Tulu(Online)
10:00-10:30	Hongwei Lin	Theodore Papamarkou	Stephan Theriault	Quanming Yao	
10:30-11:00	Coffee Break(within 30 minutes)		Group Photo(10:40-11:00)	Coffee Break	
Chair	Stephan Theriault	Vassily Gorbounov	Wojciech Chacholski	Haibao Duan	
11:00-11:30	Claudia Landi	Shailesh Lal	Wenting Zhao(Online)	Yaqing Wang	
11:30-12:00	Xiaopei Jiao	Congwei Song	Minh Quang Le	Yunguang Yue	
12:00-14:00	Lunch Break (120 minutes)				
Chair	Hongwei Lin	Theodore Papamarkou	Free Discussion 13:30-17:00	Fengling Li	
14:00-14:30	Mingming Sun	Grigory Papayanov		Matthew Burfitt	
14:30-15:00	Wenqi Pan(Online)	Rui Dong		Zehua Lai(Online)	
15:00-15:30	Changjian Fa(Online)	Xiaomeng Xu		Xin Fu	
15:30-16:00	Coffee Break (within 30 minutes)			Coffee Break	
Chair	Claudia Landi	Shailesh Lal		Quanming Yao	
16:00-16:30	Yangyang Bian(Online)	Mathieu Carriere(Online)		Huiying Gong(Online)	
17:30-19:00	Dinner		Banquet 18:00-20:00	Dinner	

February 17, 2025, Monday

Time(日期)	Name(报告人)	Title(报告题目)
7:30~8:30	Breakfast (60 minutes)	
Chair	Jie Wu(吴杰)	
9:00-9:30	Rongling Wu(邬荣领) (Online)	A statistical framework to unify Yau-Yau filter and GLMY homology
9:30-10:00	Carlile Lavor(Online)	Molecular Geometry described by the Conformal Model of 3D Space
10:00-10:30	Hongwei Lin(蒯宏伟)	Gestalt Computational Model Based on Persistence Homology
10:30-11:00	Coffee Break(within 30 minutes)	
Chair	Stephan Theriault	
11:00-11:30	Claudia Landi	Algebraic invariants for the topological analysis of vector fields
11:30-12:00	Xiaopei Jiao(焦小沛)	Yau-Yau filtering theory and novel algorithms based on deep learning
12:00-14:00	Lunch Break (120 minutes)	
Chair	Hongwei Lin(蒯宏伟)	
14:00-14:30	Mingming Sun(孙明明)	A Universal Representation for Language Analysis and Processing
14:30-15:00	Wenqi Pan(潘文祺) (Online)	miRNA-miRNA interaction mediating leaf morphogenesis in <i>Populus euphratica</i>
15:00-15:30	Changjian Fa(法昌健) (Online)	Double-Sparsity Linear Regression for Variable Selection and Parameter Estimation in IdopNetwork
15:30-16:00	Coffee Break (within 30 minutes)	
Chair	Claudia Landi	
16:00-16:30	Yangyang Bian(边洋洋) (Online)	A mathematical strategy to map the genetic architecture of a single diplotype towards genome editing
17:30-19:00	Dinner	

February 18, 2025, Tuesday

Time(日期)	Name(报告人)	Title(报告题目)
7:30~8:30	Breakfast (60 minutes)	
Chair	Rongling Wu(邬荣领)	
9:00-9:30	Vassily Gorbounov	Geometry of Electric Networks and data analysis in phylogenetic
9:30-10:00	Yulia Gel(Online)	When Witnesses Defend: A Witness Graph Topological Layer for Adversarial Graph Learning
10:00-10:30	Theodore Papamarkou	Topological deep learning
10:30-11:00	Coffee Break(within 30 minutes)	
Chair	Vassily Gorbounov	
11:00-11:30	Shailesh Lal	Deep Learning Solutions of the Yang Baxter Equation
11:30-12:00	Congwei Song(宋丛威)	From the local mean to Transformer
12:00-14:00	Lunch Break (120 minutes)	
Chair	Theodore Papamarkou	
14:00-14:30	Grigory Papayanov	A criterion for 1-generatedness of A-infinity algebras and possible applications
14:30-15:00	Rui Dong(董瑞)	Fast Computation of the Up Persistent Laplacian for Pseudomanifolds and Cubical Complexes
15:00-15:30	Xiaomeng Xu(徐潇萌)	On ℓ_p -Vietoris-Rips Complexes
15:30-16:00	Coffee Break (within 30 minutes)	
Chair	Shailesh Lal	
16:00-16:30	Mathieu Carriere(Online)	Diffeomorphic interpolation for efficient persistence-based topological optimization
17:30-19:00	Dinner	

February 19, 2025, Wednesday

Time(日期)	Name(报告人)	Title(报告题目)
7:30~8:30	Breakfast (60 minutes)	
Chair	Jianzhong Pan(潘建中)	
9:00-9:30	Wojciech Chacholski	TDA in the Life Sciences
9:30-10:00	Tamal Krishna Dey(Online)	Gril: A 2-parameter Persistence Landscaping for Machine Learning
10:00-10:30	Stephan Theriault	Distributed computing and homotopy theory
10:30-11:00	Coffee Break & Group Photo(within 30 minutes)	
Chair	Wojciech Chacholski	
11:00-11:30	Wenting Zhao(赵雯婷) (Online)	Reconstruction of facial shape from genetic variants
11:30-12:00	Minh Quang Le	Homology of Transient Convection cycles over Multiplex Markov chains
12:00-14:00	Lunch Break (120 minutes)	
Free Discussion 13:30-17:00		
18:00-20:00	Banquet	

