Research Statement

Firms or agents are connected via various linkages that have significant impact on investor behavior, firm decision, asset pricing, business cycle, and systemic risk. My current work spans several fields linked to networks – Macroeconomics, Finance, Machine Learning, Theory, and the Chinese Economy. I am also an expert in big data and especially interested in using huge dataset to reveal a micro channel to support a vivid macro picture.

Under the theme, my work can be divided into three branches: i) Equity-holding network and its implication on corporate finance, governance, and monetary policy; ii) Innovation network and its implication on business cycle, asset pricing, and investment; iii) Use machine learning to learn hidden networks or predict in a linked world, especially incorporating investors' behavior bias.

Project on innovation networks

Innovation is nothing but learning from others. In this project, I construct a new and comprehensive patent dataset of U.S back to 1911 to study the implication of the innovation network on business cycles, asset pricing, and portfolio investment.

1. Innovation Network, Production Network, and Business Cycles (Job market paper, with Yucheng Yang)
2. Networks, Linking Complexity, and Cross-Predictability.
3. Production Networks, Non-Linear Effects, and Resource Reallocation (come out soon, draft available)
4. Networks, Long Run Risk, and Asset Pricing (come out soon, draft available)

There are several things worth mentioning on the construction of the patent datasets.

1) We obtain the patent documents from google websites and extract relevant information from the document.
2) To accrue patent to the final parent companies, one need to consider the dynamic change in the parent-subsidiary relationship since a significant fraction of patents issued by subsidiaries.
3) One need to trace the transaction of patents between companies – patent assignments.
4) Match the patent assignees with the final parent companies, need to handle fuzzy match (misspell, abbreviation, mis-order of companies etc.)

Project on equity-holding networks

In this project, we construct a proprietary and dynamic updated dataset covering the universal firms registered in China till 2020. This comprehensive dataset records detailed information on firm shareholders, outside investment, and historical update. Use the information on the historical shareholders, we construct the dynamic equity-holding networks that can be traced back to 1990. Combing with other datasets, we conduct a sequence of researches trying to understand several important questions – the impact of equity-holding networks on firm growth, innovation, and monetary policy, and the driving forces of network formation. There are several related papers

1. The Network Effects of Agency Conflicts (with Yiqing Xing and Rakesh Vohra)
2. Tiered Intermediation in Business Groups and Targeted SME Support (with Yu Shi and Robert Townsend)
3. Ownership Networks and Firm Growth – What Do Five Million Companies Tell About Chinese Economy? (with Franklin Allen, Junhui Cai, Qian QJ, Xian Gu, Linda Zhao)
4. State-Owned Enterprises in China Revised (with Junhui Cai, Xian Gu, and Linda Zhao, come out

soon, draft available)

5. Innovation and Equity Holding Networks (with Allen Franklin, Junhui Cai, Xian Gu, Jun "QJ" and Linda Zhao, Wharton-China Conference)
6. Entry, Exit, and Covid-19 (with Allen Franklin, Junhui Cai, Xian Gu, Jun "QJ", and Linda Zhao, in progress)

Project on machine learning

In most of the cases, the networking data is rare available because: i) the prohibitive collection cost of such dataset; 2) large noise on the linkage between agents or firms due to measure errors or sampling basic. In this project, we use machine learning to recover the underlying linkages between firms or agents, and then use the recovered linkage to study a sequence of topics – prediction, portfolio management, and information acquisition in linking work.

1. Semi-supervised Learning with Networking Data (with Junhui Cai (Wharton Stats), Dan Yang (HKU), Linda Zhao (Wharton Stats), come out soon, draft available)
2. Information Dispersion, Acquisition, and Cross Predictability – A Network Approach (with Wayne Gao (Penn Econ), Rakesh Vohra (Penn Econ and CIS), in progress)
3. Learning in Networks, Idiosyncratic Volatility, and Predictability. (with Junhui Cai (Wharton Stats), Linda Zhao (Wharton Stats), in progress)
4. Identifying Underlying Links, and Cross Predictability. (with Junhui Cai (Wharton Stats), Linda Zhao (Wharton Stats), in progress)