[Updated: 2025.07.01]

SJTU Health started in May 2015 by a group of enthusiasts, with the leadership and support from the **SJTU Faculty and Staff Union**. Over the years, more than 1,100 online/offline sports activities were organized, with over 50,000 participations. Thousands have benefited from this community effort, by sustained outdoor physical activities, in a self-disciplined way, or with the help of other community members.

Ways of operation

With a self-owned online platform named ***SJTU-Health***, regular online activities, each with daily or accumulative step count goals, are organized on a weekly or monthly basis. Participants are asked to join in small groups. Step count data, collected on smart phones, or on wearable devices such Gamin and Huawei sport watches, are synced to the platform on a regular basis. Fulfillment rates of both individuals and groups are updated and displayed in real-time.

Team leaders are provided with the contact information of their members, and are encouraged to form private chat groups such that up-to-date information on the activities can be distributed quickly and accurately. Even though competition is not a goal in most activities, a good consensus seems to have been reached that every team member should fulfill his/her own daily goal and at the same time help other members of the team to do so. Over time, stable communities are established.

The system also provides users with an incentive to reach step count goals of 5,000, 10,000, 15,000, and so on, by allowing them to "collect" one energy point on the platform, once each of these goals has been fulfilled. The collected energy points may be used for the exchange of small gifts in SJTU's gift shop. In 2023, users collected more than 110,000 energy points, half of which were used before Dec. 31st of 2023. For some activities, the organizer may provide rewards to individuals or teams, if they fulfilled the expected activity goal.

Outcomes

Two notable outcomes are observed. First, a large number of faculty and staff members of SJTU become active and sustained users of the ***SJTU-Health*** platform, most of whom developed an active and self-disciplined lifestyle. Increased and sustained levels of everyday physical activities are observed for many users.

Second, a strong consensus is built in the community, that an active lifestyle is beneficial to everyone, regardless of one's age, gender, occupation, and health condition. Communal support is important and necessary for some people to sustain their level of physical activity. As an apparent proof of the community strength, the "SJTU-team" (of walkers), with over 200 faculty and staff members, championed in a nationwide competition **seven** years in a row, every time with a close to 100% fulfillment rate.

Publications

Our efforts in developing the community, as well as the ***SJTU-Health*** online platform, have also led to some interesting scientific publications. These publications provide insights on data acquisition and modeling, and the development of community sports organizations (CSOs).

Non-profit sports community development and network analysis

Yu, J., Ding, M., Sun, W., *Dynamic-n-Static Multiplex Graph Representation Learning for Improved Link Prediction, submitted, June 2025*

Social systems often involve multiple types of relations, each exhibiting distinct temporal characteristics. Such systems can be modeled as temporal multiplex graphs in which each graph layer represents one type of relation. In this paper, we study the link prediction problem in a particular type of multiplex graph, Dynamic-n-Static Multiplex Graphs, that have both dynamic layers and static ones. We propose DS-MGN, a representation learning model for dynamic-n-static multiplex graphs. The cross-layer neighbor encoding (CLNE) scheme in DS-MGN can capture the topological structure of static layers, enabling the model to generate better representations by leveraging information from the multiple layers. The CLNE scheme can be seamlessly integrated into other models with similar neural network architectures, offering substantial performance improvements. Due to the challenges in data collection, existing datasets are generally limited to relatively simple graphs, highlighting the need for more complex and comprehensive graph datasets. In this work, we present a high-quality dynamic-n-static multiplex graph dataset, which includes a dynamic sports social network as well as a static network representing the long-term stable colleague relationships between members. DS-MGN achieves state-of-the-art performance compared to nine strong baselines on both a public benchmark dataset and the dataset we collected.