February 20, 2025, Thursday

Time(日期)	Name(报告人)	Title(报告题目)
7:30~8:30	Breakfast (60 minutes)	
Chair	Fengchun Lei(雷锋春)	
9:00-9:30	Haibao Duan(段海豹)	Make Schubert Calculus Calculable
9:30-10:00	Fengling Li(李风玲)	A three-variable transcendental invariant of planar knotoids
10:00-10:30	Quanming Yao(姚权铭)	Knowledge-aware Parsimony Learning from Deep Networks
10:30-11:00	Coffee Break(within 30 minutes)	
Chair	Haibao Duan(段海豹)	
11:00-11:30	Yaqing Wang(王雅晴)	Few-Shot Learning in Drug Discovery
11:30-12:00	Yunguang Yue(岳云光)	Path homology of directed topological spaces
12:00-14:00	Lunch Break (120 minutes)	
Chair	Fengling Li(李风玲)	
14:00-14:30	Matthew Burfitt	Inductive construction of path homology chains
14:30-15:00	Zehua Lai(赖泽华)(Online)	Grassmannian and Flag Optimization: Algorithms and NP-hardness
15:00-15:30	Xin Fu(付鑫)	Path homology of digraphs without multisquares
15:30-16:00	Coffee Break (within 30 minutes)	
Chair	Quanming Yao(姚权铭)	
16:00-16:30	Huiying Gong(龚慧莹) (Online)	Topological change of soil microbiota networks for forest resilience under global warming
17:30-19:00	Dinner	

February 21, 2025, Friday

Time(日期)	Name(报告人)	Title(报告题目)
7:30~8:30	Breakfast (60 minutes)	
Chair	Kefeng Liu(刘克峰)	
9:00-9:30	Shiping Liu(刘世平)	Curvature, perfect matching, and local symmetry of graphs
9:30-10:00	Thomas Wetere Tulu(Online)	Machine learning-based prediction of COVID-19 mortality using immunological and metabolic biomarkers
12:00-14:00	Lunch Break (120 minutes)	
17:30-19:00	Dinner	

Titles and Abstracts

A statistical framework to unify Yau-Yau filter and GLMY homology

Rongling Wu(邬荣领)

Beijing Institute of Mathematical Sciences and Applications (BIMSA)&Yau Mathematical Sciences Center, Tsinghua University

The Yau-Yau filter is a theory to find true states of internal workings within a complex stochastic system from noisy data, whereas GLMY homology aims to find topological features of digraphs. These two theories have increasingly emerged as powerful tools to find fundamental principles behind observed data, but they have been applied independently in their own fields. In this talk, I will present a unified theory to contextualize these two theories into a cohesive whole, demonstrating how this new theory can disentangle stochastic phenomena and find natural laws from them.

Molecular Geometry described by the Conformal Model of 3D Space

Carile Lavor

State University of Campinas

One of the fundamental problems in computational chemistry is the minimization of the potential energy of molecules. In order to improve algorithms used to solve this problem, we propose the conformal model of 3D space where the calculation of interatomic distances, as well as their derivatives, are carried out more efficiently.

Gestalt Computational Model Based on Persistence Homology

Hongwei Lin(蔺宏伟)

Zhejiang University

Widely employed in cognitive psychology, Gestalt theory elucidates basic principles in visual perception. However, the Gestalt principles are validated mainly by psychological experiments, lacking quantitative research supports and theoretical coherence. In this paper, we utilize persistent homology, a mathematical tool in computational topology, to develop a unified computational model for Gestalt principles, addressing the challenges of quantification and computation. On the one hand, the Gestalt computational model presents quantitative supports for Gestalt theory. On the other hand, it shows that the Gestalt principles can be uniformly calculated using persistent homology, thus developing a coherent theory for Gestalt principles in computation. Moreover, it is anticipated that the Gestalt computational model can serve as a significant computational model in the field of computational psychology, and help the understanding of human being visual perception.

Algebraic invariants for the topological analysis of vector fields

Claudia Landi

University of Modena and Reggio Emilia

In this talk, we discuss different ways of assigning a chain complex to a vector field, and discuss

connections to known invariants from topological data analysis. In the gradient-like case, a pipeline that produces a barcode from a weighted and based chain complex can be applied to the Morse chain complex of a vector field in both the smooth and discrete settings. This pipeline is proved to be locally stable and returns the classical persistence barcode in the special case of the gradient of a scalar field. Going beyond the gradient-like case, in order to assign a chain complex to any Morse-Smale vector field, we use the filtration of the underlying manifold by unstable manifolds and consider the spectral sequence in Čech homology associated with this filtration. We present a method to rearrange the algebraic information of the spectral sequence so as to obtain a canonical chain complex, whose homology agrees with the singular homology of the manifold and yields the expected Morse inequalities. This is based on joint work with Clemens Bannwart.

Yau-Yau filtering theory and novel algorithms based on deep learning

Xiaopei Jiao(焦小沛)

Beijing Institute of Mathematical Sciences and Applications (BIMSA)

The nonlinear filtering problem, which dates back to the 1600s, aims to infer reliable state estimates from stochastic measurements. The introduction of the Kalman filter in the 1960s revolutionized fields such as aerospace engineering and navigation. However, addressing the complexities of real-world systems required the development of nonlinear filters. Numerous suboptimal filtering algorithms have since been developed, leveraging Gaussian approximations and particle reconstruction techniques. Nevertheless, achieving optimal state estimation hinges on computing the conditional density, governed by the Duncan-Mortensen-Zakai (DMZ) equation introduced in the 1970s. In the 21st century, the Yau-Yau filter, innovatively proposed by Fields Medalist Professor Shing-Tung Yau and Professor Stephen S.-T. Yau, emerged as a groundbreaking tool for nonlinear filtering. The Yau-Yau filter remains a uniquely powerful method for effectively handling complex nonlinear systems, such as those involving cubic sensors. Its key advantages include: (1) real-time computation; (2) a rigorous mathematical foundation with proven convergence properties; (3) the capability to handle general nonlinear systems; and (4) an efficient structure that separates online and offline computations. Building on the Yau-Yau framework, we introduced the Extended Direct Method (EDM) to address more general infinite-dimensional systems compared to the traditional Direct Method. EDM is supported by rigorous existence and uniqueness analyses, and numerical results demonstrate that this explicit algorithm can achieve near-optimal accuracy comparable to spectral methods. Additionally, we developed the Deep Generalized Galerkin Method based on Physics-Informed Neural Networks (PINNs), which accelerates the offline computations of the Yau-Yau filter while preserving its high accuracy. Numerical simulations validate the efficiency and precision of these advancements, highlighting their potential for broader applications in nonlinear filtering.

