Title: Structured Knowledge is Still Essential to Understand Sciences

Abstract:
Scientific knowledge is encoded in a variety of data forms such as text, tables, relational databases, ontologies, taxonomies, knowledge graphs, etc. Other than text, all the data types are structured or semi-structured, which machines can extract the knowledge easier. Recently, language models pre-trained on unstructured text have demonstrated surprising performance in high school science exams and science question answering benchmarks. So, *is structured knowledge still essential for machines to understand sciences?* In this talk, I will introduce our collaborative work with medical scientists, data scientists, and biologists, which have been published in KDD 2019, EMNLP 2019, WWW 2020, WWW 2021, and KDD 2021. I will discuss a few kinds of essential knowledge that the unstructured text could not cover and structured data were missing, and provide realistic solutions to comprehensive scientific information understanding.

Bio:
Meng Jiang is an Associate Professor in the Department of Computer Science and Engineering at the University of Notre Dame. He received B.E. and PhD from Tsinghua University. He was a visiting PhD at CMU and a postdoc at UIUC. He is interested in data mining, machine learning, and natural language processing. His data science research focuses on graph and text data for applications such as question answering, query understanding, user modeling, material discovery, online education, and mental healthcare. He received the CAREER Award from the National Science Foundation. He has delivered 14 conference tutorials and organized 7 workshops. He is a Senior Member of ACM and IEEE.