Data Science

LSE Department of Statistics

PhD Open Day, 28th November 2022

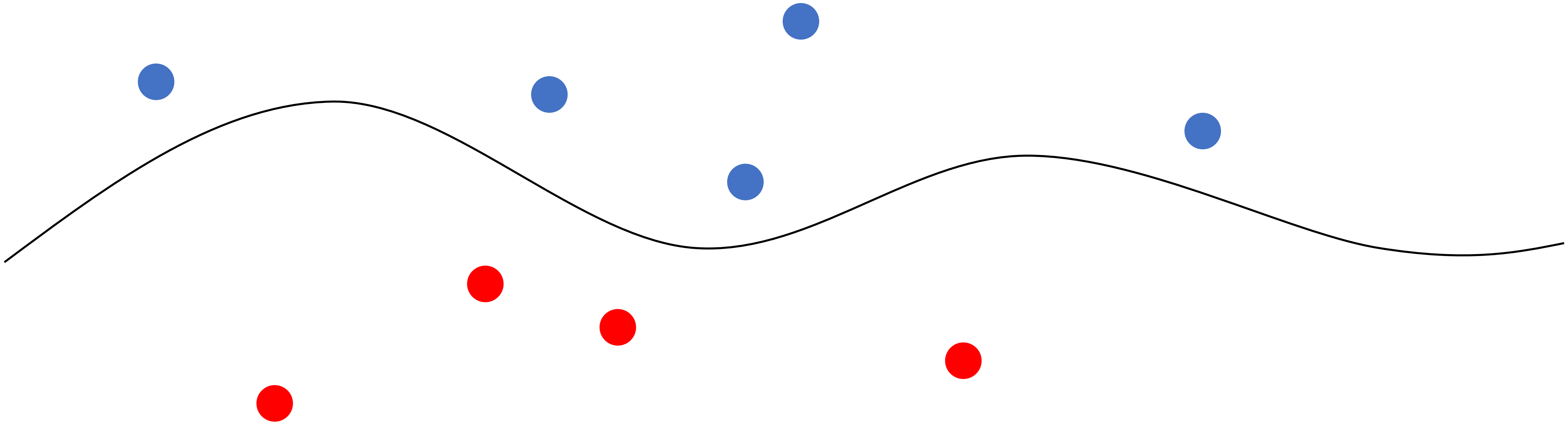
Department of Statistics: research groups

- Data Science
- Probability in Finance and Insurance
- Social Statistics
- Time Series and Statistical Learning

To probe further: https://www.lse.ac.uk/Statistics/Research

Data Science group

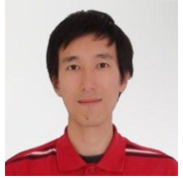
- Focus on development of machine learning and statistical methods, their theoretical foundations and applications



Data Science group: people

Mona Azdakia
Assistant Professor

Marcos Barreto
Assistant Professorial Lecturer



Yining Chen
Associate Professor

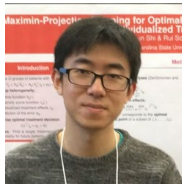
Moez Draief
Visiting Professor in Practice
Chief Scientist and VP Capgemini

Kostas Kalogeropoulos
Associate Professor

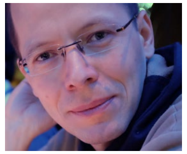
Joshua Loftus
Assistant Professor

Xinghao Qiao
Associate Professor

Francesca Panero
Assistant Professor



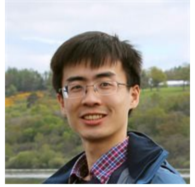
Chengchun Shi
Assistant Professor



Zoltan Szabo
Professor

Milan Vojnovic
Professor

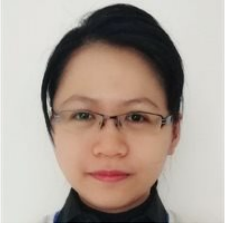
Christine Yuen
Assistant Professorial Lecturer