A Universal Representation for Language Analysis and Processing

Mingming Sun(孙明明)

Beijing Institute of Mathematical Sciences and Applications (BIMSA)

Historically, linguistic analysis and natural language processing have taken divergent paths in their approaches to language, growing increasingly distinct over time: linguistic analysis aims for fine-grained analysis of language, while natural language processing leans toward accomplishing tasks end-to-end through statistical learning. We propose a novel unified language representation

framework that enables linguistic analysis and natural language processing to operate within a single architecture, making it possible to integrate the strengths of both approaches.

miRNA-miRNA interaction mediating leaf morphogenesis in *Populus euphratica*

Wenqi Pan(潘文祺)

Beijing Forestry University&Beijing Institute of Mathematical Sciences and Applications (BIMSA)

MicroRNAs (miRNAs) are small RNA molecules that play crucial regulatory roles in a variety of biological processes and are widely distributed in both animals and plants. While the study of interactions between individual miRNAs and their target genes is well-established, a systematic analysis of miRNA interactions within biological systems remains scarce. In this study, whole-genome miRNA expression data were collected from *Populus euphratica* samples exhibiting two leaf shapes from Inner Mongolia and four leaf shapes from Xinjiang, China. A bidirectional, weighted interaction network incorporating all miRNA information was constructed using IdopNetwork. The network was further analyzed by applying GLMY homology theory to capture the information flow between miRNAs and identify core miRNAs. The findings provide valuable insights into the molecular mechanisms underlying leaf shape variation driven by miRNA interactions, and offer theoretical and practical guidance for the development of salt- and drought-resistant varieties through personalized breeding.

Double-Sparsity Linear Regression for Variable Selection and Parameter Estimation in IdopNetwork

Changjian Fa(法昌健)

Beijing Forestry University&Beijing Institute of Mathematical Sciences and Applications (BIMSA)

IdopNetwork has demonstrated strong capabilities in reconstructing informative, dynamic, omnidirectional, and personalized networks across diverse data domains and application scenarios. In this talk, I will provide a brief introduction to IdopNetwork, highlighting areas where further optimization is possible. I will then introduce the double-sparsity linear regression model, which simultaneously controls the sparsity of variables (representing edges in the network) and the sparsity of basis functions (used to capture nonlinear relationships among variables). Finally, I will discuss how the double-sparsity linear regression model can be integrated into IdopNetwork to address the challenges outlined in the first part of the talk.

A mathematical strategy to map the genetic architecture of a single diplotype towards genome editing

Yangyang Bian(边洋洋)

Beijing Forestry University&Beijing Institute of Mathematical Sciences and Applications (BIMSA)

Quantitative genetics is essential for genetic dissection of complex traits, yet its population-based theory fails to comprehend genetic architecture owned by a single individual. Here, we develop a diplotype-based mathematical strategy to assemble all genome loci into omnigenic interactome networks. These networks can not only be used to dissect the genetic architecture of a

single individual, but also leverage dominance, epistasis, and pleiotropy to be more interpretable bidirectional, signed, and weighted interactions. While traditional approaches can only estimate genetic parameters at the population level, our strategy can model and characterize how each allele acts and interacts with every other allele throughout the whole genome for a single individual, thus facilitating its genome editing and genome engineering. We apply this strategy to analyze transcriptomic data of two diplotyped cultivars of a woody plant, interpreting the genetic mechanisms underlying this species' cold resistance and inter-organ communication. Individualized genetic findings can be readily transformed to precision breeding or precision medicine. Our strategy can be generalized to establish the foundation of individualized quantitative genetics, a new theory that can make genetics more transformable.

Geometry of Electric Networks and data analysis in phylogenetic

Vassily Gorbounov

National Research University HSE Russia

We will describe the recently discovered connection of the space of electrical networks and Lagrangian Grassmannians. It turns out that electrical networks naturally parametrize the positive part of these algebraic varieties. This potentially has applications in computational biology. Namely it is well known that the electrical resistance between the nodes of the network defines a metric on the space of the nodes. For planar networks this distance has a special property called the Kalmanson metric, which is well known in phylogenetic and optimization. Using the connection with Lagrangian Grassmannians mentioned above we will give a complete description of the resistance metric in the space of all Kalmanson metrics. We will discuss as a specific application the procedure of recovering a phylogenetic network from a given Kalmanson metric on the boundary nodes.

When Witnesses Defend: A Witness Graph Topological Layer for Adversarial Graph Learning

Yulia Gel

Virginia Tech University

Capitalizing on the intuitive premise that shape characteristics are more robust to perturbations, we bridge adversarial graph learning with the emerging tools from computational topology, namely, persistent homology representations of graphs.