Mengjun Ding, Jia Yu, Danillo Barros de Souza, Serafim Rodrigues, Mathieu Desroches, Chunbo Li and Weiqiang Sun, *Growing order-heterogeneous simplicial complexes,* submitted, July 2025

Complex systems are often represented as graphs, where entities are nodes and interactions are edges. However, many real-world systems involve higher-order interactions that cannot be captured by pairwise interactions. Simplicial complexes provide a mathematical framework for modeling such higher-order systems, where interactions involving multiple nodes are represented as simplices (e.g., nodes, edges, triangles). In real-world systems, the size of interactions is often heterogeneous, with the number of participants in each interaction varying significantly. This paper proposes a growing order-heterogeneous simplicial complex model where the dimensions of added simplices follow a probability distribution. Theoretical analysis shows that the generalized degree of faces follows a power-law distribution, with exponents controllable through the simplex dimension distribution. Numerical simulations confirm these findings and demonstrate the model's ability to generate simplicial complexes with tunable structural properties. This work advances the modeling of higher-order networks by providing a flexible framework that better reflects the diversity of interactions in real-world systems.

Yu, J., Ding, M., Sun, W., Hu, W., & Wang, H. (2023). *"Understanding the power-law nature of participation in community sports organizations." IEEE Transactions on Computational Social Systems,* 2025, https://doi.org/10.1109/TCSS.2025.3532924

The improvement of living standards and awareness of chronic diseases have increased the importance of community sports organizations in promoting the physical activity levels of the public. However, limited understanding of human behavior in this context often leads to suboptimal resource utilization. In this study, we analyzed the participation behavior of 2,956 members with a time span of 6 years in a community sports organization. Our study reveals that, at the population level, the participation frequency in activities adheres to a power-law distribution. To understand the underlying mechanisms driving crowd participation, we introduce a novel behavioral model called HFBI (Habit-Formation and Behavioral Inertia), demonstrating a robust fit to the observed power-law distribution. The habit formation mechanism indicates that individuals who are more engaged are more likely to maintain participation, while the behavioral inertia mechanism suggests that individuals' willingness to participate in activities diminishes with their absences from activities. At the individual level, our analysis reveals a burst-quiet participation pattern, with bursts often commencing with incentive activities. We also find a power-law distribution in the intervals between individual participations. Our research offers valuable insights into the complex dynamics of human participation in community sports activity and provides a theoretical foundation to inform intervention design. Furthermore, the flexibility of our model enables its application to other data exhibiting power-law properties, broadening its potential impact beyond the realm of community sports.

Ding, M., Yu, J., & Sun W. *"Growing simplicial complex with face dimension selection and preferential attachment." Chaos: An Interdisciplinary Journal of Nonlinear Science* 34.10 (2024).

When simplicial complexes are used to represent higher-order systems, information regarding when and how interactions happen may be lost. In this paper, we propose the concept of temporal simplicial complexes, in which simplices with timestamps (or temporal simplices) are used to represent interactions, and faces with weights are used to represent relations. Then, we propose a growing model with two rules, face dimension selection (FDS), and preferential attachment. By properly setting the probability parameter vector q in the FDS rule, one can

balance network diameter expansion and network centrality, thus attaining more flexibility in the growing process. Our theoretical analysis and simulations that followed show the generalized degree of faces of any dimension follows a power-law distribution, with a scaling component controlled by q. Our work provides a flexible growing model and can be used to study higher-order systems with temporal properties.

Wang, Q., Yu, J., Ding, M., & Sun, W.. *"Twotier--A Layered Analysis of Backbone Members in a Moderate Sized Community Sports Organization." arXiv preprint arXiv:2307.04118.*
Backbone members are recognized as essential parts of an organization, yet their role and mechanisms of functioning in networks are not fully understood. In this paper, we propose a new framework called Twotier to analyze the evolution of community sports organizations (CSOs) and the role of backbone members. Tier-one establishes a dynamic user interaction network based on grouping relationships, and weighted k-shell decomposition is used to select backbone members. We perform community detection and capture the evolution of two separate sub-networks: one formed by backbone members and the other formed by other members. In Tier-two, the sub-networks are abstracted, revealing a core-periphery structure in the organization where backbone members serve as bridges connecting all parts of the network. Our findings suggest that relying on backbone members can keep newcomers actively involved in rewarding activities, while non-rewarding activities solidify relations between backbone members.