Tengyao Wang
Associate Professor

Research students

Sakina Hansen



Ziqing Ho

Liyuan Hu

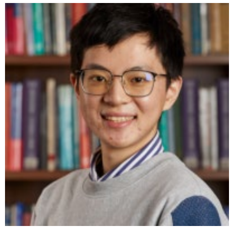
Yirui Liu

Tao Ma

Filippo Pellegrino

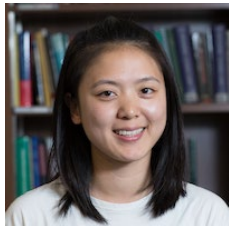
Pingfan Su

Yiliu Wang



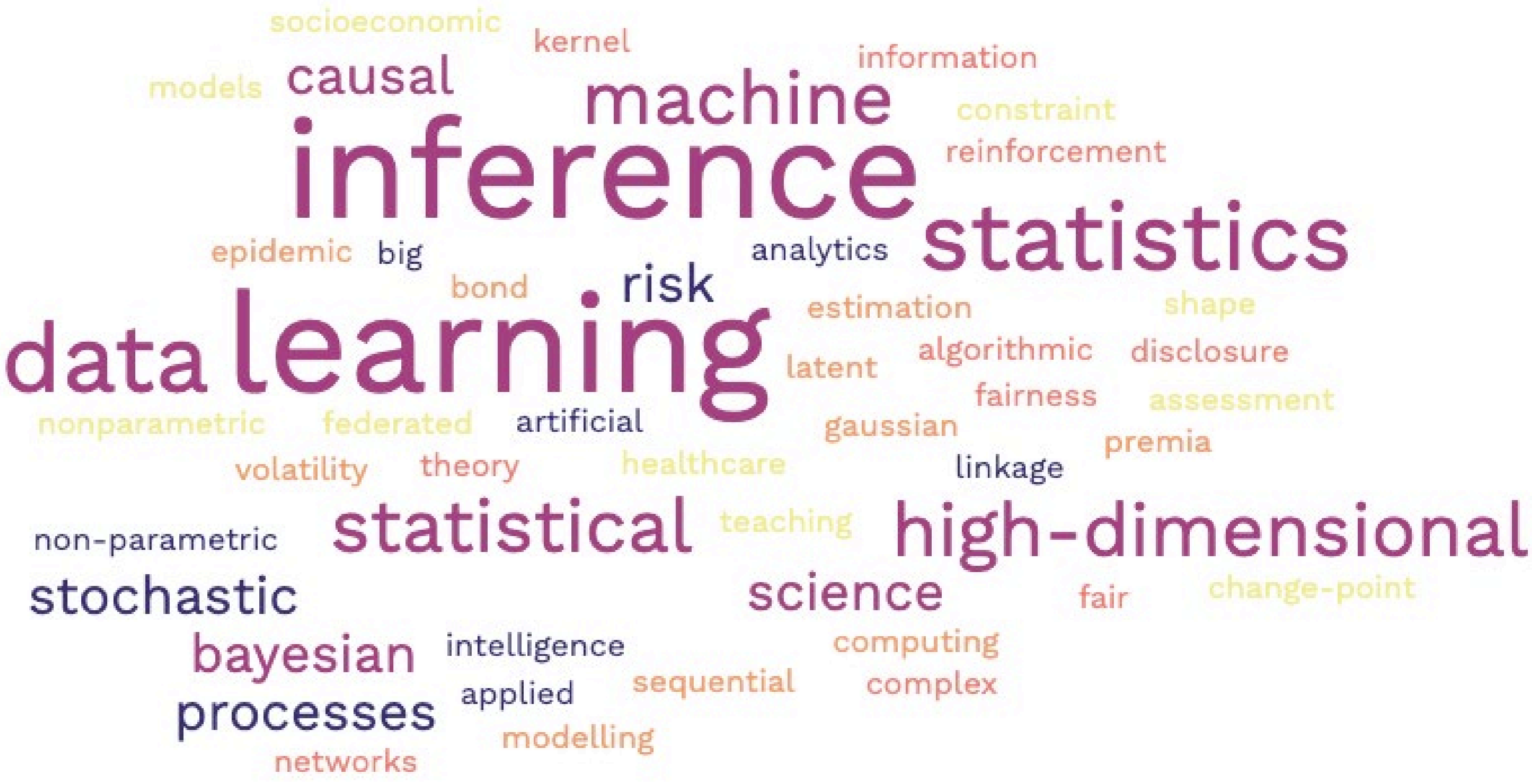
Xuzhi Yang

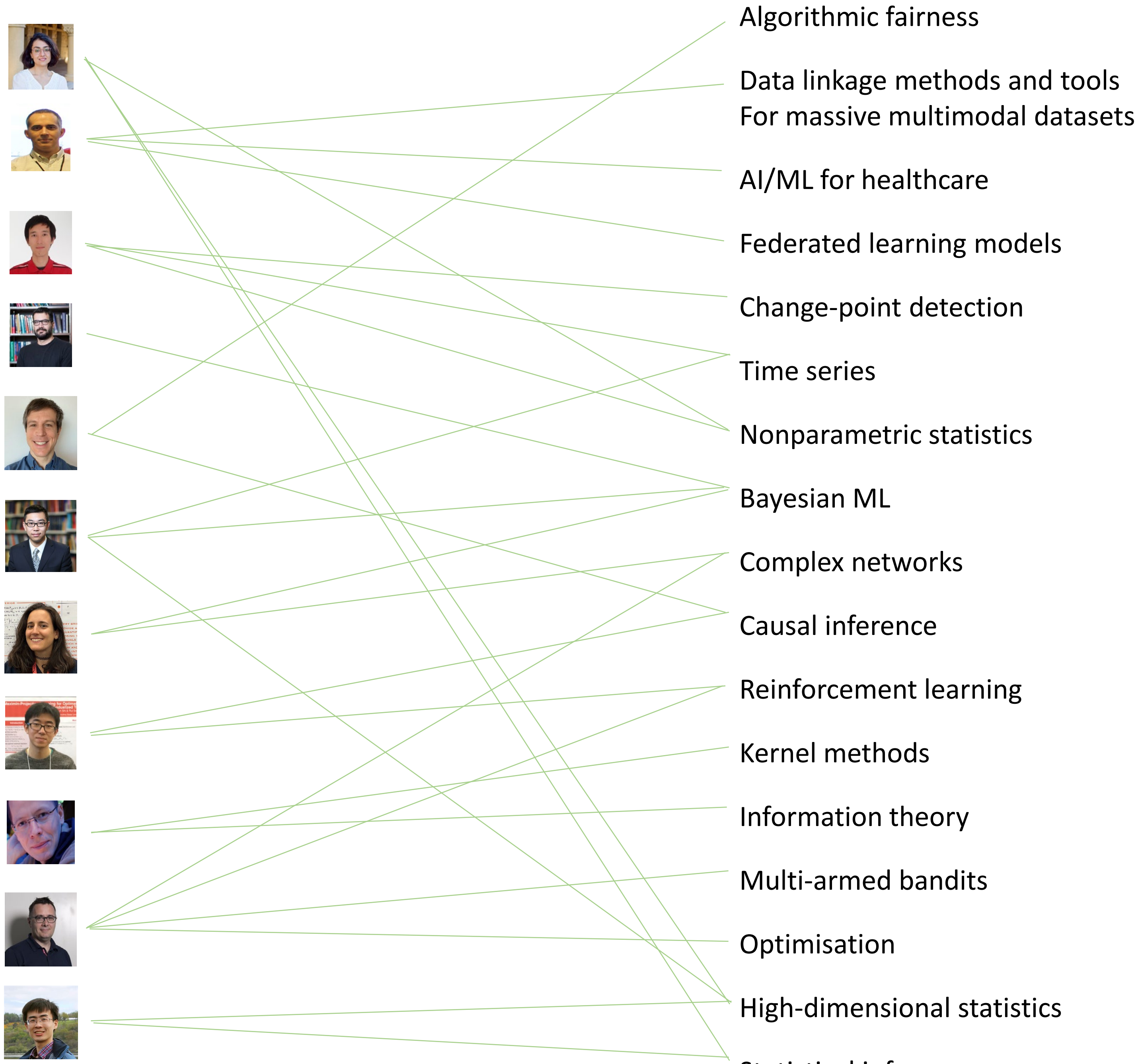
Jialin Yi



Kaifang Zhou

Research interests





Algorithmic fairness

Data linkage methods and tools
For massive multimodal datasets