We introduce the concept of witness complex to adversarial analysis on graphs, which allows us to focus only on the salient shape characteristics of graphs, yielded by the subset of the most essential nodes (i.e., landmarks), with minimal loss of topological information on the whole graph. The remaining nodes are then used as witnesses, governing which higher-order graph substructures are incorporated into the learning process. Armed with the witness mechanism, we design a Witness Graph Topological Layer (WGTL), which systematically integrates both local and global topological graph feature representations, the impact of which is, in turn, automatically controlled by the robust regularized topological loss. Given the attacker's budget, we derive the important stability guarantees of both local and global topology encodings and the associated robust topological loss. We illustrate the versatility and efficiency of WGTL by its integration with five GNNs and three existing non-topological defense mechanisms. Our extensive experiments across six datasets demonstrate that WGTL boosts the robustness of GNNs across a range of perturbations

and against a range of adversarial attacks, leading to relative gains of up to 18%. This is a joint work with Naheed Arafat, NUS, Yuzhou Chen, UCR, and Debabrota Basu, Inria.

Topological deep learning

Theodore Papamarkou
Zhejiang Normal University

Topological deep learning (TDL) has recently emerged as a field of deep learning that, among else, makes use of topology to model and analyze data and their topological properties via deep learning models. First, this talk first motivates TDL, explaining why TDL is important. Subsequently, it presents advances in TDL, particularly in the context of representation learning and message passing. The talk concludes by showcasing some applications of TDL.

Deep Learning Solutions of the Yang Baxter Equation

Shailesh Lal
Beijing Institute of Mathematical Sciences and Applications (BIMSA)

The Yang Baxter Equation is ubiquitous in Mathematical Physics with applications ranging from knot theory, spin systems, quantum field theory and string theory. We will show how deep learning is used to arrive at new solutions of the equation starting from very minimal assumptions. Time permitting, we will comment on how the framework we present extends to solving general systems of polynomial equations.

From the local mean to Transformer

Congwei Song(宋丛威)
Beijing Institute of Mathematical Sciences and Applications (BIMSA)

This report introduces a general machine learning framework, termed the localization method, which is rooted in the concept of the local mean and serves as the cornerstone for the self-attention mechanism in the Transformer architecture. The framework is rigorously defined through the establishment of the local model and localization trick, providing a strict and formal expression of its underlying principles. Furthermore, the report delves into the connections between the localization method and an array of other models, such as kernel methods, lazy learning, MeanShift, relaxation labeling, linear neighborhood propagation, and fuzzy inference. By examining these relationships, the report aims to illuminate the broader implications and potential applications of the localization method within the field of machine learning.

A criterion for 1-generatedness of A-infinity algebras and possible applications

Grigory Papayanov

Beijing Institute of Mathematical Sciences and Applications (BIMSA)

We provide a criterion for when a minimal A-infinity algebra is generated by its component H^1 . Among algebras, satisfying this criterion, there are cohomology A-infinity algebras of nilpotent Lie algebras and magnitude cohomology algebras of finite graphs. The latter provides a potentially computationally feasible algorithm for constructing all GLMY cohomology classes over a field.

Fast Computation of the Up Persistent Laplacian for Pseudomanifolds and Cubical Complexes

Rui Dong(董瑞)

Vrije Universiteit Amsterdam

We show that an orthogonal basis for the kernel of a non-branching matrix can be computed in quadratic time. Non-branching matrices are real matrices with entries in $\{-1, 0, 1\}$, where each row contains at most two non-zero entries. Such matrices naturally arise in the study of Laplacians of pseudomanifolds and cubical complexes. Building on this result, we show that the up persistent Laplacian can be computed in quadratic time for pairs of such spaces. Furthermore, we show that the up persistent Laplacian of q -non-branching simplicial complexes can be represented as the Laplacian of an associated hypergraph, thus providing a higher-dimensional generalization of the Kron reduction. Finally, we highlight the efficiency of our method on image data.

On ℓ_p -Vietoris-Rips Complexes

Xiaomeng Xu(徐潇萌)

Beijing Institute of Mathematical Sciences and Applications (BIMSA)

In this talk, I will introduce the notion of ℓ_p -Vietoris-Rips complexes for $1 \leq p \leq \infty$, which is a generalization of the ordinary Vietoris-Rips complexes. We will see from the example of a circle the differences of these complexes while p varying. I will also present several results parallel to the ordinary Vietoris-Rips complexes: (1) a stability theorem for the corresponding version of the persistent homology; (2) for a compact Riemannian manifold and a sufficiently small scale parameter, all the ℓ_p -Vietoris-Rips complexes are homotopy equivalent to the given manifold; (3) the ℓ_p -Vietoris-Rips complexes are invariant (up to homotopy) under taking the metric completion. At the end of the talk, I will introduce the simplicial version of this theory, and describe its relation with blurred magnitude homology.

Diffeomorphic interpolation for efficient persistence-based topological optimization

Mathieu Carriere

Centre Inria d'Université Côte d'Azur

Topological Data Analysis (TDA) provides a pipeline to extract quantitative topological descriptors from structured objects. This enables the definition of topological loss functions, which assert to what extent a given object exhibits some topological properties. These losses can then be used to perform topological optimization via gradient descent routines. While theoretically sounded, topological optimization faces an important challenge: gradients tend to be extremely sparse, in the sense that the loss function typically depends on only very few coordinates of the input object, yielding dramatically slow optimization schemes in practice. Focusing on the central case of topological optimization for point clouds, we propose in this work to overcome this limitation using diffeomorphic interpolation, turning sparse gradients into smooth vector fields defined on the whole space, with quantifiable Lipschitz constants. In particular, we show that our approach combines efficiently with subsampling techniques routinely used in TDA, as the diffeomorphism derived from the gradient computed on a subsample can be used to update the coordinates of the full input object, allowing us to perform topological optimization on point clouds at an unprecedented scale.

