

# YONGSHEN HE

12132901@mail.sustech.edu.cn | +86 177 2470 2474 | 1088 Xueyuan Avenue, Shenzhen, 518055 | [megasunny](#)  
Research interest: text mining, scientific mobility, computational social science, applied statistics.

## EDUCATION

M. Sc., Mathematics, **Southern University of Science and Technology**  
GPA: 3.45/4

Shenzhen, China  
2021.9 - 2023.12

B. Sc. - Statistics, **Shenzhen University**  
GPA: **4.15**/4.5, ranking **1**/46.

Shenzhen, China  
2017.9 - 2021.6

## RESEARCH EXPERIENCE

**Knowledge graph construction and inference based on literature data**  
Investigator, led by assistant professor [Yifang Ma](#)

Shenzhen, China  
2024.1 - ongoing

- Recognized entities (such as medicines and drugs in medical fields and models and algorithms in STEM fields) through BERT-based models.
- Extract relations (such as indications of drugs and models and the based-on relation between models and algorithms) via prompt techniques.
- constructs knowledge graphs based on the extracted triples and analyzes their structures using statistics and network science methods.

**Semantic-Based Study on the Mobility of Researchers and the Shift of Research Interests**  
Master thesis, supervised by assistant professor [Yifang Ma](#)

Shenzhen, China  
2022.7 - 2023.11

- Implemented topic classification task combining document embedding techniques of TF-IDF, word2vec, and SPECTER with the classifier of logistics regression, random forest, and supporting vector machine.
- Embedded institutions into a low dimensional space combining word2vec and UMAP based on the scientists' moving trajectories.
- Conducted an empirical analysis of the effect of scientific mobility on the shift in research interests via propensity score weighting.

**Short Text Clustering Based on Word2vec Model**

Bachelor thesis, Supervised by associate professor [Songchen Li](#)

Shenzhen, China  
2020.11 - 2022.5

- Compared different document representation methods of word2vec embeddings clustering methods in clustering titles of articles.

**Health Information Recommendation Decision Based on Machine Learning**

Group leader, Supervised by associate professor [Yingjue Fang](#)

Shenzhen, China  
2018.11 - 2020.5

- Designed neural networks of Fasttext and TextCNN to distinguish the health news from the news on other topics.

## COURSE PROJECT

**Statistical Graph Models on Community Detection**

Instructor: associate professor [Qingyan Hu](#)

Shenzhen, China  
2022.2 - 2022.6

- Detected communities in the binary network with the probability-base model.

**High Dimensional Statistics on Square-root LASSO**

Instructor: associate professor [Zeng Li](#)

Shenzhen, China  
2022.2 - 2022.6

- Derive Conic Programming and CCD algorithm to solve the square-root LASSO.
- Implement simulate study to compare the rates of precise and recall of both LASSO and square-root LASSO.

## Categorical Data Analysis on American National Election Survey

Instructor: associate professor [Xiyun Jiao](#)

Shenzhen, China  
2022.2 - 2022.6

- Construct binomial regressions of different link functions to estimate the effect on the vote in the presidential election.
- Utilize AIC criteria to select of the vital factors of the vote.

## PUBLICATIONS

**Yongshen He**, Yurui Huang, Chaolin Tian, Shibing Xiang, Yifang Ma\*. Neural embeddings of scientific mobility reveal the stratification of institutions in China. Information Processing & Management, 61(3), 103702.

Yurui Huang, Jialong Guo, Chaolin Tian, Shibing Xiang, **Yongshen He**, Yifang Ma\*. Domestic Brain Circulation in China. Scientometrics. Under review.

## HONOR

Outstanding Graduate in Shenzhen University  
First Prize Scholarship in Shenzhen University

2021  
2018-2021

## TEACHING EXPERIENCE

Network Science and Computing, Teaching Assistant, SUSTech.  
Bayesian Statistics, Teaching Assistant, SUSTech.

2022.2 - 2022.6  
2022.2 - 2022.6

## RESEARCH INTEREST, LANGUAGE AND SKILL

Languages: IELTS 6, Chinese (Mandarin, Cantonese).  
Skills: Python, R, Latex.