AI/ML for healthcare

Federated learning models

Change-point detection

Time series

Nonparametric statistics

Bayesian ML

Complex networks

Causal inference

Reinforcement learning

Kernel methods

Information theory

Multi-armed bandits

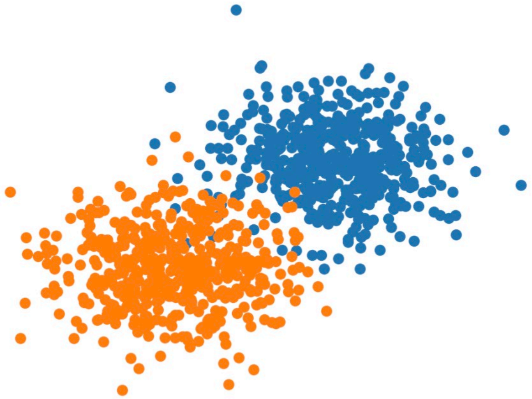
Optimisation

High-dimensional statistics

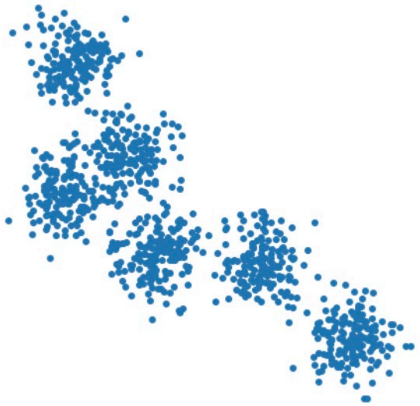
Statistical inference

Disclaimer: information may be incomplete

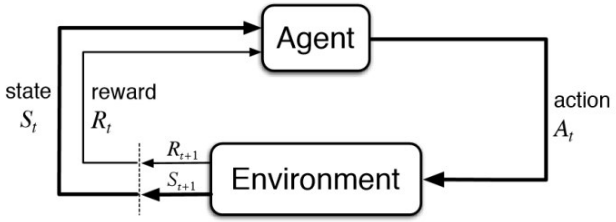