Joint work with Theo Lacombe and Marc Theveneau, based on https://papers.nips.cc/paper_files/paper/2024/hash/2feff80094b297bcfb42dbb01f34b875-Abstract-Conference.html

TDA in the Life Sciences

Wojciech Chacholski

KTH Royal Institute of Technology

The development of informative vector representations for text, languages, protein interactions, and other complex data has revolutionized how we analyze, organize, and extract insights, exemplified by advancements in large language models and generative AI. Similarly, creating effective vectorizations for geometry holds the promise of equally transformative and far-reaching impacts. This potential is already evident in the growing adoption of TDA methods, which uses homology to translate geometrical properties of data into meaningful vector representations. The aim of my talk is to present one such representation, developed by the TDA group at KTH, called stable rank. I will introduce various novel ways of using it to encode geometry, and provide several illustrative examples of how to apply them to find meaningful results in biological data, for example how to use expression data to identify a set of responders in a cancer therapy.

Gril: A 2-parameter Persistence Landscaping for Machine Learning

Tamal Krishna Dey

Purdue University

1-parameter persistent homology has been applied to enhance deep learning models, such as Graph Neural Networks (GNNs). To enrich the representations of topological features, we propose to study 2-parameter persistence modules induced by bi-filtration functions. In order to incorporate these representations into machine learning models, we introduce a novel vector representation called

Generalized Rank Invariant Landscape (Gril) for 2-parameter persistence modules. This vector representation is stable and can be easily integrated into machine learning models. We present an algorithm to compute Gril efficiently and test our methods on synthetic and benchmark graph datasets. Finally, we adapt Gril to ZZ-Gril operating on "Quasi Zigzag Bifiltrations" that enables us to handle time-varying data with "quasi zigzag persistence". Some test results on time-series data show the usability of ZZ-Gril.

Distributed computing and homotopy theory

Stephan Theriault

University of Southampton

A distributed computing system is a collection of inter-communicating components located on different computers that run as a single system. Algebraic topology has been successfully used to classify certain distributed computing systems. This talk will focus on how more advanced techniques in homotopy theory can be used to show the existence of solutions when the topology of the underlying network is more complex.

Reconstruction of facial shape from genetic variants

Wenting Zhao(赵雯婷)

Institute of Forensic Science of China (IFSC)

Variance of facial phenotype underlies the uniqueness of human individuals and also associated with hundreds of genetic syndromes. Understanding the genetic architecture that form the diversity of facial shape remains challenging partly because of the high dimensionality and complexity of craniofacial morphology. Different research fields have diverse strategies on facial morphological phenotyping. 2D landmarks and 3D point clouds are widely applied to measure facial shape in anthropology and genetics, and computer vision uses embedding for face recognition and face topology for animate 3D modeling. Based on the strong fitting capability of deep neural network, can we directly predict face embeddings through genetic variation data?

Homology of Transient Convection cycles over Multiplex Markov chains

Minh Quang Le

Ho Chi Minh City Open University

Studying multi-layer network is one of concerned topic in network theory; however it is less understood under the view of Markov chain or multiplex Markov chains (MMCs). In this article, Topological data analysis (TDA), specifically via its tool: persistent homology (PH), is utilized to study homological features of MMCs. PH is used here to detect and quantify convection cycles arising for the stationary distributions of MMCs. We apply our approach to investigate the effect on convection of various systems properties, including the coupling parameters of the PageRank algorithm, the optimum of transitive probability that give rise to Markov Decision Process (MDP) and the coupling strength for MMCs (a recent generalization whereby a set of MCs are coupled together by another set of MCs). We illustrate for these diverse applications how such parameters can act as homological regularizers of convection, which we summarize through homological

bifurcation diagrams. Our findings place the study of convective flows, both real and simulated, over Markov Decision Process (MDP) as a special case of multilayer networks and model of Alzheimer's patients on a stronger mathematical and computational footing, which has numerous potential applications throughout the sciences and engineering.

Make Schubert Calculus Calculable

Haibao Duan(段海豹)

Institute of Mathematics, Academy of Mathematics and Systems Sciences

In problem 15 Hilbert called for a rigorous foundation of Schubert's enumerative calculus, of which a long standing and challenging part is the Schubert problem of characteristics. In the course of securing a foundation for algebraic geometry, Van der Waerden and Andrei Weil attributed this problem to the intersection theory of flag manifolds.

This talk surveys the background, content, and solution of the 15th problem. Our main results are a unified formula for computing the characteristics, and a systematic description for the intersection theory of flag manifolds. We illustrate the effectiveness of the formula and the algorithm by explicit examples.

A three-variable transcendental invariant of planar knotoids

Fengling Li(李风玲)

Dalian University of Technology

As a generalization of the classical knots, knotoids are equivalence classes of immersions of the oriented unit interval in a surface. In recent years, a variety of invariants have been proposed for spherical and planar knotoids as extensions of invariants of classical and virtual knots. In this talk, we will introduce a three-variable transcendental invariant of planar knotoids and discuss some properties of it.

Knowledge-aware Parsimony Learning from Deep Networks

Quanming Yao(姚权铭)

Tsinghua University

The scaling law, which involves the brute-force expansion of training datasets and learnable parameters, has become a prevalent strategy for developing more robust learning models. However, due to bottlenecks in data, computation, and trust, the sustainability of the scaling law is a serious concern for the future of deep learning. In this paper, we address this issue by developing next-generation models in a parsimonious manner (i.e., achieving greater potential with simpler models). The key is to drive models using domain-specific knowledge, such as symbols, logic, and formulas, instead of relying on the scaling law. This approach allows us to build a framework that uses this knowledge as "building blocks" to achieve parsimony in model design, training, and interpretation. Empirical results show that our methods surpass those that typically follow the scaling law. We also demonstrate the application of our framework in AI for science, specifically in the problem of drug-drug interaction prediction. We hope our research can foster more diverse technical roadmaps in the

era of foundation models.