J. Yu, M. Ding, Q. Wang, W. Sun and W. Hu, *"Community Sports Organization Development From a Social Network Evolution Perspective— Structures, Stages, and Stimulus," in IEEE Transactions on Computational Social Systems,* vol. 10, no. 3, pp. 878-889, June 2023, doi: 10.1109/TCSS.2021.3135809.
As people become more health-conscious, more community sports organizations (CSOs) are emerging in society. The development of CSOs is often subject to very limited resources, and understanding its underlining principles is critical to maximizing the utility of available resources. Existing studies of CSOs are often empirical and adopt top-down methods of organizational management. This is partly because the inputs needed for such studies, for example, the interactions between community members, are often difficult to record. In this article, we perform an objective social network analysis on a real CSO with 2073 active members, with data collected over a five-year span on a mobile platform. Our study shows that the CSO members' network has a low density and weak densification and is assortative and obvious in community structure. With a classification algorithm called SC-CEE, we observe that the development of the CSO can be robustly divided into stages. We further observe that the relationship between specific stimulus and community evolution varies across development stages, implying that an organization's response strategy to change should be aligned with its stage of development. We then give practical guidance on the developmental priorities and response strategy of CSOs at different stages.

Zhou, Q., Yu, J., and Sun, W., *"Formation of a Community: in the Case of a Particular Non-profit Sports Organization, " 2020 International Conference on Computing, Networking and Communications (ICNC),* Big Island, HI, USA, 2020, pp. 844-848, doi: 10.1109/ICNC47757.2020.9049688.
Non-profit sports organizations play a more important role in our society than ever before, as the public is now more aware of their personal health and have recognized the importance of a community for persistent sport exercise. The long-term interactions of members in such organizations allow researchers to observe the structure of the organizations from the perspective of a dynamic social network. With this paper, we study the formation of a non-profit sports organization, with data collected over more than four years. Our main goal is to understand the dynamics in the formation and to find how the formation is affected by external stimulus. We build a network based on members' interactions in organizational activities. And mining the dynamic characteristics of the community for the purpose of looking for the relationship between the dynamics and external incentives. Community evolution-based method is used in analyzing the dynamic network. The results of the study indicate that specific incentive behaviors have a definite impact on the communities' evolution, and provide useful insights in building similar non-profit sports organizations, especially when the available resource for doing that is limited.

Modeling and Analysis of physical activity data (and other data)

Zhihua Ding, Keqin Shi, Zhen Chen, Zida Liang, Weiqiang Sun and Weisheng Hu, *Detecting Change Points in Time Series with Cosine Similarity,* submitted, June 2025
This paper introduces DCS-CPD, a novel method for change point detection in time series using cosine similarity. The method constructs a cosine similarity series that represents the statistical properties of the original series, allowing for the detection of change points. When no change points are present, the series follows a specific distribution, while deviations indicate potential change points. DCS-CPD derives the relationship between the change point probability and cosine similarity, enabling precise localization and probability estimation. Compared to traditional methods, DCS-CPD offers superior computational efficiency and broader applicability. Experiments on synthetic and real-world datasets validate its accuracy, robustness, and ability to provide reliable probability estimates for detected change points.

Shi, K., Ding, Z., Chen, Z., Sun, W., ***A Multi-scale Time Series Forecasting Framework with Temporal Hierarchical Information Fusion and Reconciliation,*** **Data Mining and Knowledge Discovery, 39, 4 (2025). https://doi.org/10.1007/s10618-025-01103-9**

Time series forecasting plays a crucial role in numerous applications. Views at different scale may offer complementary information and contain temporal coherence of a time series, but are largely neglected by existing studies. In this paper, we aim to take advantage of the complementarity and consistency among multiple scales of a time series to enhance accuracy of forecasting models. To accomplish this, we propose a Multi-Scale forecasting Framework (MSF) that comprises two key components: the Temporal Hierarchical Information Fusion (THIF) module and the Reconciliation Network (ReconNet). The THIF module facilitates the exchange of complementary information among scales, enabling a more comprehensive representation of each scale. The ReconNet, in a top-down manner, dynamically revises predictions to ensure coherence and enhance accuracy. The framework is designed to be flexible, such that different forecasting models can be plugged in, and is capable of joint training of the components. Extensive experiments on real-world datasets indicate the effectiveness of our approach.

