

# YONGSHEN HE

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

## WORK EXPERIENCE

Research Assistant  
supervised by assistant professor [Yifang Ma](#)

Shenzhen, China  
2024.1 - ongoing

## RESEARCH EXPERIENCE

**AProject on Neural EmbeddingReveal the genderdifference in research topics**  
Collaborated with assistant professor [Yifang Ma](#)

Shenzhen, China  
2024.8 - ongoing

- Represented papers, authors and topics through Word2vec embedding.
- Calculated interdiscipline indexes and other axes to distinguish disciplines like nature and social, hard and soft.
- measuring the difference of research topics in genders interacting with research fields, countries, and other factors.

**Multidimensional feature Analysis, validation, and early identification of original science**  
Investigator, led by assistant professor [Yifang Ma](#)

Shenzhen, China  
2025.1 - 2025.7

- Constructed a validation dataset using award-winning scientific achievements and a comprehensive literature database to compare differences in originality metrics, like novelty and disruption.
- Developed early identification machine learning frameworks (both tree-based methods and logistic regression) for highly original research using the originality metrics.
- Analyzed feature importances, individual and interactive effects of predictive features with SHAP values.

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

Shenzhen, China  
2024.1 - 2024.8

- 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 basing 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's 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 classifiers of logistic regression, random forest, and support 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.

## 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. Journal of Data and Information Science. 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.