Few-Shot Learning in Drug Discovery

Yaqing Wang(王雅晴)

Beijing Institute of Mathematical Sciences and Applications (BIMSA)

In the current field of AI-assisted scientific research (AI for Science), particularly in drug discovery and biomedicine, we often face the challenge of scarce labeled data. Few-shot learning has become a key technology to address this challenge, as it can effectively leverage limited data for learning and prediction. In this report, I will introduce a series of machine learning algorithms developed specifically to improve data efficiency and prediction accuracy in AI for Science under data scarcity. I will discuss the application of few-shot learning techniques in molecular property prediction, reviewing existing technologies and presenting our proposed Property-Aware Relationship Network (PAR) (NeurIPS 2021, TPAMI 2024) and parameter-efficient Graph Neural Network Adapter (PACIA) (IJCAI 2024). PAR optimizes the relationship representations between molecules by introducing a property-aware molecular encoder and a dependency-query-based relational graph learning module, thereby improving prediction accuracy for various chemical properties. Meanwhile, PACIA enhances few-shot molecular property prediction performance by generating a small number of adaptive parameters to modulate the information propagation process in graph neural networks. Finally, I will share the vision of applying few-shot learning techniques in broader scientific research.

Path homology of directed topological spaces

Yunguang Yue(岳云光)

Shihezi University

In this talk, we will give a construction of path homology of directed topological spaces. Then, we will show that homology enjoys some desirable properties. For example, it is invariant under isomorphisms of directed spaces, and it is a powerful invariant used in classification of directed topological spaces.

Inductive construction of path homology chains

Matthew Burfitt

Beijing Institute of Mathematical Sciences and Applications (BIMSA)

Path homology plays a central role in digraph topology and GLMY theory more general. Unfortunately, the computation of the path homology of a digraph is a two-step process, and until now no complete description of even the underlying chain complex has appeared in the literature. In particular, our understanding of the path chains is the primary obstruction to the development of fast path homology algorithms, which in turn would enable the practicality of a wide range of applications to directed networks.

In this talk I will introduce an inductive method of constructing elements of the path homology chain modules from elements in the proceeding two dimensions. When the coefficient ring is a finite

field the inductive elements generate the path chains. Moreover, in low dimensions the inductive elements coincide with naturally occurring generating sets up to sign, making them excellent candidates to reduce to a basis.

Inductive elements provide a new concrete structure on the path chain complex that can be directly applied to understand path homology, under no restriction on the digraph. During the talk I will demonstrate how inductive elements can be used to construct a sequence of digraphs whose path Euler characteristic can differ arbitrarily depending on the choice of field coefficients. In particular, answering an open question posed by Fu and Ivanov.

Grassmannian and Flag Optimization: Algorithms and NP-hardness

Zehua Lai(赖泽华)

University of Texas at Austin

We will talk about two recent works about optimization on Grassmannian and flag manifolds. In the first half of the talk, we will explain the basics of Grassmannian optimization and flag optimization, and propose an efficient block coordinate descent algorithm on the flag manifolds. In the second half, we will show that unconstrained quadratic optimization over a Grassmannian is NP-hard. As an addendum we demonstrate the NP-hardness of unconstrained quadratic optimization over the Cartan manifold, i.e., the positive definite cone regarded as a Riemannian manifold, another popular example in manifold optimization.

Path homology of digraphs without multisquares

Xin Fu(付鑫)

Beijing Institute of Mathematical Sciences and Applications (BIMSA)

A digraph is said to have no multisquares if, for any pair of vertices x, y with $d(x, y) = 2$, there are at most two shortest paths from x to y . For such a digraph and a field F , we construct a basis of the vector space of path n chains for $n \geq 0$, generalizing the basis for $n = 3$, previously constructed by Grigor'yan. We consider the F -path Euler characteristic defined as the alternating sum of dimensions of path homology groups with coefficients in F . If the vector space of the path chains is a bounded chain complex, we can use these bases to compute the F -path Euler characteristic. We provide an explicit example of a digraph \mathcal{G} where the F -path Euler characteristic changes depending on whether the characteristic of F is two. This illustrates differences between GLMY theory and the homology theory of spaces, leading us to conclude that no topological space X can have homology $H^*(X; K)$ isomorphic to path homology $\text{PH}_*(\mathcal{G}; K)$ simultaneously for $K = \mathbb{Z}$ and $K = \mathbb{Z}/2\mathbb{Z}$. This is joint work with Sergei O. Ivanov.

Topological change of soil microbiota networks for forest resilience under global warming

Huiying Gong(龚慧莹)

Beijing Forestry University

Forest management by thinning can mitigate the detrimental impact of increasing drought

caused by global warming. Growing evidence shows that the soil microbiota can coordinate the dynamic relationship between forest functions and drought intensity, but how they function as a cohesive whole remains elusive. We outline a statistical topology model to chart the roadmap of how each microbe acts and interacts with every other microbe to shape the dynamic changes of microbial communities under forest management. To demonstrate its utility, we analyze a soil microbiota data collected from a two-way longitudinal factorial experiment involving three stand densities and three levels of rainfall over a growing season in artificial plantations of a forest tree – larch (*Larix kaempferi*). We reconstruct the most sophisticated soil microbiota networks that code maximally informative microbial interactions and trace their dynamic trajectories across time, space, and environmental signals. By integrating GLMY homology theory, we dissect the topological architecture of these so-called omnidirectional networks and identify key microbial interaction pathways that play a pivotal role in mediating the structure and function of soil microbial communities. The statistical topological model described provides a systems tool for studying how microbial community assembly alters its structure, function and evolution under climate change.