Some problems studied by our group



supervised learning



unsupervised learning

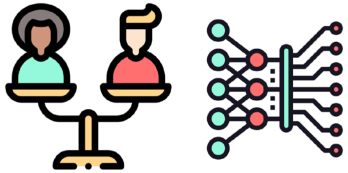


reinforcement learning

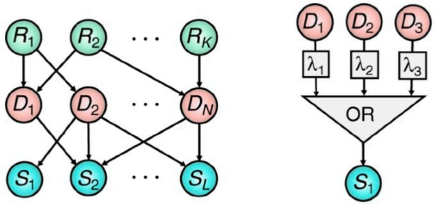
Some challenges:

- high-dimensional models (many parameters)
- complex models (e.g. deep neural networks)
- complex problems (e.g. function classes)
- unknown fundamental limits of learning, need for new algorithms

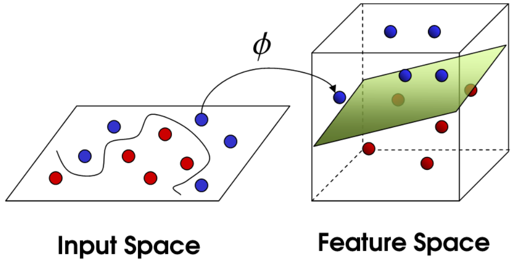
Some problems studied by our group (cont'd)