Shi, K., Chen, Z., Sun, W. et al. ***"Measuring regularity of human physical activities with entropy models." Journal of Big Data 11, 37 (2024). https://doi.org/10.1186/s40537-024-00891-z***

Regularity is an important aspect of physical activity that can provide valuable insights into how individuals engage in physical activity over time. Accurate measurement of regularity not only advances our understanding of physical activity behavior but also facilitates the development of human activity modeling and forecasting. Furthermore, it can inform the design and implementation of tailored interventions to improve population health outcomes. In this paper, we aim to assess the regularity of physical activities through longitudinal sensor data, which reflects individuals' all physical activities over an extended period. We explore three entropy models, including entropy rate, approximate entropy, and sample entropy, which can potentially offer a more comprehensive evaluation of physical activity regularity compared to metrics based solely on periodicity or stability. We propose a framework to validate the performance of entropy models on both synthesized and real-world physical activity data. The results indicate entropy rate is able to identify not only the magnitude and amount of noise but also macroscopic variations of physical activities, such as differences on duration and occurrence time. Simultaneously, entropy rate is highly correlated with the predictability of real-world samples, further highlighting its applicability in measuring human physical activity regularity. Leveraging entropy rate, we further investigate the regularity for 686 individuals. We find the composition of physical activities can partially explain the difference in regularity among individuals, and the majority of individuals exhibit temporal stability of regularity.

Chen, Z., Shi, K., and Sun, W., ***"Step Count Print: A Physical Activity-Based Biometric Identifier For User Identification and Authentication." IEEE Transactions on Biometrics, Behavior, and Identity Science (2024). https://doi.org/ 10.1109/TBIOM.2024.3466269***

Step count is one of the most widely used physical activity data and is easily accessible through smart phones and wearable devices. It records the intensity and happening time of a user's physical activities, and often reflects a users' unique way of living. Incorporation of step count into biometric systems may thus offer an opportunity to develop innovative, user-friendly and non-invasive strategies of user identification and authentication. In this paper, we propose Step Count Print (SCP), a physical activity-based novel biometric identifier. Extracted from coarse-grained minute-level physical activity data (step counts), SCP contains features, including user step cadence distribution and average step distribution etc., that reflect an individual's physical activity behavior. With data collected from 100 users in a five-year long period, we conducted an ablation study to demonstrate the non-redundancy of SCP in user identification and authentication scenarios using commonly used machine learning algorithms. The results show that SCP can achieve a Rank-1 rate of up to 75.0% in user identification scenarios and an average accuracy of 92.3% in user authentication scenarios. In different classification algorithms, the user's accuracy histogram is drawn to demonstrate the universality of SCP and its effectiveness across a range of scenarios and use cases.

Z. Chen, K. Shi and W. Sun, ***Walking to Authenticate: Identifying Robust Behavioral Biometrics from Step Count Data,*** **2023 International Conference on Cyber-Enabled Distributed Computing and Knowledge Discovery (CyberC)*, Jiangsu, China, 2023, pp. 393-396, doi: 10.1109/CyberC58899.2023.00067.***

Step count data, readily available from smartphones and wearable devices, offers valuable insights into a user's physical activities and lifestyle patterns, positioning it as a viable contender for integration into biometric authentication systems. In this study, we introduce Step Count Print (SCP), a novel behavioral biometric derived from coarse-grained minute-level step count data, featuring daily step count distribution to capture an individual's unique physical activity pattern. We conduct an ablation study with data from 100 users collected over a five-year period, utilizing mainstream machine learning algorithms. Our experimental results demonstrate SCP's non-redundancy in user authentication scenarios, achieving an impressive average accuracy of 92.3%. We showcase the universality and effectiveness of SCP across various classification algorithms through user accuracy

histograms. Our proposed strategy holds promise for user-friendly and reliable biometric authentication, leveraging step count data to enhance security and usability in diverse applications.