Curvature, perfect matching, and local symmetry of graphs

Shiping Liu(刘世平)
University of Science and technology of China

There is a long history on the study of diameter bounds of graphs. We will explain how curvature-related ideas come into the picture. We will explain how to obtain sharp diameter bounds via calculating discrete Ricci curvature. Often, calculating discrete Ricci curvature involves showing existence of certain local perfect matchings. We prove a weaker version of a conjecture proposed by Qiao, Park and Koolen on diameter bounds of amply regular graphs and make new progress on Terwilliger's conjecture on finiteness of amply regular graphs. Furthermore, we obtain a volume estimate of amply regular graphs which is sharp for hypercubes. This is based on joint works with Kaizhe Chen, Chunyang Hu and Heng Zhang.

Machine learning-based prediction of COVID-19 mortality using immunological and metabolic biomarkers

Thomas Wetere Tulu
Beijing Institute of Mathematical Sciences and Applications (BIMSA)

Background: COVID-19 has become a major global public health problem, despite prevention and efforts for the last few years. The daily number of COVID-19 cases rapidly increases, and the time and financial costs associated with testing procedure are burdensome.

Method: To overcome this, we aim to identify immunological and metabolic biomarkers to predict COVID-19 mortality using a machine learning model. We included inpatients from public hospitals between January 1, and September 30, 2023, who were diagnosed with COVID-19 using RT-PCR. We developed three machine learning models to predict the mortality of COVID-19 patients based on data in their electronic medical records. We performed statistical analysis to compare the trained machine learning models which are Deep Neural Networks (DNN), Random Forest Classifier (RF) and Support Vector Machine (SVM) using data from a cohort of 5,059 patients (median age = 46 years; 49.3% male) who had tested positive for COVID-19 based on electronic health records and data from 532,427 patients as controls.

Result: We identified the top 20 immunological and metabolic biomarkers that can accurately

predict the risk of mortality from COVID-19 with ROC-AUC of 0.98 (95% CI 0.96-0.98). Of the three models used, our result demonstrate that the random forest (RF) model achieved the most accurate prediction of mortality among COVID-19 patients with age, glomerular filtration, albumin, urea, procalcitonin, c-reactive protein, oxygen, bicarbonate, carbon dioxide, ferritin, glucose, erythrocytes, creatinine, lymphocytes, PH of blood and leukocytes among the most important biomarkers identified. A cohort from Kwong Wah Hospital (131 patients) was used for model validation with ROC-AUC of 0.90 (95% CI 0.84-0.92).

Conclusion: We recommend physicians closely monitor hematological, coagulation, cardiac, hepatic, renal and inflammatory factors for potential progression to severe conditions among COVID-19 patients.

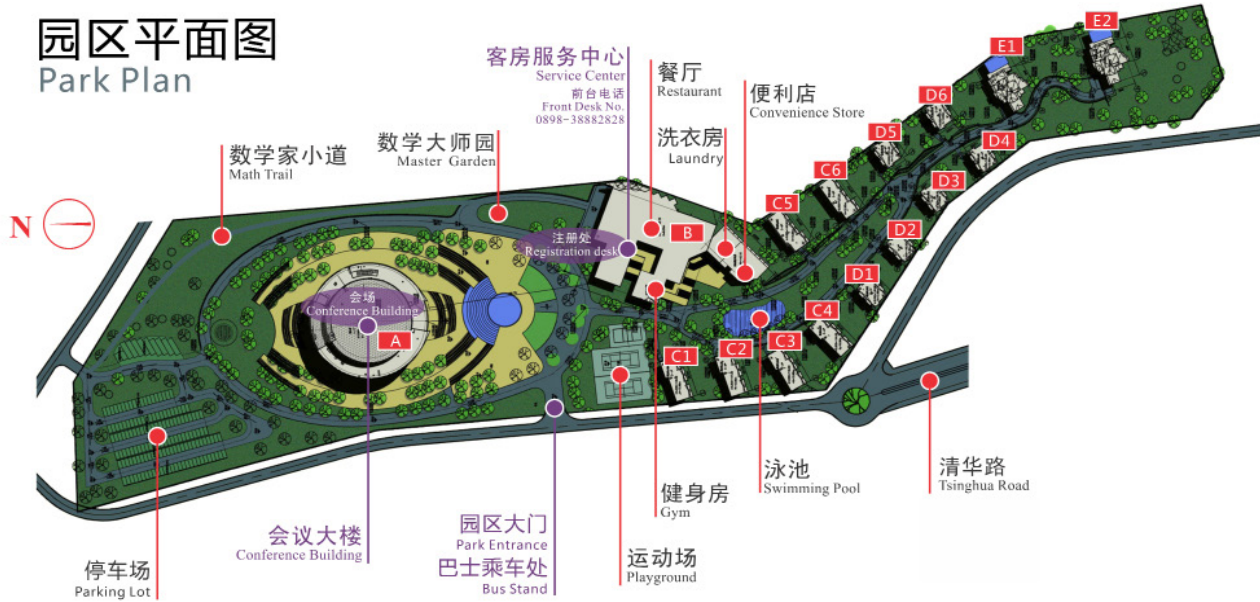
Welcome to TSIMF



The facilities of TSIMF are built on a 23-acre land surrounded by pristine environment at Phoenix Hill of Phoenix Township. The total square footage of all the facilities is over 29,000 square meter that includes state-of-the-art conference facilities (over 10,000 square meter) to hold many international workshops simultaneously, two reading rooms of library, a guest house (over 10,000 square meter) and the associated catering facilities, a large swimming pool, gym and sports court and other recreational facilities.