Algorithmic fairness

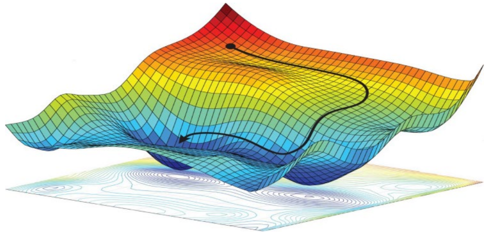


Causal inference

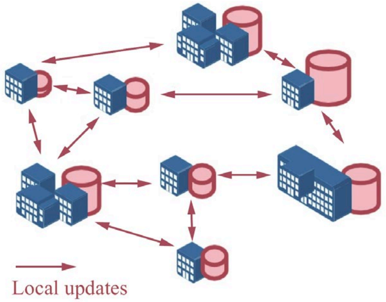


Kernel methods

Network data



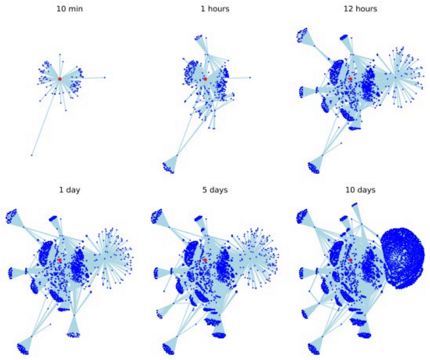
Optimization



Federated learning

Some applications of our research

- Computer and information systems
  - Anomaly detection in computer systems
  - Meta content moderation platform



- Healthcare
  - Precision medicine
  - The Cambridge kidney assessment tool

- Information Theoretical Estimators – Python toolbox

ST510 Foundations of Machine Learning

- A PhD level course, taught by several data science group members

- Foundations of supervised learning
- Convex optimization
- Non-convex optimization
- Support vector machines
- Decision trees and random forests
- Neural networks
- Unsupervised learning – clustering
- Unsupervised learning – dimensionality reduction
- Online learning and optimization
- Reinforcement learning

ST510 **Half Unit**

Foundations of Machine Learning

Teaching programmes

- MSc Data Science
- MSc Health Data Science (jointly with Department of Health Policy)
- BSc Data Science
- BSc Data Science and Business Analytics online degree programme

* Our data science courses are taken by students from different LSE departments

MSc Data Science

- Compulsory modules
  - ST443 Machine Learning and Data Mining
  - ST445 Managing and Visualising Data
  - ST447 Data Analysis and Statistical Methods
  - ST498 Capstone Project
- Optional modules
  - ST444 Computational Data Science
  - ST446 Distributed Computing for Big Data
  - ST449 Artificial Intelligence
  - ST451 Bayesian Machine Learning
  - ST455 Reinforcement Learning
  - ST456 Deep Learning
  - …