Z. Ding, K. Shi and W. Sun, "GRU-CP: Non-stationary Time Series Forecasting with Change Point Detection," *2023 International Conference on Cyber-Enabled Distributed Computing and Knowledge Discovery (CyberC)*, Jiangsu, China, 2023, pp. 1-9, doi: 10.1109/CyberC58899.2023.00012. (Best Student Paper)

Time series forecasting finds numerous applications in real-world scenarios such as finance, healthcare, and weather prediction. However, a significant portion of real-world time series data exhibits non-stationarity. While some methods like Prophet and STL can decompose non-stationary time series due to seasonality and periodicity for accurate predictions, they may falter when faced with unforeseeable factors, such as structural breaks induced by government interventions in exchange rates. This study asserts that non-stationary time series can be segmented into multiple stationary segments via change points, and these segments may possess latent temporal characteristics. By employing change point detection algorithms to identify non-stationary features, these features can be incorporated into prediction models. The proposed GRU-CP model is introduced in this context. Compared to traditional models, GRU-CP yields lower errors when forecasting time series data.

Shi, Keqin, et al. *"Inferring Activity Patterns from Sparse Step Counts Data with Recurrent Neural Networks." ACM Transactions on Computing for Healthcare (ACM Health) 4.1 (2023): 1-20, also presented on CHASE – IEEE/ACM Conference on Connected Health: Applications, Systems and Engineering Technologies, Delaware, USA, June 2024*

As an accurate measurement of physical activity, step counts data can be collected expediently by smartphones and wearable devices. Complete and high time-resolution step counts data record the time and intensity of individuals' physical activity in a day, and can be used to mine activity habits or to recommend customized workout plans. However, sparse step counts data are common in practice due to hardware and software limitations. Understanding the value of sparse step counts data can contribute to its application in healthcare, and also can help us design cost-effective hardware and software. In this article, we aim to infer activity patterns from sparse step counts data. We design a deep learning model based on recurrent neural networks, namely MLP-GRU, which considers bidirectional short-term dependency and long-term regularity of sparse step counts data, and implements data-driven imputation and classification. We also develop an interpretable and elastic method to obtain sparse step counts data labeled with multi-granular activity patterns to train MLP-GRU. Evaluations on real-world datasets reveal that MLP-GRU outperforms other strong baseline methods. The results also show that activity patterns can be inferred from extremely sparse step counts data with high accuracy, provided that proper granularity is used for data of different sparsity.

website: http://health.sjtu.edu.cn/infer/

Dong, Jiapeng, Pengju Wang, and Weiqiang Sun. "Cost of Dietary Data Acquisition with Smart Group Catering." *Intelligent Computing: Proceedings of the 2020 Computing Conference*, Volume 1. Springer International Publishing, 2020.

The need for dietary data management is growing with public awareness of food intakes. As a result, there are increasing deployments of smart canteens where dietary data is collected through either Radio Frequency Identification (RFID) or Computer Vision(CV)-based solutions. As human labor is involved in both cases, manpower allocation is critical to data quality. Where manpower requirements are underestimated, data quality is compromised. This paper has studied the relation between the quality of dietary data and the manpower invested, using numerical simulations based on real data collected from multiple smart canteens. We found that in both RFID and CV-based systems, the long-term cost of dietary data acquisition is dominated by manpower. Our study provides a comprehensive understanding of the cost composition for dietary data acquisition and useful insights towards future cost effective systems.

Contact: Prof. Weiqiang Sun (sunwq@sjtu.edu.cn)
The school of Information and Electronic Engineering, Shanghai Jiao Tong University, Shanghai, China
To get updated version of this document: http://health.sjtu.edu.cn/papers/SJTU-Health-Project-Information.pdf