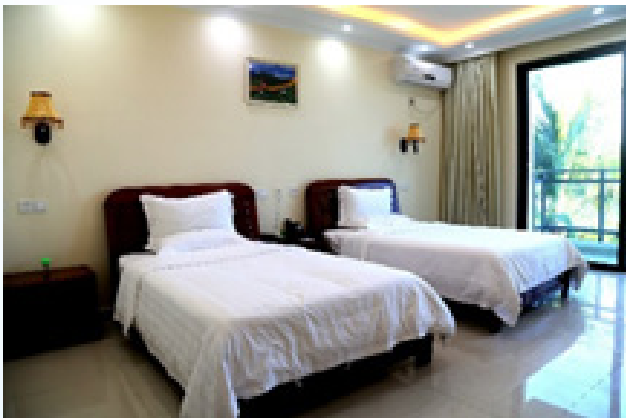
Management Center of Tsinghua Sanya International Forum is responsible for the construction, operation, management and service of TSIMF. The mission of TSIMF is to become a base for scientific innovations, and for nurturing of innovative human resource; through the interaction between leading mathematicians and core research groups in pure mathematics, applied mathematics, statistics, theoretical physics, applied physics, theoretical biology and other relating disciplines, TSIMF will provide a platform for exploring new directions, developing new methods, nurturing mathematical talents, and working to raise the level of mathematical research in China.

About Facilities



Registration

Conference booklets, room keys and name badges for all participants will be distributed at the front desk. Please take good care of your name badge. It is also your meal card and entrance ticket for all events.



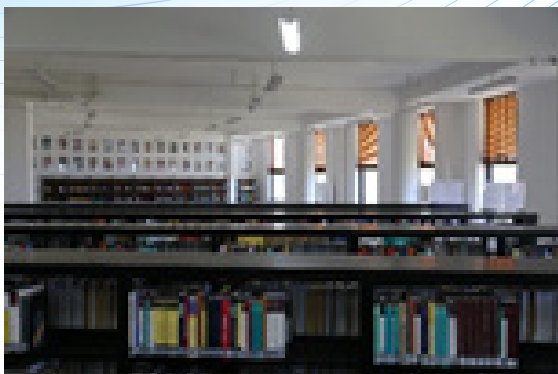
Guest Room

All the rooms are equipped with: free Wi-Fi (Password:tsimf123), TV, air conditioning and other utilities.

Family rooms are also equipped with kitchen and refrigerator.



Library



Opening Hours: 09:00am-22:00pm

TSIMF library is available during the conference and can be accessed by using your room card. There is no need to sign out books but we ask that you kindly return any borrowed books to the book cart in library before your departure.



In order to give readers a better understanding of the contributions made by the Fields Medalists, the library of Tsinghua Sanya International Mathematics Forum (TSIMF) instituted the Special Collection of Fields Medalists as permanent collection of the library to serve the mathematical researchers and readers.

So far, there are 271 books from 49 authors in the Special Collection of Fields Medalists of TSIMF library. They are on display in room A220. The participants are welcome to visit.



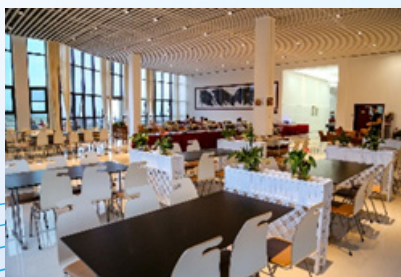
Restaurant

All the meals are provided in the restaurant (Building B1) according to the time schedule.

Breakfast 07:30-08:30

Lunch 12:00-13:30

Dinner 17:30-19:00



Laundry

Opening Hours: 24 hours

The self-service laundry room is located in the Building(B1).



Gym

Opening Hours: 24 hours

The gym is located in the Building 1 (B1), opposite to the reception hall. The gym provides various fitness equipment, as well as pool tables, tennis tables etc.



Playground

Playground is located on the east of the central gate. There you can play basketball, tennis and badminton. Meanwhile, you can borrow table tennis, basketball, tennis balls and badminton at the reception desk.

Swimming Pool

Please enter the pool during the open hours, swimming attire and swim caps are required, if you feel unwell while swimming, please stop swimming immediately and get out of the pool. The depth of the pool is 1.2M-1.8M.

Opening Hours: 13:00-14:00 18:00-21:00



Free Shuttle Bus Service at TSIMF

We provide free shuttle bus for participants and you are always welcome to take our shuttle bus, all you need to do is wave your hands to stop the bus.

Destinations: Conference Building, Reception Room, Restaurant, Swimming Pool, Hotel etc.



Contact Information of Administration Staff

Location of Conference Affairs Office: Room 104, Building A

Tel: 0086-898-38263896

Conference Affairs Manager: Shouxi He 何守喜

Tel:0086-186-8980-2225

Email: heshouxi@tsinghua.edu.cn

Location of Accommodation Affairs Office: Room 200, Building B1

Tel: 0086-898-38882828

Accommodation Manager: Ms. Li YE 叶莉

Tel: 0086-139-7679-8300

Email: yel@tsinghua.edu.cn

IT

Yuanhang Zhou 周远航

Tel: 0086-133-6898-0169

Email: 13368980169@163.com

*Reception duty hours: 7:00-23:00, chamber service please call: 0086-38882828 (exterior line) 80000 (internal line)

*Room maintainer night duty hours: 23:00-7:00, if you need maintenance services, please call: 0086-38263909 (exterior line) 30162 (internal line)

Director Assistant of TSIMF

Kai CUI 崔凯

Tel/Wechat: 0086- 136-1120-7077

Email :cuik@tsinghua.edu.cn

Director of TSIMF

Prof.Xuan GAO 高瑄

Tel: 0086-186-0893-0631

Email: gaoxuan@tsinghua.edu.cn

清华大学三亚国际论坛管理中心 | 三亚清华数学论坛管理中心
Tsinghua Sanya International Mathematics Forum (TSIMF)



0086-898-38882828

0086-898-38883896



0086-898-38883895



tsimf@tsinghua.edu.cn



572000



<https://www.tsimf.cn>



海南省三亚市天涯区清华路 100 号, 清华三亚国际数学论坛

No.100, Tsinghua Road, Tianya District, Sanya, Hainan, P. R. China.