MSc Data Science: capstone project partners

alpha Telefónica

AstraZeneca

Capgemini

facebook

Microsoft

SAMSUNG

Adobe

Google

IBM

Microsoft Research

proximity

TESCO

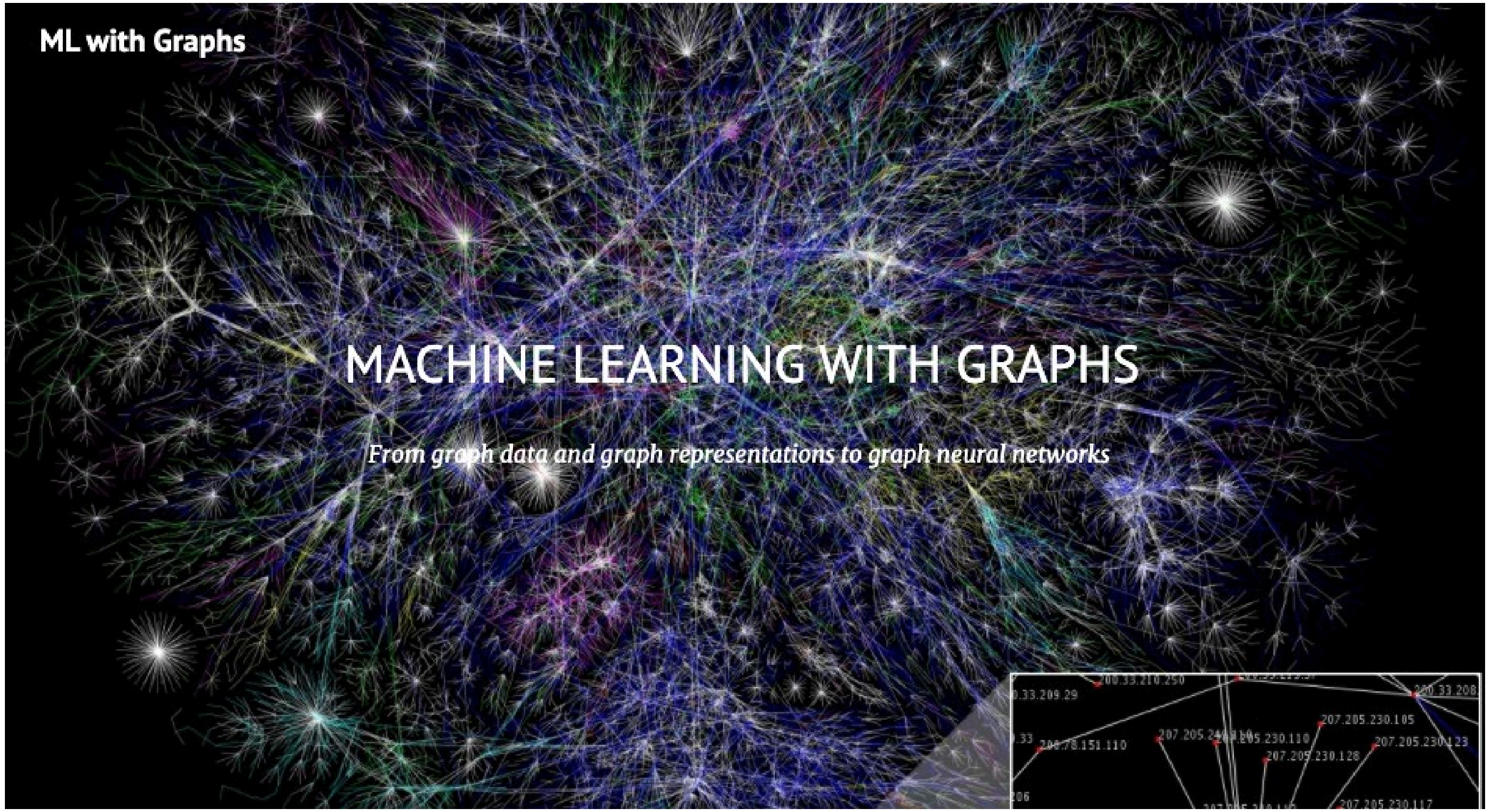
Data science seminar series

- Goal: promote research related to machine learning, computer science, statistics and their interface

- Some speakers:
  - Caroline Uhler, MIT
  - Vladimir Vovk, Royal Holloway, University of London
  - Anastasia Borovykh, Imperial College London
  - Erwan Scornet, Ecole Polytechnique

- To probe further: https://www.lse.ac.uk/Statistics/Seminars/Data-Science-Seminar-Series

Course on machine learning with graphs



ML with Graphs

MACHINE LEARNING WITH GRAPHS

From graph data and graph representations to graph neural networks

- Dr. Moez Draief

- 16-30th March 2021
- 6-7th May 2021

To probe further: https://sites.google.com/view/lsegraphrepresentations/home

Awards, grants and industry collaborations

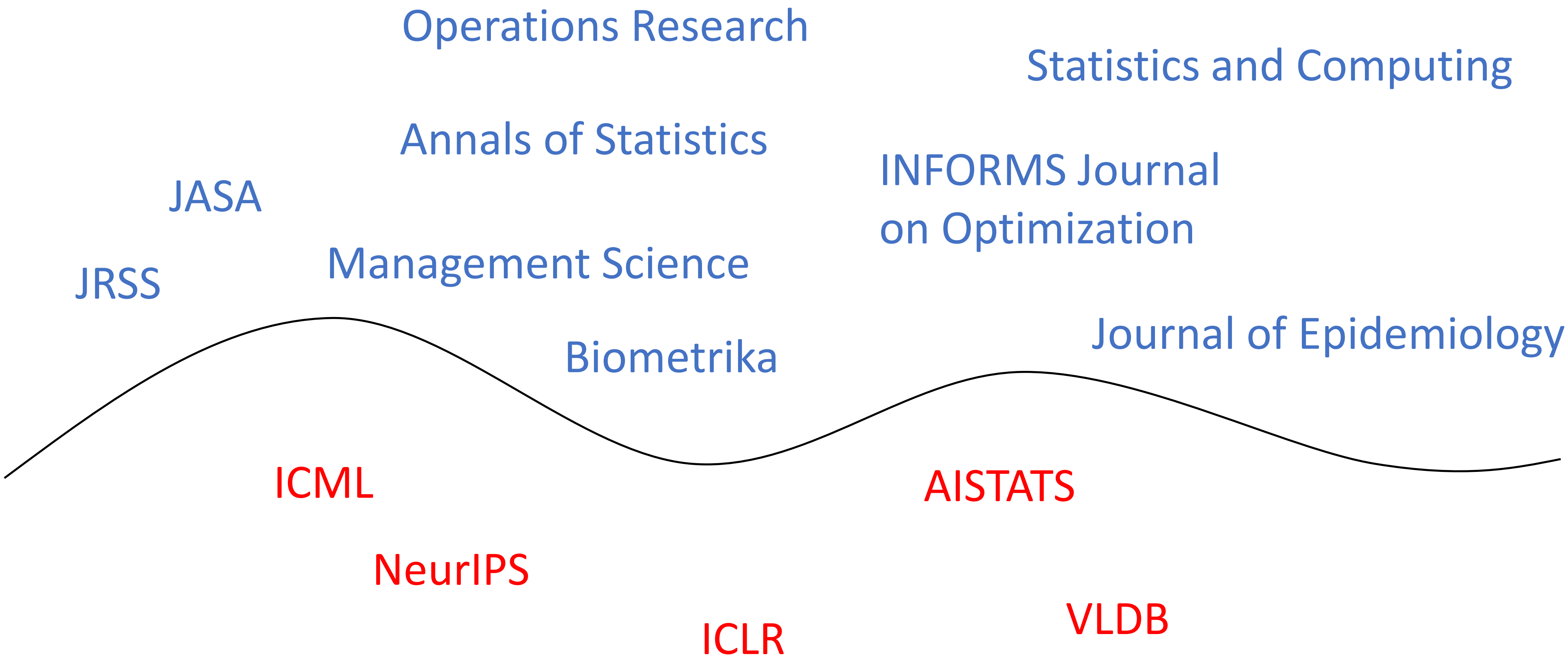
- Chengchun Shi, 2021 Royal Statistical Society Research Prize
- Zoltan Szabo, 2017 NeurIPS Best Paper Award

- Andurand Capital, 2021 – forecasting price movements
- Facebook, Systems for ML, 2020 – scheduling jobs with complex delay costs in data processing platforms
- Criteo AI, 2018 – matching learning algorithms
- Huawei, 2018 – anomaly detection

- Facebook / Meta 2019 -, Core Data Science – machine learning algorithms for content moderation in online platforms, active learning

- Our PhD students undertaking internships, e.g. Microsoft Research

Some publication outlets



Operations Research

Statistics and Computing

Annals of Statistics

INFORMS Journal
on Optimization

JASA

Management Science

JRSS

Journal of Epidemiology

Biometrika

ICML

AISTATS

NeurIPS

VLDB

ICLR

Future research directions

- Fair and responsible machine learning methods
- Safety-critical learning
- Causal inference, targeted learning
- Interpretable machine learning
- Object-oriented data analysis
- Machine learning for complex time series
- Data-driven algorithms
- Reinforcement learning
- Transfer learning
- …